

INSTITUTO FEDERAL DO ESPÍRITO SANTO
GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

HENRIQUE HILUEY RORIZ PERIN

**GERAÇÃO DE VALOR COM BIG DATA E A INFLUÊNCIA DA CULTURA DE
DADOS NA TOMADA DE DECISÃO EM GRANDES EMPRESAS**

Cariacica

2024

HENRIQUE HILUEY RORIZ PERIN

**GERAÇÃO DE VALOR COM BIG DATA E A INFLUÊNCIA DA CULTURA DE
DADOS NA TOMADA DE DECISÃO EM GRANDES EMPRESAS**

Trabalho de Conclusão de Curso apresentado à
Coordenadoria do Curso de Engenharia de
Produção do Instituto Federal do Espírito Santo
como requisito parcial para obtenção do título de
Bacharel em Engenharia de Produção

Orientador: Prof.º.D.Sc. Guilherme Guilhermino
Neto

Cariacica

2024

(Biblioteca do *Campus* Cariacica do Instituto Federal do Espírito Santo)

P445g Perin, Henrique Hiluey Roriz.

Geração de valor com *Big Data* e a influência da cultura de dados na tomada de decisão em grandes empresas / Henrique Hiluey Roriz Perin. – 2024.

91 f. : il.; 30 cm.

Orientador: Guilherme Guilhermino Neto.

Monografia (graduação) – Instituto Federal do Espírito Santo, campus Cariacica, Curso Superior em Engenharia de Produção, 2024.

1. Big Data. 2. Processo decisório. 3. Grandes empresas. I. Guilhermino Neto, Guilherme. II. Instituto Federal do Espírito Santo. Campus Cariacica. III. Título.

CDD-21:

658.403

(Bibliotecária: Luciana Dumer CRB6-ES nº 662)

HENRIQUE HILUEY RORIZ PERIN

**GERAÇÃO DE VALOR COM BIG DATA E A INFLUÊNCIA DA CULTURA DE
DADOS NA TOMADA DE DECISÃO EM GRANDES EMPRESAS**

Trabalho de Conclusão de Curso apresentado à
Coordenadoria do Curso de Engenharia de
Produção do Instituto Federal do Espírito Santo
como requisito parcial para obtenção do título de
Bacharel em Engenharia de Produção

Aprovado em 06 de dezembro de 2024.

COMISSÃO EXAMINADORA

Prof. Guilherme Guilhermino Neto, D.Sc.
Instituto Federal do Espírito Santo
Orientador

Prof. Tiago José Menezes Gonçalves, D.Sc.
Instituto Federal do Espírito Santo
Membro da banca avaliadora

Prof. Luiz Henrique Lima Faria, D.Sc.
Instituto Federal do Espírito Santo
Membro da banca avaliadora

FOLHA DE APROVAÇÃO-TCC Nº 22/2024 - CAR-CCEP (11.02.19.01.08.03.10)

(Nº do Protocolo: NÃO PROTOCOLADO)

(Assinado digitalmente em 10/12/2024 12:47)
GUILHERME GUILHERMINO NETO
PROFESSOR DO ENSINO BASICO TECNICO E TECNOLOGICO
CAR-CCEP (11.02.19.01.08.03.10)
Matricula: 2151589

(Assinado digitalmente em 12/12/2024 15:01)
TIAGO JOSE MENEZES GONCALVES
PROFESSOR DO ENSINO BASICO TECNICO E TECNOLOGICO
CAR-CCEP (11.02.19.01.08.03.10)
Matricula: 2073974

(Assinado digitalmente em 11/12/2024 05:16) LUIZ
HENRIQUE LIMA FARIA
PROFESSOR DO ENSINO BASICO TECNICO E TECNOLOGICO
CAR-CCTA (11.02.19.01.08.03.02)
Matricula: 1670417

Visualize o documento original em <https://sipac.ifes.edu.br/documentos/> informando seu número: 22, ano: 2024, tipo: **FOLHA DE APROVAÇÃO-TCC**, data de emissão: 10/12/2024 e o código de verificação: **f5ad8bf47b**

AGRADECIMENTO

À Deus, que me deu um coração inquieto, que nunca cessa de buscar. Ele, que deu-me forças para continuar nos momentos mais desafiadores, e se mostrou presente em minha vida com seu silêncio entre as palavras.

À minha família, por todo amor e dedicação doados na minha educação. Em especial ao meu pai e minha mãe, que em muitos momentos de dificuldades não abriram mão de me proporcionar a melhor educação.

À minha namorada, Ruskaya, cuja paciência e incentivo foram essenciais ao longo deste último ano de graduação. Agradeço por compreender a distância dos domingos, dedicados à realização deste trabalho.

Aos professores e servidores do IFES. Tenho orgulho por ter estudado em uma Instituição Federal de Ensino e ter o privilégio de aprender com a grande bagagem de vocês.

Ao meu orientador, Prof °. Dr °. Guilherme Guilhermino Neto, por ter se disponibilizado em conduzir minha orientação, dando grande apoio e sugestões valiosas para esta pesquisa.

RESUMO

O objetivo deste estudo é analisar como as organizações lidam com o desafio de extrair valor da dimensão complexa do *big data*, e verificar se empresas que possuem uma cultura de dados são mais eficientes na tomada de decisão. Para isso, realizou-se um levantamento por meio de um questionário com 10 itens aplicado a 60 empresas previamente selecionadas com base em seu modelo de estruturação de dados. Essas empresas foram divididas em duas amostras intencionais: o primeiro grupo, composto por “Organizações Avançadas em Dados”, as quais viabilizam a cultura *data-driven*, e o segundo grupo, denominado como “Organizações Iniciais em Dados”, referente aquelas que ainda não viabilizam. A análise foi conduzida com o uso de análise exploratória de dados, testes de proporcionalidade Qui-quadrado e teste exato de Fisher. A análise exploratória, a partir de representações gráficas, permitiu avaliar de forma comparativa os desafios enfrentados por esses dois grupos em relação à geração de valor com *big data*, os benefícios obtidos e as percepções dessas empresas sobre aspectos relevantes da cultura de dados, como governança e estratégia de dados. Aplicaram-se os testes de proporção para validar se as organizações avançadas apresentam processos decisórios com maior utilização de dados e se a capacidade de geração de valor com dados está associada ao seu grau de maturidade técnica e analítica. Concluiu-se que a consolidação de uma cultura de dados é essencial para organizações que buscam eficiência na tomada de decisão e maior geração de valor a partir do *big data*. Apesar disso, este estudo apresentou limitações, como uma amostra restrita a 60 organizações e a ausência de métricas financeiras que poderiam aprofundar a análise dos impactos econômicos dessa cultura. Para estudos futuros, sugere-se ampliar a diversidade da amostra, incorporar indicadores financeiros como ROI e faturamento, e explorar as diferenças entre decisões operacionais e estratégicas.

Palavras-chave: *Big data*. Análise de *big data*. Cultura de dados. *Data-driven*. Tomada de decisão.

ABSTRACT

The objective of this study is to analyze how organizations address the challenge of extracting value from the complex dimension of big data and to determine whether companies with a data-driven culture are more efficient in decision-making. To achieve this, a survey was conducted using a 10-item questionnaire applied to 60 companies, previously selected based on their data structuring models. These companies were divided into two intentional samples: the first group, composed of "Data-Advanced Organizations," which enable a data-driven culture, and the second group, referred to as "Data-Beginner Organizations," representing those that have not yet implemented such a culture. The analysis was carried out using exploratory data analysis, Chi-square proportionality tests, and Fisher's exact test. The exploratory analysis, supported by graphical representations, provided a comparative assessment of the challenges faced by these two groups in generating value from big data, the benefits obtained, and their perceptions regarding key aspects of data culture, such as governance and data strategy. Proportion tests were applied to validate whether data-advanced organizations demonstrate decision-making processes with greater data utilization and whether their ability to generate value from data is associated with their level of technical and analytical maturity. The study concluded that the consolidation of a data-driven culture is essential for organizations seeking efficiency in decision-making and greater value generation from big data. Nonetheless, the study faced limitations, such as a restricted sample size of 60 organizations and the absence of financial metrics that could deepen the analysis of the economic impacts of this culture. For future studies, it is suggested to expand the sample diversity, incorporate financial indicators such as ROI and revenue, and explore the differences between operational and strategic decision-making processes.

Keywords: Big data. Big data analytics. Data culture. Data-driven. Decision-making.

LISTA DE FIGURAS

Figura 1 – Arquitetura do Big Data.....	24
Figura 2 – Fluxo de ETL.....	25
Figura 3 – Funcionamento do MapReduce	26
Figura 4 – Infraestrutura em nuvem.....	29
Figura 5 – Fluxo do CGBDA.....	31
Figura 6 – Modelo de processo para tomada de decisão de Simon.....	34
Figura 7 – Modelo de processo para tomada de decisão de Mintzberg.....	36
Figura 8 –Estrutura da cultura de dados.....	41

LISTA DE TABELAS

Tabela 1 – Classificação do processo decisório pelas OAD e OID.....	52
Tabela 2 – Classificação da assimilação do big data pelas OAD e OID.....	54
Tabela 3 – Classificação da maturidade técnica e analítica pelas OAD e OID.....	56
Tabela 4 – Ranking dos setores com maior utilização de dados.....	60
Tabela 5 – Ranking dos benefícios do DD.....	62
Tabela 6 – Frequência e percentual das linhas para a primeira associação.....	71
Tabela 7 – R - Estatística de teste para a primeira associação.....	71
Tabela 8 – R - Percentual das linhas para a segunda associação.....	73
Tabela 9 – R - Estatística de teste para a segunda associação.....	73

LISTA DE GRÁFICOS

Gráfico 1 – Previsão de dados gerados no mundo após 2021.....	15
Gráfico 2 – Representatividade do big data por setores econômicos.....	20
Gráfico 3 – Segmento de mercado das empresas participantes.....	50
Gráfico 4 – Diferenças na tomada de decisão entre OADs e OIDs.....	51
Gráfico 5 – Diferenças de assimilação e geração de valor com dados.....	54
Gráfico 6 – Comparativo do nível de maturidade para lidar com big data.....	55
Gráfico 7 – Concordância das OAD com as premissas da cultura de dados.....	57
Gráfico 8 – Concordância das OID com as premissas da cultura de dados.....	57
Gráfico 9 – Comparativo da utilização de dados por departamento.....	59
Gráfico 10 – Comparativo dos desafios da geração de valor com dados.....	61
Gráfico 11 – Comparativo dos benefícios obtidos com dados.....	62
Gráfico 12 – Importância atribuída pelas OAD aos pilares da cultura de dados.....	63
Gráfico 13 – Importância atribuída pelas OID aos pilares da cultura de dados.....	65
Gráfico 14 – Comparativo dos desafios para implementar a cultura de dados.....	66
Gráfico 15 – Principais ferramentas e abordagens no trabalho com dados.....	67

LISTA DE SIGLAS

BD - Big Data

IoT - Internet of Thing

IA - Inteligência Artificial

BDA - Big Data Analytics

BI - Business Intelligence

BV - Business Value

BI&A - Business Intelligence e Analytics

DD - Data-Driven

HDFS - Hadoop Distributed File System

SQL - Structured Query Language

ACID - Atomicidade, Consistência, Isolamento e Durabilidade

CGBDA - Capacidade de Gerenciamento de Big Data Analytics

TDBD - Tomada de Decisão Baseada em Dados

OAD - Organizações Avançadas em Dados

OID - Organizações Iniciantes em Dados

IFES - Instituto Federal do Espírito Santo

ROI - Retorno sobre o Investimento

SUMÁRIO

1	INTRODUÇÃO	15
1.1	PROBLEMA DE PESQUISA.....	15
1.2	JUSTIFICATIVA	17
1.3	OBJETIVOS.....	18
1.4	ORGANIZAÇÃO DO TRABALHO.....	19
2	REFERENCIAL TEÓRICO	20
2.1	<i>BIG DATA</i>	20
2.1.1	6vs big data	21
2.1.2	Tecnologia	23
2.1.2.1	Arquitetura do big data.....	23
2.1.2.2	Coleta de dados.....	24
2.1.2.3	Processamento.....	25
2.1.2.4	Armazenamento.....	27
2.1.2.5	Computação em nuvem	28
2.1.3	<i>Big data analytics</i>	30
2.1.3.1	Capacidades de gerenciamento de big data analytics.....	31
2.2	TOMADA DE DECISÃO.....	33
2.2.1	Processo decisório	33
2.2.1.1	Decisão baseada em dados.....	38
2.2.2	Cultura de dados	40
3	MATERIAL E MÉTODO	43

3.1	CARACTERIZAÇÃO DA PESQUISA.....	43
3.2	INSTRUMENTO DE COLETA DE DADOS.....	43
3.3	DEFINIÇÃO DA POPULAÇÃO E AMOSTRAS.....	45
3.4	PROCEDIMENTOS DE COLETA E ANÁLISE DOS DADOS.....	47
4	APRESENTAÇÃO E ANÁLISE DOS DADOS.....	50
4.1	ANÁLISE DO PERFIL DOS RESPONDENTES.....	50
4.2	ANÁLISE EXPLORATÓRIA DOS RESULTADOS.....	51
4.2.1	Processo decisório.....	51
4.2.2	Capacidade de assimilação do big data.....	53
4.2.3	Maturidade técnica para lidar com big data.....	55
4.2.4	Concordância com as premissas da cultura de dados.....	56
4.2.5	Utilização de dados por departamento.....	58
4.2.6	Desafios na geração de valor com dados.....	60
4.2.7	Benefícios obtidos através do data-driven.....	61
4.2.8	Cultura de Dados.....	63
4.2.9	Obstáculos para implementação da cultura de dados.....	66
4.2.10	Ferramentas mais utilizadas.....	67
4.3	ANÁLISE INFERENCIAL DOS DADOS.....	68
4.3.1	Teste Qui-quadrado.....	68
4.3.1.1	Teste exato de Fisher.....	69
4.3.2	Tipo de organização e o processo decisório.....	70
4.3.3	Nível de maturidade e capacidade de assimilação.....	72

5	CONSIDERAÇÕES FINAIS.....	75
	REFERÊNCIAS	78
	APÊNDICE A	84

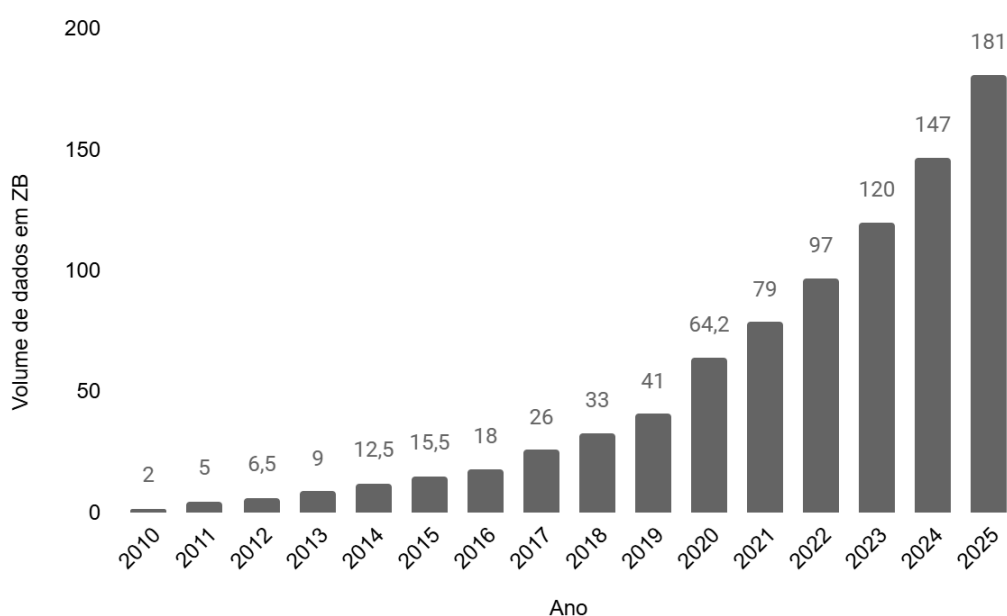
1 INTRODUÇÃO

Este capítulo é dedicado à apresentação do problema que motivou a realização deste trabalho, assim como à compreensão dos objetivos estabelecidos que justificam a busca por resultados através deste estudo.

1.1 PROBLEMA DE PESQUISA

A era moderna é marcada pela grande quantidade de dados circulantes nos meios digitais. A evolução tecnológica global tornou possível o compartilhamento de informações em velocidade instantânea. Segundo Chauhan e Sood (2021), os dados estão a ser criados em um ritmo exponencial. Prevê-se que o volume total de dados criados e consumidos em todo o mundo seja de 149 *zettabyte* (ZB) em 2024, quase duas ordens de grandeza acima dos 9 ZB em 2013 (Chauhan; Sood, 2021), como é apresentado no gráfico 1 (Berisha; Meziu; Shabani, 2022). Para colocar a dimensão em perspectiva, seriam necessários 181 milhões de anos se todos os dados existentes fossem descarregados (Chauhan; Sood, 2021).

Gráfico 1: Previsão de dados gerados no mundo após 2021



Fonte: adaptado de (Berisha; Meziu; Shabani, 2022)

Singularmente, o crescimento desenfreado dos dados enquadra-se no conceito de *Big Data* (BD), descrito em Sharda et al. (2019) por “6Vs”: volume, variedade, velocidade, veracidade, variabilidade e valor. O BD abrange tanto os dados estruturados quanto os não estruturados. Os conjuntos de dados estruturados são constituídos por tipos de dados definidos, facilmente organizados em bancos de dados relacionais. Os dados não-estruturados são aqueles que não são facilmente pesquisáveis ou armazenados em formato de base estrutural (Chauhan; Sood, 2021). Estas enormes quantidades de dados exigem métodos complexos para a sua manipulação e armazenamento, bem como abordagens estatísticas intensivas e modelos de programação para extrair informações relevantes (Misra et al., 2022).

Fenômenos recentes como a Internet das coisas (IoT - *Internet of Things*) e a evolução da Inteligência artificial (IA) refletem a ascensão da geração de valor através do *big data*. IoT pode ser definida como a capacidade de permitir uma comunicação robusta entre o mundo físico e o mundo digital a partir de um sistema condicionado por aprendizado de máquina e processamento de dados (Misra et al., 2022). O crescimento do BD a partir dos dispositivos de IoT proporcionou grandes oportunidades de melhorias nas operações comerciais, além de inovar os setores de pesquisa (Chauhan; Sood, 2021).

Para Dwivedi et al. (2021), o campo da IA envolve o desenvolvimento de sistemas computacionais capazes de realizar tarefas que normalmente requerem a inteligência humana, como a percepção sensorial e a tomada de decisão. A análise de *big data* (BDA - *Big Data Analytics*) combinado com a IA, tem o potencial de transformar diferentes áreas da manufatura, da saúde e da Inteligência de negócios (BI - *Business Intelligence*), oferecendo insights avançados num contexto preditivo (Alvarez et al., 2018).

Entretanto, Paradza e Daramola (2021), mostram que a real vantagem da análise de decisão dos negócios está, na verdade, atrelada ao quanto de valor as empresas conseguem extrair dos seus dados. Relatórios publicados pela Forrester em 2020 sugerem que entre 60 a 73% de todas as informações nunca são usadas para fins analíticos (Gualtieri, 2016) e, atualmente, cerca de 0.5% de todos os dados são analisados corretamente pelas empresas (Myler, 2017). A partir de estimativas, entre

80 a 85% dos dados organizacionais encontram-se em algum formato não estruturado (Sharda et al., 2019), um formato inacessível para propósitos analíticos. Nesse sentido, volume e variedade de dados obtidos nem sempre representam ganhos na perspectiva da tomada de decisão.

Uma hipótese para o surgimento dessa problemática decorre da falta de infraestrutura interna nas organizações para implementar o *big data analytics* e da carência de uma cultura de inteligência orientada por dados. A infraestrutura de BDA é projetada para facilitar a interação entre dados históricos e os atuais, permitindo a extração de *insights*. Isso envolve o uso de ferramentas e plataformas que suportam a integração dos dados e permitem a sua análise em tempo real, ou mediante demanda (Awan et al., 2021). Pesquisas indicam que as empresas com um alto nível de capacidades de BDA tendem a ter mais facilidade na geração de conhecimento útil (Acharya et al., 2018). No que tange à inteligência de dados, abordagens mais específicas são necessárias, como é o caso do BI.

BI é um termo que combina ferramentas analíticas e metodologias aplicadas. O principal objetivo do BI é possibilitar acesso interativo aos dados e oferecer aos analistas a capacidade de produzir análises informativas através da manipulação de dados (Sharda et al., 2019). Através da adoção do BI, as organizações obtêm o valor de negócio (BV - *Business Value*) adequado para manter sua eficiência no longo prazo (Paradza; Daramola, 2021).

1.2 JUSTIFICATIVA

À medida que as organizações avançam cada vez mais ao *big data*, a expansão heterogênea dos dados, tanto em termos de quantidade como de variedade, exigirá a evolução da infraestrutura e da análise de dados para processá-los (Chauhan; Sood, 2021). Um relatório publicado pela NewVantage Partners em 2022 revela que 91% das empresas entrevistadas aumentaram o seu investimento em iniciativas de dados, 26% conseguiram estabelecer uma cultura de dados, porém apenas 19% efetivamente a incorporaram em suas rotinas (Davenport; Bean, 2022). Esse fator vai em direção oposta à filosofia de Grant (1996), a qual aponta o conhecimento como o recurso estratégico mais importante das organizações.

Nesse contexto, muitas empresas reconhecem a importância de progredir tecnicamente com sua infraestrutura de dados, porém nem sempre direcionam o mesmo esforço para o desenvolvimento dos setores analíticos. Os números apontam para uma lacuna preocupante, especialmente quando comparados com os benefícios identificados por Akter et al. (2016) em empresas que aplicam eficazmente a análise de *big data*. Estas empresas evidenciaram aumentos de 70% na expansão de novos mercados, elevação de 79% na satisfação dos clientes, e melhoria de 76% nas vendas. Nesse cenário, existe uma oportunidade tangível de vantagem competitiva.

O presente estudo, a partir do entendimento de que em muitos casos as organizações são ineficientes na geração de informação útil com os dados, propõe fornecer a resposta para a seguinte questão: “Como as organizações podem enfrentar os desafios associados à extração de valor do *big data*, e de que maneira a cultura de dados influencia o sucesso na geração de ganhos e na melhoria do processo de tomada de decisão?”. Para isso, serão realizadas pesquisas exploratórias em diferentes organizações, com o objetivo de aprofundar a compreensão sobre como o processo decisório é estruturado nessas empresas, incluindo o uso da inteligência de dados e as práticas de governança associadas.

Cabe ressaltar que este trabalho se adequa às competências do engenheiro de produção, como listadas na resolução nº 235 de nove de outubro de 1975 do Conselho Federal de Engenharia, Arquitetura e Agronomia (CONFEA) que discerne a ocupação do Engenheiro de Produção (CONFEA, 1975). Além de estar inserido na área de Pesquisa Operacional e subárea de Processos Decisórios, de acordo com a Associação Brasileira de Engenharia de Produção (ABEPRO).

1.3 OBJETIVOS

Objetivo geral

Analisar como as organizações lidam com o desafio de extrair valor do amplo volume de dados disponíveis e verificar se a cultura de dados é um fator de sucesso para uma empresa atingir ganhos e melhorar sua tomada de decisão.

Para isso, espera-se alcançar os seguintes objetivos específicos:

- Explorar o conceito de big data e seu papel no processo decisório;
- Avaliar a influência de uma cultura data-driven na eficácia da tomada de decisão;
- Investigar a maturidade técnica e analítica das organizações na geração de valor a partir dos dados;
- Analisar a percepção sobre governança e cultura de dados em organizações de grande porte.

1.4 ORGANIZAÇÃO DO TRABALHO

Cada seção delineada aqui é fundamental para a abrangência e propósito deste estudo. Nos próximos parágrafos, serão apresentados os objetivos de cada seção.

O capítulo a seguir concentra-se no embasamento teórico, compilando os estudos relevantes acerca do tema, os pensamentos de autores pioneiros e contemporâneos, acompanhados de explicações sobre pesquisas anteriores.

O terceiro capítulo descreve o método e os recursos da pesquisa, onde detalha-se as etapas e atividades realizadas. Por conseguinte, é abordado a aplicação de uma pesquisa exploratória e inferencial em uma amostra de empresas selecionadas, bem como a análise dos resultados obtidos. Neste tópico, são condensadas as visões de diferentes organizações sobre os desafios e benefícios obtidos com o BD na tomada de decisão empresarial e um enquadramento dessa visão com o conceito de cultura de dados.

No quinto e último capítulo, são tratadas as conclusões e recomendações finais, as quais trazem considerações acerca deste estudo e a sua contribuição.

2 REFERENCIAL TEÓRICO

Este capítulo objetiva apresentar o referencial teórico dos principais conceitos analisados neste estudo. Inicialmente, será apresentado o conceito de *big data* e suas tecnologias. Por conseguinte, será analisado o conceito da tomada de decisão junto aos fundamentos do processo decisório e o seu vínculo com a cultura de dados.

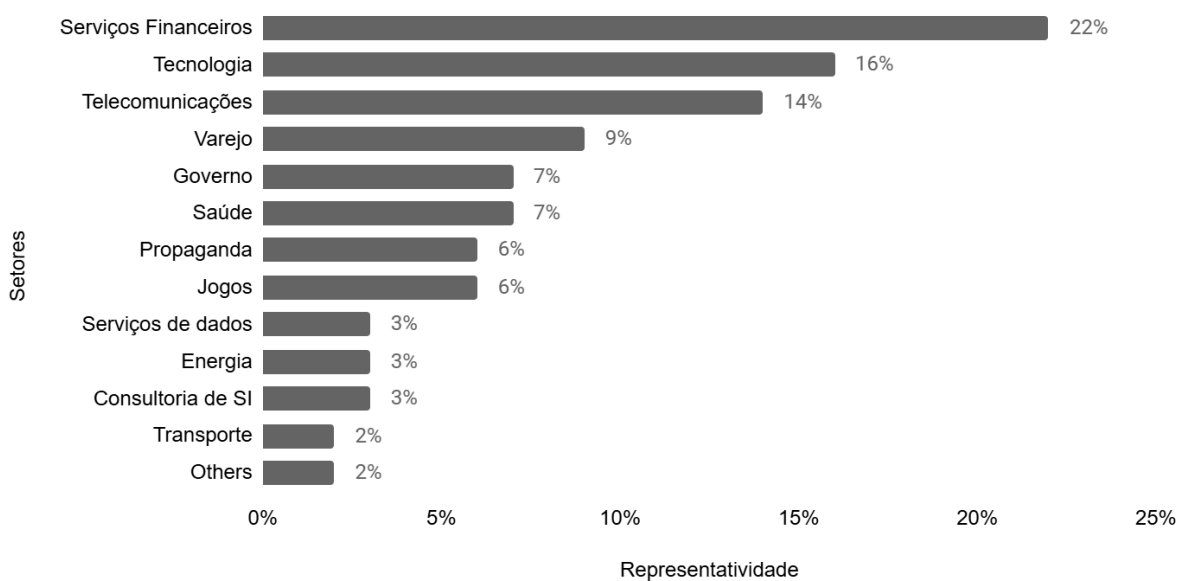
2.1 BIG DATA

De acordo com Sharda et al. (2019), BD tornou-se um termo subjetivo. Isso porque “*Big*” refere-se a uma das premissas do BD que é apresentar alto volume mas não existe uma delimitação quantitativa do tamanho em GB ou TB que se estabelece o BD. O termo “*big data*” surgiu no ano de 1997 quando Michael Coxand e David Ellsworth, investigadores da Nasa, o utilizaram pela primeira vez em um estudo apresentado na conferência do Instituto de Engenheiros Eletricistas e Eletrônicos (IEEE) para explicar grandes conjuntos de dados (Wang; Kung; Byrd, 2018). Neste estudo, os pesquisadores abordaram os desafios associados ao processamento dos conjuntos de dados de entrada (*input*), os quais eram superiores a 100GB. Na época, essa dimensão ultrapassava a capacidade de memória dos computadores, originando o problema de “*Big Data*” (Cox; Ellsworth, 1997).

Segundo Oussous et al. (2017) *big data* pode ser definido como grandes conjuntos de dados em crescimento, que incluem formatos heterogêneos: estruturados, não estruturados e semiestruturados. O BD tem uma natureza complexa que exige tecnologias e algoritmos avançados .

O período de 2001 a 2008 foi marcado como a fase evolutiva do BD, no qual foram desenvolvidos softwares para atender às necessidades de lidar com o crescente número de dados gerados. Após 2009, a inovação do BDA se tornou uma revolução no ambiente empresarial e econômico (Wang; Kung; Byrd, 2018). O gráfico 2 oferece uma visão panorâmica da representatividade do uso do BD por setores econômicos, evidenciando aqueles que mais se desenvolveram nesses anos.

Gráfico 2: Representatividade do *big data* por setores econômicos



Fonte: adaptado de (Zafar, 2019)

2.1.1 6vs *big data*

O BD é caracterizado por 6 variáveis descritivas a seguir:

a) Volume:

O volume é a propriedade que determina o tamanho dos conjuntos de dados, normalmente relatada em Terabyte ou Petabyte. Por exemplo, redes sociais como o Facebook armazenam cerca de 250 milhões de fotografias e mais de 2,5 trilhões de mensagens dos seus utilizadores. Trata-se de uma quantidade grande de dados que precisa de ser armazenada e processada. O volume é a característica mais importante do BD (Berisha; Meziu; Shabani, 2022)

b) Variedade:

Os dados são gerados em diversos formatos por meio de redes sociais, smartphones ou dispositivos. Essas ferramentas geram dados em forma de registros, imagens, vídeos, áudio, documentos ou texto (Sandhu, 2022). Essa ampla gama de fontes contribui para a geração de dados em formatos multidimensionais, incluindo dados estruturados, não estruturados ou semi-estruturados (Acharya et al., 2018).

c) Velocidade:

A velocidade diz respeito à rapidez com que os dados são gerados ou à velocidade necessária para processá-los e analisá-los. Por exemplo, os usuários do Facebook carregam mais de 900 milhões de fotos por dia, o que equivale a aproximadamente 10400 fotos carregadas por segundo. Portanto, as redes sociais devem processar, armazenar e recuperar essas informações em tempo real para seus usuários (Berisha; Meziu; Shabani, 2022).

Para Sharda et al. (2019), essa propriedade representa um desafio para as organizações:

A velocidade talvez seja a característica mais desprezada do *Big Data*. Reagir com agilidade suficiente para lidar com a velocidade é um desafio para a maioria das organizações. Em ambientes que correm contra o tempo, o relógio do custo de oportunidade dos dados dispara no momento em que os dados são criados (Sharda et al., 2019, p. 442).

d) Veracidade:

A veracidade estabelece a qualidade e confiabilidade dos dados (Sandhu, 2022) e está relacionada ao grau necessário de confiabilidade que as fontes de BD devem ter (Wamba; Mishra, 2017). De acordo com Cappa et al. (2020), altos índices de veracidade no BD proporcionam vantagens adicionais ao desempenho das empresas, já que decisões baseadas em dados confiáveis garantem uma maior certeza de retorno das ações.

e) Variabilidade:

Além da rapidez e a diversidade dos dados, os fluxos de informações podem ser notavelmente irregulares, com períodos de intensidade intermitentes. Fatores sazonais ou eventos específicos podem desencadear picos variáveis nas fontes de dados (Sharda et al., 2019). Esses outliers muitas vezes são encontrados através de análises de detecção de anomalias ou pela visualização gráfica de uma análise estatística. Rialti et al. (2018) resume a variabilidade como variações na taxa de fluxo de dados ou processamento.

f) Valor:

O valor dos dados está associado à capacidade das organizações de gerar valor social ou econômico. O grau de valor depende da qualidade dos processos que envolvem as tecnologias do *BD* (Berisha; Meziu; Shabani, 2022). Esse ponto está relacionado a como os dados podem ser tratados e transformados em informações relevantes (Sandhu, 2022). Recentemente, as organizações têm investido cada vez mais em tecnologias de *BD* sob o pretexto de adquirir melhorias de resultados. Nesse contexto, espera-se que os projetos de *BD* garantam o retorno desse investimento (Gao; Sarwar, 2022).

2.1.2 Tecnologia

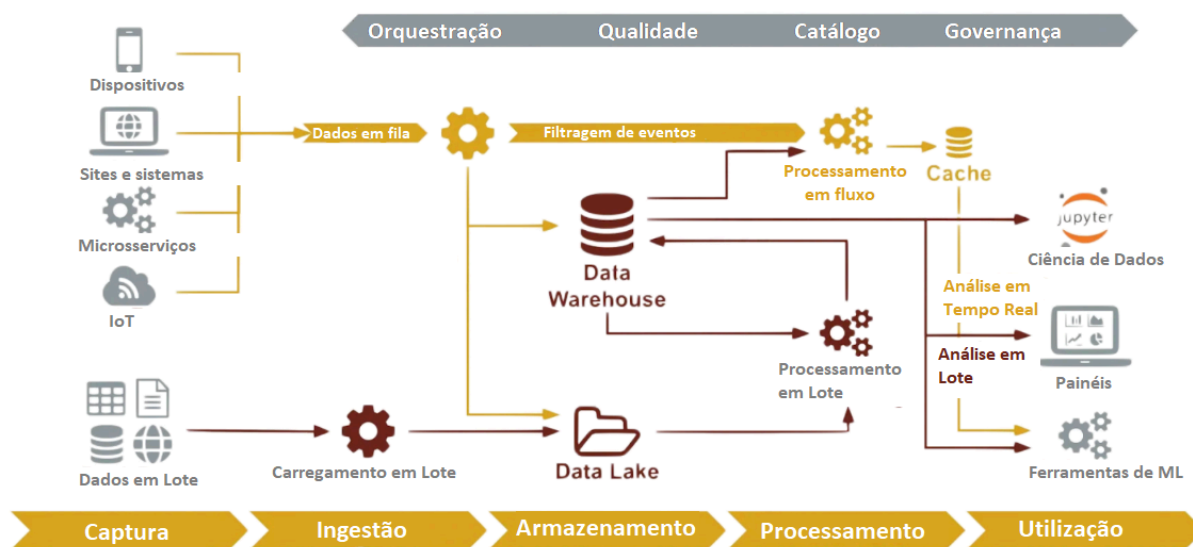
As tecnologias do *BD* são classificadas em diferentes níveis, incluindo: coleta, processamento, armazenamento, segurança e análise. Muitas vezes, a quantidade de informações e fontes distribuídas ultrapassam a nossa capacidade de aproveitamento. Embora o tamanho do *BD* continue aumentando exponencialmente, a capacidade tecnológica atual de lidar com esses conjuntos de dados está nos níveis mais baixos de petabytes, exabytes e zettabytes (Oussous *et al.*, 2017).

2.1.2.1 Arquitetura do *big data*

O processamento do *BD* difere do processamento de dados tradicionais. Em ambientes tradicionais, a abordagem envolve inicialmente a exploração dos dados, seguida pela concepção do modelo de processamento e pela criação da estrutura do banco de dados que vai armazenar os mesmos. A Figura 1 ilustra o fluxo de arquitetura do *BD* para dados em lote e dados em tempo real (distribuído em filas ou micro-lotes). Como demonstrado, ele se inicia com a coleta de dados provenientes de fontes internas ou externas, os quais são então armazenados na chamada "zona de destino", que é uma infraestrutura capaz de lidar com o volume e a variedade dos dados, frequentemente implementada por meio de um sistema de arquivos distribuídos. Após o armazenamento dos dados, as transformações são aplicadas para garantir sua eficiência e escalabilidade. Posteriormente, os dados são

empregados em diferentes contextos analíticos ou são enviados para um banco de dados relacional (Berisha; Meziu; Shabani, 2022).

Figura 1: Arquitetura do *Big Data*



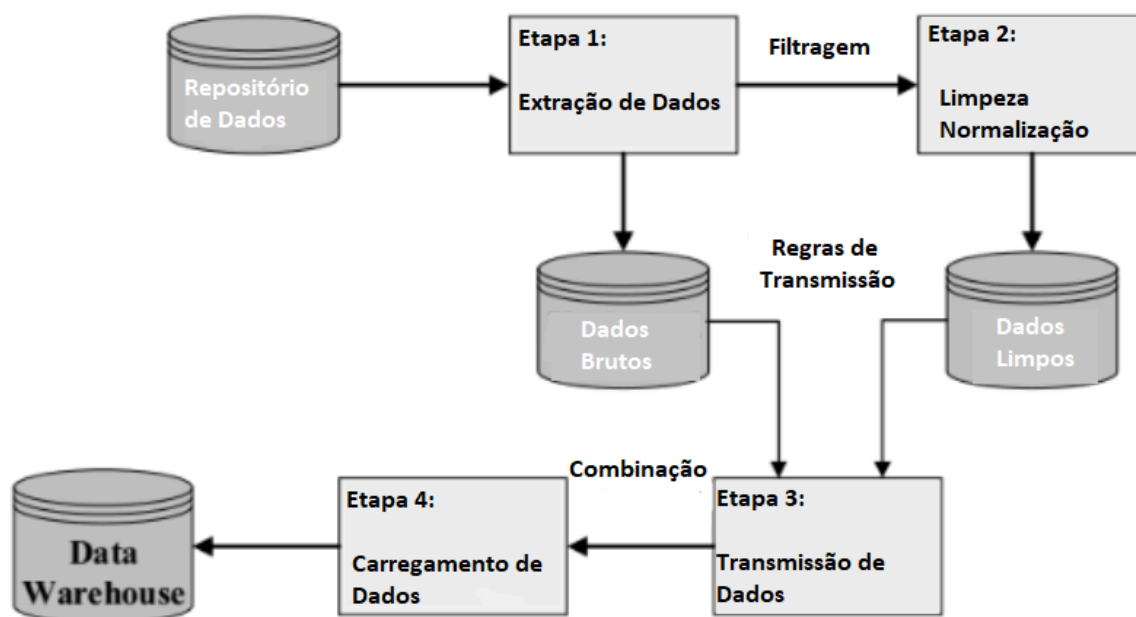
Fonte: Adaptado de Gupta (2020)

2.1.2.2 Coleta de dados

Plataformas como dispositivos IoTs e sistemas gerenciais assimilam dados em tempo real (Sandhu, 2022). Essas fontes são organizadas na arquitetura para o próximo estágio do processo, que envolve a ingestão dos dados relevantes. Independentemente do formato, os dados podem ser integrados assim que estiverem disponíveis. As formas de BD compreendem dados em lote ou dados de fluxo dinâmico. Os dados em lote são armazenados de forma estática, enquanto os dados de fluxo são uma sequência contínua de filas de dados em tempo real. Os dados dinâmicos não são completamente armazenados, com muitos elementos sendo descartados após o processamento (Wang et al., 2020).

A fase de coleta e processamento fundamentam-se nos pilares do ETL (Extração, Transformação e Carregamento) representado na figura 2. Segundo Berisha, Meziu e Shabani (2022), o ETL é um conjunto de procedimentos essenciais para unir dados de diversas fontes em um único repositório central, conhecido como data warehouse.

Figura 2: Fluxo de ETL



Fonte: adaptado de Wang et al. (2020)

O processo de ETL consiste em extrair os dados das fontes e carregá-los em destinos de armazenamento. Os dados passam por transformações, incluindo a limpeza e sua normalização, de acordo com o critério do esquema de destino. Por fim, os dados são carregados em um banco de dados operacional, também chamado de *data warehouse* (Bansal; Kagemann, 2015). No fluxo dinâmico, uma ferramenta de coleta que possa garantir a instantaneidade e estabilidade do fluxo é essencial (Wang et al., 2020).

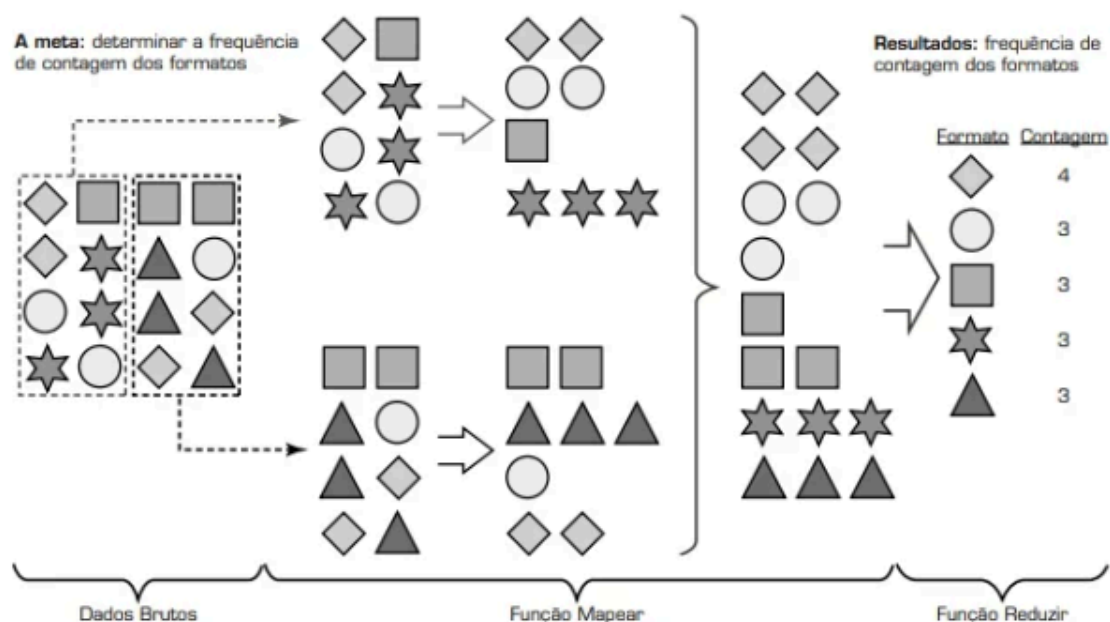
2.1.2.3 Processamento

A etapa de armazenamento dos dados demanda procedimentos de pré-processamento e limpeza, que incluem a combinação e filtragem das informações para otimizar o envio dos dados e garantir a sua consistência. Essas etapas são desafiadoras devido à diversidade de fontes de dados presentes nesse processo. Adicionalmente, as fontes de dados podem conter erros ou estarem incompletas, o que intensifica o desafio de limpar os conjuntos de dados e garantir sua confiabilidade (Sandhu, 2022).

Devido à natureza estática e ao imenso volume dos dados em lote, o processamento de dados é comumente conduzido através de métodos de computação off-line distribuída. Em 2004, o Google introduziu o MapReduce, um modelo de programação distribuída amplamente utilizado para o processamento de grandes conjuntos de dados. O MapReduce possibilita aos usuários elaborar algoritmos complexos em conjuntos de dados extensos sem se preocupar com sincronização, tolerância a falhas ou disponibilidade (Wang et al., 2020). Um programa baseado em computação distribuída é um sistema de armazenamento que distribui os dados entre vários nós de uma rede, garantindo redundância e escalabilidade, permitindo o processamento eficiente de grandes volumes de dados. Tais programas são executados em paralelo por um grande cluster de máquinas (Sharda et al., 2019).

No geral, o MapReduce pode ser descrito de acordo com a figura 3: A função Map divide os dados de entrada em partições independentes, formando pares chave-valor. Esses pares são enviados para o Mapper, que os processa individualmente em múltiplas tarefas de mapeamento paralelas no cluster. Cada partição de dados é atribuída a um nó de computação exclusivo, gerando pares de chave-valores intermediários. A função Reduce é então utilizada para processar os dados intermediários. Para cada chave única, a função Reduce agrega os valores associados de acordo com um programa predefinido, como filtragem, resumo, classificação, média ou encontrar o máximo. Isso resulta em um ou mais pares de chave-valores. Por fim, são armazenados todos os pares de chave-valores em um arquivo de saída (Oussous et al., 2017).

Figura 3: Funcionamento do MapReduce



Fonte: Sharda et al. (2019)

O Hadoop, por sua vez, é um outro programa de arquivos distribuídos utilizado no processamento de dados, principalmente os não estruturados (Oussous et al., 2017). O sistema, composto pelo Hadoop Distributed File System (HDFS), é capaz de armazenar grandes volumes de dados não estruturados e semiestruturados, uma vez que não exige o formato relacional. Devido a essa característica, também é aplicável a fluxos de dados dinâmicos. No Hadoop, cada segmento de dados é replicado várias vezes e distribuído pelo sistema de arquivos, assegurando que, se um nó falhar, haverá uma cópia dos dados em outro nó. Assim que os dados são abastecidos no cluster, estão prontos para serem processados via MapReduce (Sharda et al., 2019).

2.1.2.4 Armazenamento

De acordo com Wang et al. (2020), nos últimos anos houve uma mudança significativa nas técnicas de armazenamento de dados devido ao crescimento do BD. Entre as opções adotadas, destacam-se os sistemas de arquivos distribuídos, NoSQL, NewSQL e outras linguagens de bancos de dados. Bancos de dados NoSQL podem lidar com dados não estruturados com alta disponibilidade e durabilidade. Um dos principais benefícios dos bancos de dados NoSQL é a

escalabilidade horizontal, a qual permite o escalonamento contínuo de uma única tabela em centenas de servidores de baixo custo (Chauhan; Sood, 2021).

O artigo "Relational vs NoSQL Data", publicado pela Microsoft em 2022, descreve que os bancos de dados relacionais oferecem armazenamento de tabelas que podem ser combinadas. Essas tabelas possuem um esquema fixo, usam Structured Query Language (SQL) para gerenciar dados e oferecem suporte a garantias de Atomicidade, Consistência, Isolamento e Durabilidade (ACID). Os bancos de dados NoSQL normalmente não fornecem garantias ACID e referem-se a armazenamentos de dados não relacionais de alto desempenho. Eles se destacam por suas características de facilidade de uso, escalabilidade e disponibilidade. Em vez de unir tabelas de dados normalizados, o NoSQL armazena dados não estruturados ou semiestruturados, geralmente em pares de valores-chave ou documentos JSON (Microsoft, 2022).

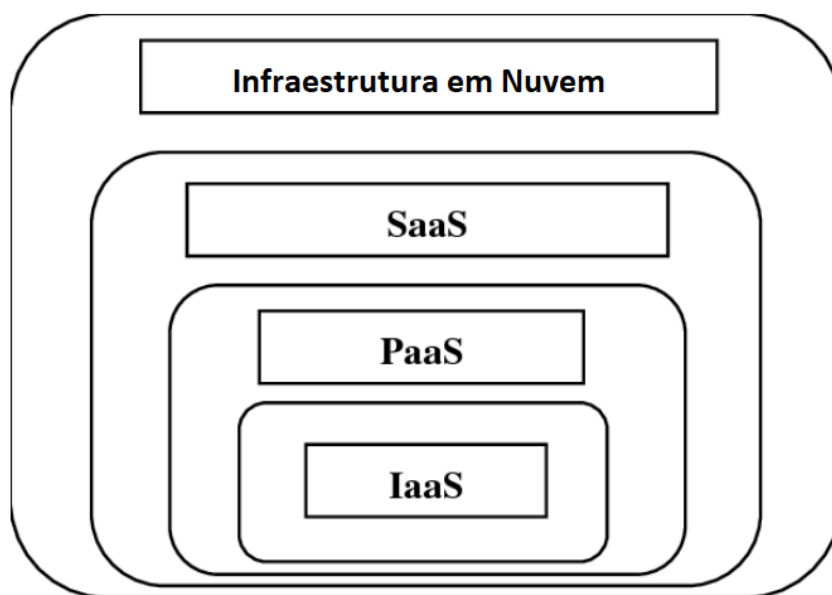
Já o NewSQL pode ser considerado uma tecnologia emergente de banco de dados relacional com escalabilidade distribuída do NoSQL, serviço SQL e garantia de transações ACID (Wang et al., 2020). Essas tecnologias podem ser aplicadas a sistemas baseados na computação em nuvem.

2.1.2.5 Computação em Nuvem

A evolução do BD se intensificou ainda mais nos últimos anos com o advento da computação em nuvem (Dwivedi et al., 2021). A computação em nuvem facilitou o armazenamento e o processamento de dados, fornecendo armazenamento quase ilimitado e facilidade de acesso a servidores de maior performance (Berisha; Meziu; Shabani, 2022). Diante das vantagens oferecidas pela computação em nuvem como a elasticidade, modelo de pagamento sob demanda e baixo investimento inicial, tornou-se uma opção viável para empresas armazenarem seus dados, principalmente diante da imprevisibilidade do volume de dados no BD (Berisha; Meziu; Shabani, 2022).

Os serviços de computação em nuvem podem ser classificados em três categorias: IaaS, SaaS e PaaS (Sandhu, 2022), exibidos em escala na figura 4.

Figura 4: Infraestrutura em nuvem



Fonte: adaptado de Shah et al. (2010)

1) Infraestrutura como serviço (IaaS): Esses serviços seguem o princípio de "pagar pelo que você precisa", fornecendo computação de alto desempenho aos clientes. Exemplos incluem a Amazon Web Services (AWS) e o Simple Storage Service (S3). A AWS e o S3 oferecem armazenamento online, permitindo acesso aos maiores *data centers* do mundo a preços acessíveis (Sandhu, 2022).

2) Software como Serviço (SaaS) permite aos usuários acessar estruturas de processamento de dados para diferentes requisitos por meio de serviços de computação em nuvem (Wang et al., 2020). Nele, os usuários não precisam manter softwares, apenas entregam os dados a serem processados para os serviços SaaS específicos e pagam conforme a conclusão da tarefa, sob demanda. Os aplicativos são executados em infraestrutura de nuvem remota, oferecendo aos usuários um ambiente de desktop virtual onde todos os programas estão instalados (Sandhu, 2022).

3) Plataforma como Serviço (PaaS): Oferece uma plataforma escalável na nuvem para o processamento de BD, permitindo aos usuários criar, testar e executar aplicativos da Web. Exemplos incluem Microsoft Azure e o Google Cloud (Sandhu, 2022).

2.1.3 *Big data analytics*

Segundo Dwivedi et al. (2021), o *big data analytics* é uma metodologia criada para analisar os dados gerados no BD. O BDA também simplifica a coleta de dados operacionais e econômicos à medida que o escopo dos dados mais utilizados nas análises é gradualmente esclarecido (Mirarab; Mirtaheri; Asghari, 2019). De acordo com Mirarab, Mirtaheri e Asghari (2019), o BDA é utilizado em diversas áreas dentro das organizações, dentre elas: financeira, performance e vendas, gerenciamento de risco e detecção de fraude.

Com o crescimento do BDA, “a conversão de uma enorme quantidade de dados diversos em informações úteis tem gerado benefícios anteriormente inatingíveis” (Gao; Sarwar, 2022, p.6). Nesse contexto, o *big data analytics* surgiu como um dos fatores mais importantes na geração de *insights* para a tomada de decisões (Awan et al., 2021).

Nesse sentido, existem 4 tipos de análise de dados:

1. Análise Descritiva: “a análise de dados descritiva diz respeito a conhecer o que está acontecendo na organização e entender tendências e causas subjacentes de tais ocorrências” (Sharda et al., 2019, p 25).
2. Análise Diagnóstica: busca determinar a causa de ocorrência para um fenômeno ou evento (Cote, 2021) <https://online.hbs.edu/blog/post/diagnostic-analytics>
3. Análise Preditiva: “visa determinar o que é mais provável de acontecer no futuro com base na análise histórica e de tendência” (Sharda et al., 2019).
4. Análise Prescritiva: define os melhores caminhos a seguir a partir das previsões realizadas. O objetivo desse tipo de análise é chegar a uma decisão para otimização de algum desempenho (Sharda et al., 2019).

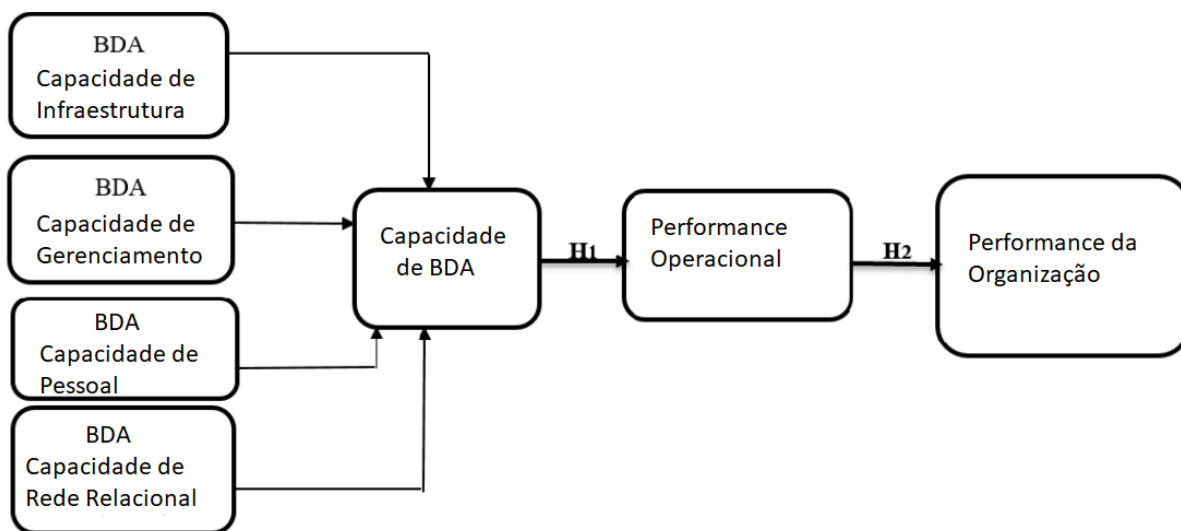
No contexto do BDA é muito aparente a necessidade de prever comportamentos e tomar decisões. Dessa forma, a análise preditiva e prescritiva são as mais utilizadas, embora o diagnóstico e a descrição também sejam muito importantes (Chauhan; Sood, 2021). No âmbito dos negócios, as técnicas de predição conseguem prever, por exemplo, se o cliente está propenso a migrar para um concorrente, o que o

cliente tende a comprar, a quais promoções o cliente reagiria, qual o risco de crédito do cliente, e assim por diante (Sharda et al., 2019). A análise prescritiva entra em vigor após o entendimento dessas tendências de forma a auxiliar na elaboração de um plano-ação (Awan et al., 2021).

2.1.3.1 Capacidades de gerenciamento de *big data analytics*

A capacidade de gerenciamento de *big data analytics* (CGBDA) começou a receber atenção na literatura mediante a necessidade de entendimento sobre como as organizações devem promover o desenvolvimento do BDA internamente de forma a usar adequadamente os seus recursos (Rialti et al., 2019). Para muitos autores, a CGBDA aumenta a capacidade das empresas de transformar os processos existentes e gerar valor comercial. Há uma relação positiva entre o desenvolvimento da CGBDA e os ganhos de performance (Gao; Sarwar, 2022). Essa influência é representada no fluxo da CGBDA, de acordo com a figura 5.

Figura 5: Fluxo do CGBDA



Fonte: adaptado de Garmaki, Boughzala e Wamba (2016)

A CGBDA fundamenta-se em três frentes estratégicas: infraestrutura, gerenciamento e expertise da equipe de BDA (Awan et al., 2021). De acordo com Akter et al. (2016), a infraestrutura é o domínio físico, composto pela estrutura tecnológica da empresa para realizar a análise de BD. O gerenciamento compõe o nível organizacional do

BDA na companhia e está associado a questões de planejamento, investimento, coordenação e segurança. A falta de capacidade de gerenciamento é uma das principais barreiras à operacionalização do BDA (Gao; Sarwar, 2022). No gerenciamento, destaca-se a importância da abordagem de *Business Intelligence e Analytics* (BI&A) como a frente de transformação do BDA em uma abordagem estratégica baseada em dados (*data-driven*) (Awan et al., 2021).

O pilar da expertise refere-se ao nível de conhecimento da equipe para desenvolver processos de BDA. A capacidade de pessoal do BDA serve como catalisador para mobilizar a gerência no entendimento das diferentes regras de negócios de acordo com os objetivos da organização (Awan et al., 2021). Nesse sentido, Mirarab, Mirtaheri e Asghari (2019) enfatizam a importância de formar uma equipe de BDA com profissionais capacitados, com domínio técnico e observância teórica, capazes de conduzir corretamente um projeto de BDA. Entre as falhas comuns da operação voltada ao BDA situam-se os dados analisados de forma errônea, sem critérios de qualidade. Uma análise conduzida por Rialti et al. (2019) aponta que os efeitos positivos do BDA só existem ao se trabalhar com análises confiáveis:

Na verdade, o uso de análises de baixa qualidade pode gerar previsões errôneas que levam a decisões erradas. Apesar disso, os riscos relacionados à análises de baixa qualidade não estão relacionados apenas à qualidade dos dados analisados pelos sistemas de BDA. (...) Foi observado que as pessoas também podem se opor a usar uma tecnologia corretamente quando não são capazes de compreender os benefícios de sua aplicação (Rialti et al., 2019, p. 9).

Nesse sentido, é imperativo a realização de programas de treinamento para desenvolver o time dedicado ao BDA em toda a organização. Esses treinamentos podem ser ministrados com a ajuda de consultores especializados (Rialti et al., 2019).

2.2 TOMADA DE DECISÃO

Mintzberg, Raisinghani e Théorét (1976) definem a decisão como um desenvolvimento do raciocínio com foco na escolha entre diferentes cursos de ação, ou seja, uma escolha entre alternativas. A definição de processo decisório é apresentada pelos autores como um conjunto de ações e fatores dinâmicos que começam com a identificação de um estímulo para a ação e terminam com o compromisso específico para a ação. Na perspectiva de Nutt (1984), o processo decisório é constituído por um conjunto de atividades que começa com a identificação de um problema e termina com a ação. Essencialmente, para haver uma decisão nesse caso é preciso existir um problema.

O conceito da tomada de decisão foi explorado inicialmente pelo renomado economista alemão Herbert Simon a partir do ano de 1947. A teoria da decisão de Simon oferece uma visão sobre como os seres humanos encaram o desafio de tomar decisões em um mundo complexo e cheio de incertezas. Essa teoria introduziu o conceito de "racionalidade limitada", o qual reconhece as restrições inerentes às capacidades cognitivas e ao ambiente em constante mudança que influenciam as escolhas. Em alguns casos, diante da vasta quantidade de alternativas disponíveis, os tomadores de decisão não podem analisar exaustivamente todas as opções. Em vez disso, recorrem a estratégias simplificadas, conhecidas como "heurísticas", para encontrar soluções satisfatórias (Shimizu, 2001). No presente estudo, o processo da tomada de decisão será tratado sob o enfoque da racionalidade dentro da perspectiva teórica, com foco no ambiente corporativo.

2.2.1 Processo decisório

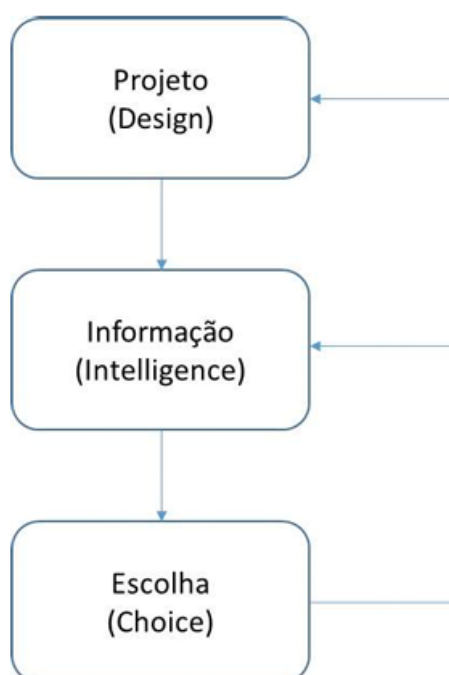
De acordo com Shimizu (2001), é comum que as organizações se deparem com desafios na tomada de decisão. Esses desafios se caracterizam por sua amplitude e complexidade, envolvendo riscos e incertezas. Além disso, os empecilhos aparecem em diversos níveis funcionais dentro de uma empresa. As decisões podem ser classificadas em programadas e não programadas. As decisões programadas são repetitivas e rotineiras, as quais podem ser automatizadas nos processos de

negócios. Decisões não programadas exigem que o decisor empregue julgamento, intuição e criatividade (Simon, 1977).

A “boa decisão” é resultado de um processo racional que pode ser replicado para obter os mesmos resultados em condições semelhantes. Não se avalia uma decisão apenas pelo resultado final da mesma, mas sim pela qualidade do processo que a fundamenta. Os decisores precisam fazer uma escolha de forma imparcial, evitando influências de ideias preconcebidas. Eles são responsáveis por envolver os agentes relevantes no processo decisório com base em fatos e dados. A escolha de uma ação resulta da avaliação das alternativas, guiada por critérios definidos previamente. Para isso, é essencial estabelecer uma estrutura analítica que oriente a tomada de decisão com base na escolha do melhor curso de ação, considerando as principais variáveis envolvidas no processo (Nascimento *et al.*, 2011). O decisor é aquele que, entre os atores, está munido de poder institucional para ratificar uma decisão (Mintzberg; Raisinghani; Theorêt, 1976).

Para minimizar erros relacionados às perspectivas individuais dos decisores e atingir melhores resultados, Simon (1977) introduziu o modelo de processo sistemático do processo decisório, demonstrado pelo fluxo a seguir:

Figura 6: Modelo de processo para tomada de decisão de Simon



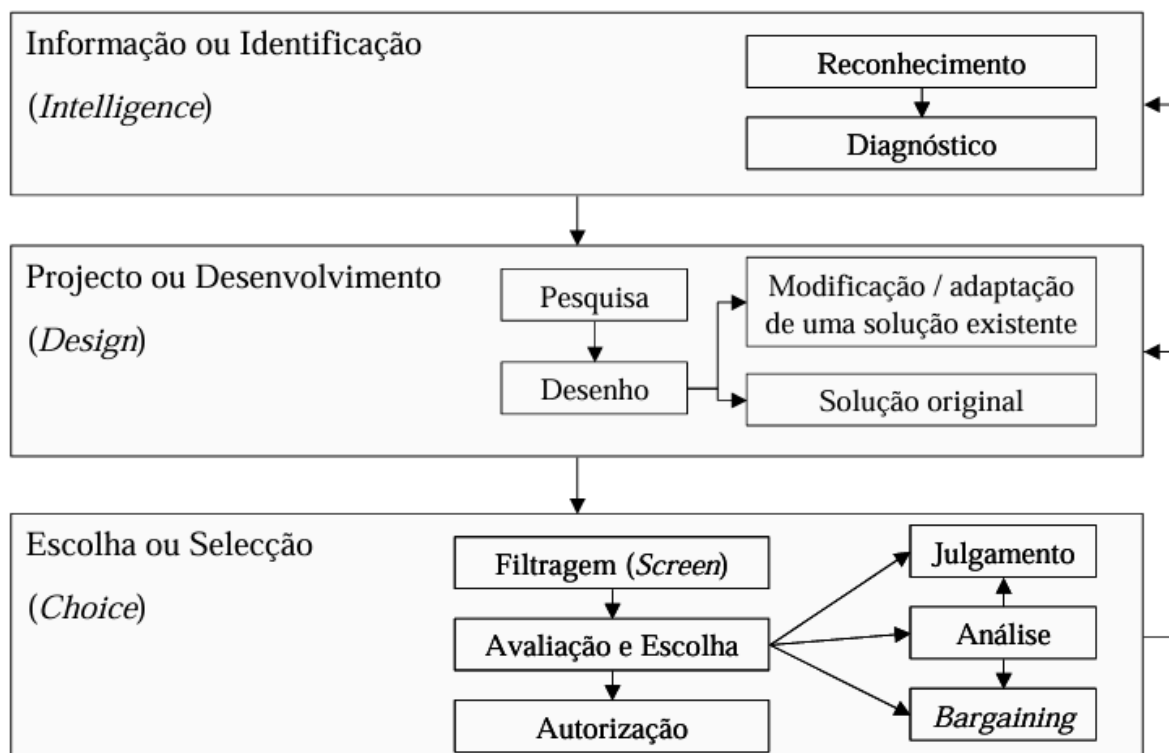
Fonte: Simon (1977)

Herbert Simon (1977) propõe uma estrutura tridimensional para o processo decisório, que se desdobra ao longo do tempo, mostrando como a decisão é um resultado desencadeado de várias etapas. Essas fases podem ser divididas em: *Design* (projeto), Inteligência (informação) e Escolha.

Na fase inicial de Projeto, o foco é identificar a decisão a ser tomada e as diversas opções disponíveis. Por exemplo, decidir se a empresa deve criar uma nova campanha promocional em determinado período no ano para alavancar suas vendas. Na etapa seguinte, Inteligência, o objetivo é coletar informações relevantes para investigar essa ideia. Idealmente, essas informações auxiliam na melhoria da qualidade da decisão. Simon observa que a fase de Inteligência pode impactar a fase anterior de Design, pois a coleta de informações pode introduzir novas alternativas ou mesmo modificar a decisão original. Na terceira fase, Escolha, a decisão é tomada ao selecionar pelo menos uma das alternativas delineadas na fase de Design. É importante ressaltar que não tomar uma decisão também é considerado uma decisão (De Souza; Alturas, 2019).

Mintzberg, Raisinghani e Theorêt (1976) realizaram um estudo de campo que consistiu na observação de vinte e cinco processos de tomada de decisão “não estruturada”. Este estudo deu origem à organização de um processo de tomada de decisão estratégica que, tal como o de Simon, é constituído pelas mesmas três etapas. Uma decisão estruturada acontece quando existe um certo grau de clareza das possíveis consequências de uma ação (Simon, 1965). O modelo desestruturado de tomada de decisão, formulado por Mintzberg, descreve o processo decisório em situações de alta incerteza. Este modelo reconhece a natureza incremental da tomada de decisão, caracterizada por avanços graduais que eventualmente convergem para uma decisão. Nas organizações, quando surgem obstáculos, é comum reconsiderar alternativas e até retroceder, tornando o processo não linear e imprevisível. Essa abordagem enfatiza a natureza desestruturada das decisões, provocadas devido às constantes mudanças do ambiente. Um estudo elaborado por Matos (2006) descreve a composição do modelo de Mintzberg, Raisinghani e Theorêt (1976), indicado por cada etapa na figura 7:

Figura 7: Modelo de processo para tomada de decisão de Mintzberg



Fonte: Mintzberg, Raisinghani e Theorêt (1976 apud Matos, 2006)

O modelo de processo para tomada de decisão de Mintzberg, Raisinghani e Theorêt (1976) descreve as seguintes etapas:

1. Identificação: Esta fase envolve a identificação de problemas a serem resolvidos ou oportunidades em uma determinada situação:
 - 1.1. Reconhecimento: identificação de oportunidades, problemas e crises que exigem decisões.
 - 1.2. Diagnóstico: compreensão das variáveis envolvidas no caso e análise das consequências com uma possível decisão.
2. Desenvolvimento: Nesta fase, são criadas alternativas de solução para o problema. Esta fase inclui duas etapas:
 - 2.1. Pesquisa: busca por soluções possíveis com base em informações armazenadas pela empresa ou em bancos de dados.

2.2. Desenho: adaptação de soluções encontradas na pesquisa ou criação de soluções originais quando as encontradas na pesquisa não são viáveis.

3. Seleção: Nesta fase, um curso de ação é escolhido entre as alternativas desenvolvidas na fase de Desenvolvimento. Esta fase consiste em três etapas:

3.1. Filtragem: eliminação de alternativas impraticáveis quando a pesquisa gera mais opções do que podem ser avaliadas.

3.2. Avaliação e Escolha: as soluções são avaliadas para que seja selecionada a mais adequada. Pode-se utilizar três modos de avaliação e escolha:

3.2.1. Julgamento: decisão tomada individualmente, usando procedimentos que às vezes não podem ser explicados.

3.2.2. Barganha: escolha feita por um grupo de indivíduos, cada um com suas próprias metas e julgamentos.

3.2.3. Análise: avaliação das alternativas por um técnico ou especialista, baseado em métodos matemáticos e dados. Normalmente resulta em um relatório comparativo das alternativas, que é fornecido aos responsáveis pela decisão para escolher um curso de ação.

3.3. Autorização: quando o indivíduo ou grupo de decisão não tem autorização para decidir um curso de ação, a decisão deve seguir um processo de aprovação hierárquica dentro da organização.

Certo (2005) também sugere a avaliação da pós-decisão. Ao implementar uma ação decorrida do processo sistemático, é preciso reunir feedbacks para descobrir se a alternativa implementada está sendo eficaz para o problema identificado. Se não estiver funcionando adequadamente, o processo deve ser reiniciado.

Ao entender os possíveis caminhos de avaliação e escolha do processo decisório proposto por Mintzberg, Raisinghani e Theorêt (1976), o presente estudo procura

estabelecer o critério de avaliação da tomada de decisão sob a ótica analítica (explicado no tópico 3.2.3), devido a sua natureza racional e não intuitiva.

A diferença entre racionalidade e intuição está na proporção de informação, de um lado, e opinião e sentimentos, de outro. Quanto maior a base de informação, mais racional é o processo. Quanto maior a proporção de opiniões e sentimentos, mais intuitivo se torna (Maximiano, 2009, p.71).

A melhor estratégia para a avaliação de uma decisão é o seu julgamento com base em dados. O processo analítico deve ser pautado em indicadores mensuráveis, na identificação de padrões e na busca por insights potenciais, visando desenvolver decisões estratégicas que maximizem o retorno de valor, principalmente no contexto do BD (Mirarab; Mirtaheri; Asghari, 2019). Dessa forma, a perspectiva do processo decisório baseado em dados é analisada no tópico seguinte.

2.2.1.1 Decisão baseada em dados

“Baseado em dados” (DD - *data-driven*) é a ênfase dada ao uso de dados como parte principal do processo decisório de uma organização (PUCPR, 2024). O *data-driven* surgiu com o advento da Indústria 4.0, a qual refere-se à transformação digital de diversos setores dos mercados industriais. Ela representa a chamada quarta revolução industrial na manufatura. A indústria 4.0 visa aumentar a eficiência na produção industrial a partir da conexão inteligente de processos e sistemas digitais em áreas como a logística e a cadeia de suprimentos (Misra et al., 2022). O data driven emerge nesse cenário mediante a alta combinação de dados provenientes das tecnologias IoT, inteligência artificial e o próprio *big data* (Dwivedi et al., 2021).

Diante da compreensão dos benefícios gerados pelo DD, algumas organizações adotaram a metodologia como uma base para a inteligência de negócio (BI). Isso porque as informações provenientes de experimentações com dados são valorizadas acima de opiniões. Dessa forma, assegura-se decisões holísticas a partir da análise do passado, da compreensão do presente e da predição do futuro (PUCPR, 2024). Um exemplo proeminente na área pode ser observado com o crescimento da empresa Netflix, um serviço de streaming que adotou estratégias baseadas em dados e obteve resultados significativos com essa abordagem. O

diretório de pesquisa da empresa revela a utilização do DD como um pilar estratégico:

A Netflix tem sido uma empresa focada em dados desde o seu início. A análise na Netflix aproveita um conjunto diversificado de habilidades (enquadramento de problemas, engenharia de dados, ciência de dados, pesquisa de consumo, engenharia de visualização e muito mais) para conectar pontos entre domínios e responder a desafios de negócios complexos com soluções analíticas inovadoras. Acreditamos que os insights derivados de dados e pesquisas rigorosas nos capacitarão a tomar decisões inteligentes sobre nossos negócios e a experiência dos membros (Netflix, 2024).

Uma pesquisa publicada no site da Investing em 2024 mostra que desde 2009 a Netflix registrou um crescimento médio anual de receita e valor de mercado superior a 24% (Manzoni, 2024). Outras empresas como Amazon, Spotify e Airbnb também são exemplos notáveis de como estratégias baseadas em dados têm impulsionado o sucesso nos negócios. A Amazon utiliza dados para oferecer recomendações personalizadas, elevando a experiência do cliente. O Spotify usa algoritmos baseados em dados para criar playlists personalizadas, melhorando o engajamento dos usuários com a plataforma. O Airbnb ajusta dinamicamente os preços do aplicativo com base em dados de mercado. Tais aplicações demonstram o uso inteligente dos dados na transformação da experiência do consumidor, estabelecendo novos padrões de competitividade e inovação (InfinitScale, 2023).

Nesse sentido, a Tomada de Decisão Baseada em Dados (TDBD) se caracteriza pelo uso de dados para orientar decisões de negócios. A TDBD se inicia a partir da definição de um objetivo, o que envolve compreender a meta a ser buscada pela organização com uma decisão. Esse método orienta a escolha de indicadores-chave de desempenho (KPIs) e as métricas que serão analisadas para apoiar esses objetivos (Tableau, 2024).

O entendimento dos objetivos de cada projeto inicia-se a partir de uma análise preliminar diagnóstica dos principais fatores envolvidos no processo. Nesse aspecto, podem ser utilizadas ferramentas presentes na literatura da gestão da qualidade para investigar as causas e as variáveis imperativas de um problema. O diagrama de Ishikawa e o gráfico de Pareto podem contribuir para esses objetivos (Maximiano, 2009). Estabelecida a compreensão dos objetivos que guiarão o processo analítico até a tomada de decisão, inicia-se o processo de concepção da análise dos dados

(Mirarab; Mirtaheri; Asghari, 2019). Nesse cenário, o analista busca entender em quais fontes pode-se obter os dados para a criação da análise. O resultado final pode ser desde uma análise documentada até um *dashboard*.

No contexto do *big data*, o grande desafio da TDBD reside no mapeamento das informações que de fato são relevantes para o negócio. Converter os problemas, metas e objetivos corporativos em dados torna-se complexo devido à vasta gama de possíveis escolhas disponíveis para cada situação (Paradza; Daramola, 2021). É nesse aspecto que reside o ponto de contato entre os dois universos: BD e a tomada de decisão. Esse processo de conversão só se torna viável com um nível adequado de centralização e normalização dos dados dentro das empresas. O primeiro referencial faz destaque para a escala do BD, especialmente em termos de volume, variedade e velocidade dos dados. Como mencionado, as organizações não apenas necessitam de uma infraestrutura tecnológica robusta e capaz de suportar o BD, como também de uma governança de dados eficaz. Com o ambiente estruturado, ainda assim os dados provenientes só terão valor se os seus consumidores estiverem preparados para recebê-los e entendido a importância de pautar suas decisões em dados (Rialti et al., 2018). Quando esse cenário se estabelece, o *data-driven* torna-se cultural dentro de uma organização, e em todo caso o DD será como uma premissa básica dentro do processo decisório (Awan et al., 2021).

2.2.2 Cultura de dados

A cultura de dados representa as convicções compartilhadas por pessoas que valorizam o uso de dados para aprimorar a tomada de decisões. Conseqüentemente, o DD é integrado às operações, à mentalidade e à identidade de uma organização. A cultura de dados capacita todos os membros da organização a buscarem uma abordagem orientada por fatos (Tableau, 2024).

De acordo com o relatório “*Data-Driven Business*” publicado em 2021 pela Idera, um aspecto importante para a obtenção de sucesso na implementação da cultura de dados é a confiabilidade dos dados. Se os colaboradores não tiverem confiança nos dados, é impossível criar uma cultura em que os dados sejam usados ativamente. Todas as análises, desde estatísticas simples até modelos avançados de aprendizado de máquina dependem de informações confiáveis. Além da qualidade,

os dados devem estar centralizados em um repositório principal para que se tenha uma fonte única da verdade e para que possam ser encontrados sem dificuldades. Empresas apoiadas no data-driven democratizaram e tornaram os dados acessíveis em toda a organização (Almquist; Gjengsto; Nordli, 2021).

Uma pesquisa de mercado feita pela Barc e pelo Tableau em 2023 estruturou a cultura de dados em seis áreas de atuação: Liderança de dados, Estratégia de Dados, Governança de Dados, Alfabetização em Dados, Acesso a Dados e Comunicação de Dados (Bange; Grosser, 2023). A estrutura é representada na figura 8.

Figura 8: Estrutura da cultura de dados



Fonte: adaptado de Bange e Grosser (2023).

Liderança de dados: refere-se a indivíduos que coordenam e gerenciam os dados dentro de uma empresa. Essa liderança deve ser responsável pela governança dos dados, estratégias, arquitetura, tecnologias e confiabilidade (Tera, 2020).

Estratégia de Dados: plano de longo prazo que define as tecnologias, os processos, as pessoas e as regras necessárias para potencializar o ganho de valor através dos dados em uma organização. Atualmente, vários tipos de empresas coletam grandes volumes de dados brutos. No entanto, elas precisam de um plano bem elaborado de gerenciamento se quiserem usar essas informações para tomar decisões. Uma estratégia de dados descreve os objetivos da organização em relação aos seus dados (Amazon, 2024).

Governança de Dados: abordagem fundamentada em princípios para gerenciar dados ao longo do seu ciclo de vida, desde a aquisição até o seu descarte. Abrange todas as medidas tomadas para garantir a segurança, privacidade, disponibilidade e qualidade dos dados. Isso envolve tanto as ações realizadas pelas pessoas quanto os processos, além da utilização das tecnologias de forma correta (Google, 2024).

Alfabetização em Dados: capacidade de uma pessoa trabalhar e interpretar dados (Bange; Grosser, 2023). Segundo um estudo feito pela Forrester em 2021, 82% dos gestores esperam que seus funcionários saibam trabalhar com dados da forma correta.

Acesso a Dados: relacionado à democratização dos acessos aos dados. As pessoas devem ter acesso às plataformas que disponibilizam os dados, limitadas pelo nível adequado de segurança do acesso (Bange; Grosser, 2023). Em algumas corporações adota-se o modelo “*Analytics Self-Service*, que é um tipo de BI por meio do qual os indivíduos podem analisar os dados por conta própria sem depender de analistas ou de uma equipe de TI para obter suporte” (Targit, 2024).

Comunicação de dados: A comunicação de dados é a capacidade de trabalhar em equipe de forma conjunta e eficiente com dados, além de comunicar os resultados de maneira correta. Fornece uma base sólida para discussões e decisões (Barc, 2024).

3 MATERIAL E MÉTODO

Este capítulo expõe a metodologia empregada neste estudo. Inicialmente, apresenta-se a caracterização da pesquisa, seguida pela descrição do instrumento de coleta de dados e pela definição do público-alvo e das amostras. Em seguida, detalham-se os procedimentos empregados nas etapas de coleta e análise dos dados.

3.1 CARACTERIZAÇÃO DA PESQUISA

A pesquisa que busca informações diretamente com um grupo de interesse é conhecida como *survey*. Este método foi escolhido pelo autor por ser eficaz em estudos exploratórios (Santos, 1999). O *survey* envolve a coleta de informações sobre características ou opiniões de um grupo específico considerado representativo de uma população-alvo. Normalmente, questionários são os principais instrumentos de investigação (Fonseca, 2002). Nesse sentido, o presente estudo utilizou abordagem quantitativa com caráter descritivo. A escolha deste método fundamenta-se na necessidade de quantificar e analisar de forma detalhada os dados coletados. Isso permite uma compreensão mais objetiva do fenômeno abordado.

O método quantitativo oferece uma visão ampla e generalizável dos resultados, além de possibilitar a identificação de padrões e tendências nos dados coletados. Este método é adequado para testar hipóteses e estabelecer relações causais entre variáveis (Gil, 2010), o que exige do pesquisador um conhecimento aprofundado do tema em estudo, de modo a identificar com precisão os resultados que se pretende obter ou mensurar (Mattar, 2001).

3.2 INSTRUMENTO DE COLETA DE DADOS

O instrumento de coleta de dados refere-se ao meio pelo qual as questões são apresentadas aos participantes da pesquisa e através do qual obtêm-se os resultados (Mattar, 2001). No presente estudo, foi aplicado um questionário, por ser uma ferramenta de pesquisa amplamente aceita e utilizada em levantamentos acadêmicos. De acordo com McDaniel Jr. e Gates (2005), além de possibilitar uma

ampla gama de participantes, o questionário é uma ferramenta eficaz para identificar quem são os respondentes, por que eles adotam certos comportamentos e como tomam suas decisões. Dessa forma, a pesquisa em questão busca obter uma visão abrangente dos participantes sobre aspectos voltados à gestão de dados e como as empresas alinham a realidade de crescimento do BD e suas tecnologias com a tendência de geração de valor.

De acordo com Gil (2010), o processo de elaboração do questionário consiste em traduzir os objetivos específicos da pesquisa em questões cuidadosamente formuladas. Esta transformação visa garantir que cada pergunta contribua diretamente para a obtenção das informações necessárias. A pesquisa por meio de um questionário estruturado permite que as informações sejam claramente definidas e organizadas (Malhotra, 2012). Para Gunter (2003), as perguntas devem ser posicionadas obedecendo uma sequência lógica. A primeira pergunta possui o objetivo de deixar o respondente à vontade para prosseguir no questionário. Para isso, indica-se iniciar a pesquisa com uma pergunta mais aberta e direta sobre o objeto de estudo, de forma que o respondente possa conhecer o tema principal ainda no seu início Mattar (1994). O modo da aplicação do questionário se deu por meio online, o que possibilitou a liberação do acesso para empresas selecionadas por meio das diversas redes de comunicação disponíveis na internet. Entre elas, destaca-se o compartilhamento da pesquisa através das plataformas whatsapp e linkedin. O autor optou por utilizar a plataforma de criação de formulários Google Forms da Google, devido à confiabilidade do serviço e salvamento automático em nuvem dos resultados em formatos estruturados.

O formulário (Apêndice A) foi elaborado com referência ao survey "BARC Data Culture Survey 23: How to Liberalize Data Access to Empower Data Users", desenvolvido pela BARC (Business Application Research Center) em parceria com a Tableau (Bange; Grosser, 2023). As perguntas e respostas foram adaptadas para o problema de pesquisa deste estudo. Nesse sentido, a utilização de um instrumento já validado e reconhecido no campo de pesquisa confere maior robustez à metodologia aplicada (Gil, 2010). A adaptação do survey original permitiu preservar a integridade das questões essenciais, garantindo que as informações coletadas fossem consistentes com estudos anteriores reconhecidos na área. Além disso,

foram adicionadas novas perguntas cuidadosamente elaboradas para abordar as especificidades deste estudo, de forma a trazer uma análise mais direcionada ao contexto investigado. Cada pergunta foi analisada individualmente na seção de resultados.

Para as possíveis respostas, foram elaboradas diferentes escalas de mensuração das variáveis de estudo para melhor atender ao escopo da pesquisa. Nesse contexto, foram utilizadas escalas de Likert de quatro ou cinco pontos com diferentes ancoragens, como a de concordância, variando de "1 - discordo totalmente" a "5 - concordo totalmente", e a de intensidade, variando de "1 - muito baixo" a "5 - muito alto" (Mcdaniel Jr.; Gates, 2005). Além disso, foram aplicadas escalas personalizadas que melhor se adequam ao tipo de resposta esperada. Nesses casos, encontram-se também seleções qualitativas. Algumas dessas escalas são de múltipla escolha, permitindo que os participantes selecionem mais de uma opção quando há a possibilidade de múltiplos fatores estarem associados a uma determinada ocorrência.

3.3 DEFINIÇÃO DA POPULAÇÃO E AMOSTRAS

A população desta pesquisa é composta por empresas de médio a grande porte que lidam com o BD e utilizam dados em suas operações diárias. A pesquisa se concentra em empresas que possuem uma infraestrutura mínima para se trabalhar com dados. Essas empresas, no geral, pertencem a setores variados como tecnologia, finanças, saúde, varejo, manufatura e telecomunicações (Zafar, 2019).

A seleção da amostra de empresas foi realizada por meio de amostragem intencional, também conhecida como amostragem por julgamento. Esse tipo de amostragem é uma técnica não probabilística, na qual "a seleção dos elementos da população para compor a amostra depende ao menos em parte do julgamento do pesquisador ou do entrevistador no campo" (Mattar, 1996, p. 132). De acordo com Aaker et al. (1995), o procedimento de amostragem por julgamento pode ser útil quando é necessário incluir um pequeno número de unidades na amostra. Diante desse cenário, o autor optou pelo método de amostragem por julgamento devido à limitação de tempo disponível para a execução de um método pautado em amostragem probabilística. A aplicação da amostragem aleatória exigiria um

tamanho amostral elevado, o que se mostrou inviável considerando o cronograma estabelecido para o desenvolvimento deste trabalho. A amostragem por julgamento foi considerada a abordagem mais eficaz para alcançar os objetivos da pesquisa dentro do tempo estipulado, uma vez que a fase de coleta dos dados durou aproximadamente dois meses.

Nesse caso, a seleção de amostras intencionais é realizada com base no discernimento do pesquisador. Para Aaker et al. (1995), na amostragem intencional escolhem-se elementos típicos e representativos para compor a amostra. Em alguns casos, é comum recorrer à escolha de profissionais especializados para definição das características da amostra (Kish, 1965). Neste estudo, a seleção dos elementos foi guiada pela expertise e experiência de mercado do autor, engenheiro de dados sênior na Vix Logística S/A, juntamente com o auxílio do professor orientador, especialista em Estatística e doutor em Modelagem Computacional pelo Instituto Federal do Espírito Santo (IFES). O método de julgamento é amplamente utilizado quando se busca selecionar uma amostra representativa, a qual, em alguns casos, é preferível em relação a métodos aleatórios. Em determinadas situações, a amostragem por julgamento pode ser mais fidedigna do que uma amostra probabilística (Aaker et al., 1995). Segundo Kish (1965) e Aaker et al. (1995), essa prática é frequentemente usada quando se tratam de universos de estudos mais amplos.

Após uma análise minuciosa, o processo de seleção das amostras foi dividido em dois grupos distintos de empresas:

1. Empresas que já adotaram sistemas de informação que viabilizam a cultura *data-driven*: Este grupo inclui empresas que, além de lidar com BD, possuem uma arquitetura de sistemas de informação estabelecida e estrutura organizacional que promove a cultura orientada a dados. Também abrange empresas que estão em processo de implementação de uma cultura *data-driven*, ou seja, que estão desenvolvendo e integrando práticas e sistemas para tornar a tomada de decisões baseada em dados como parte de suas operações e estratégias.

2. Empresas que ainda não adotaram sistemas de informação que viabilizam a cultura *data-driven*: Este grupo inclui empresas que, apesar de lidarem com BD, ainda não possuem sistemas de informação adequados para tal, além de não possuírem estrutura que promove a cultura orientada a dados ou são principiantes nesse quesito.

Essa abordagem foi adotada com o objetivo de validar se o grande volume de dados disponível em ambientes de BD atua como um catalisador significativo na geração de valor para a organização; e, segundo, se empresas com uma cultura *data-driven* são mais eficientes na tomada de decisões dos seus negócios. Com base nestas amostras distintas, o estudo busca realizar uma análise comparativa dos resultados obtidos. Busca-se compreender a visão das empresas com diferentes estágios de maturidade de implementação da cultura *data-driven*. Esta análise comparativa permite explorar como as empresas que já adotaram práticas *data-driven* estão utilizando seus dados sob a ótica das decisões estratégicas, além de oferecer *insights* sobre os fatores de sucesso e empecilhos à implementação da cultura de dados.

3.4 PROCEDIMENTOS DE COLETA E ANÁLISE DOS DADOS

Inicialmente, foi realizado um teste piloto de pesquisa com atribuição de 10 testes para validar a ausência de erros no processo de preenchimento do formulário e certificar a aderência das perguntas. Nesse cenário, a pesquisa foi disponibilizada para participantes seletos, com experiência comprovada na área de dados, de forma a coletar feedbacks sobre a qualidade das perguntas e estrutura do questionário. Em seguida, as potenciais empresas que se adequaram às amostras previamente definidas foram selecionadas e, em seguida, a pesquisa foi disponibilizada. De forma a manter a confiabilidade das opiniões coletadas como representação da visão organizacional daquela empresa, foram escolhidos somente gestores de áreas ou diretores dessas organizações. Após a conclusão da coleta de dados, que levou cerca de dois meses, iniciou-se o processo de tabulação dos resultados, que consiste na conversão das respostas em dados numéricos para análises (Samara; Barros, 2007). Esta etapa foi realizada com o auxílio de planilhas eletrônicas, permitindo a organização sistemática das informações coletadas. Em seguida,

procedeu-se à limpeza dos dados, sendo essa uma etapa fundamental para garantir a qualidade e confiabilidade dos dados (Malhotra, 2012). Durante este processo, foram identificados e excluídos quatro questionários que apresentavam padrões de resposta inconsistentes ou que não atendiam aos critérios de inclusão estabelecidos para o estudo. Foram removidos questionários nos quais os respondentes selecionaram opções mutuamente excludentes, indicando possível falta de engajamento ou compreensão inadequada das perguntas. Após a conclusão deste processo de validação e limpeza, o conjunto final de dados para análise foi estabelecido, compreendendo 62 questionários preenchidos e validados. Para fins de visualização dos resultados, foram adotados 60 questionários, 30 para a amostra A e o restante para amostra B. A etapa seguinte compreende a análise exploratória dos dados, na qual para facilitar a visualização dos resultados, a amostra A, que inclui empresas que já implementaram sistemas de informação que sustentam uma cultura orientada por dados, será denominada “Organizações Avançadas em Dados” (OAD). Por outro lado, a amostra B, que abrange empresas que ainda não adotaram tais sistemas, será chamada de “Organizações Iniciantes em Dados” (OID).

Em seguida, foi realizada a análise inferencial dos dados para validação das hipóteses da pesquisa. O método de inferência estatística, segundo Akter et al. (2019), se destaca no contexto da análise de dados por permitir conclusões sobre uma população a partir das amostras do estudo. Isso significa que os resultados obtidos podem ser generalizáveis para um grupo representativo. Testes estatísticos como o teste t de Student, o teste Qui-quadrado e a análise de variância são comumente utilizados nesse campo de estudo (Rehman et al., 2016). Um dos métodos mais relevantes de inferência estatística na engenharia é o teste de hipóteses, amplamente aplicado em problemas de tomada de decisão (Montgomery; Runger, 2018). Conforme Devore (2019, p. 296), "uma estatística de teste é uma função dos dados amostrais utilizada como base para decidir se a hipótese nula deve ser rejeitada". Outro elemento fundamental na análise de hipóteses é o valor-p, que indica a probabilidade de se observar nos dados disponíveis um resultado da estatística de teste tão ou mais contraditório em relação à hipótese nula (Devore, 2019). A decisão final depende da comparação entre o valor-p e um nível de significância pré-definido, o que determina a aceitação ou rejeição da hipótese nula (Devore, 2019). Dada a natureza qualitativa das variáveis coletadas no questionário,

optou-se pela aplicação do teste de proporções Qui-quadrado neste estudo, de maneira a comparar as OAD com as OID e identificar se existe uma associação significativa entre as características desses grupos e a forma como seus processos decisórios são embasados. Nesse sentido, pretende-se validar se empresas com uma cultura *data-driven* são mais eficazes na tomada de decisão e se o grau de maturidade técnica é um fator importante para obtenção de valor com BD. A análise inferencial foi elaborada com a utilização do software R-Project, um programa amplamente usado em análises estatísticas e visualização de dados (Venables; Smith, 2014).

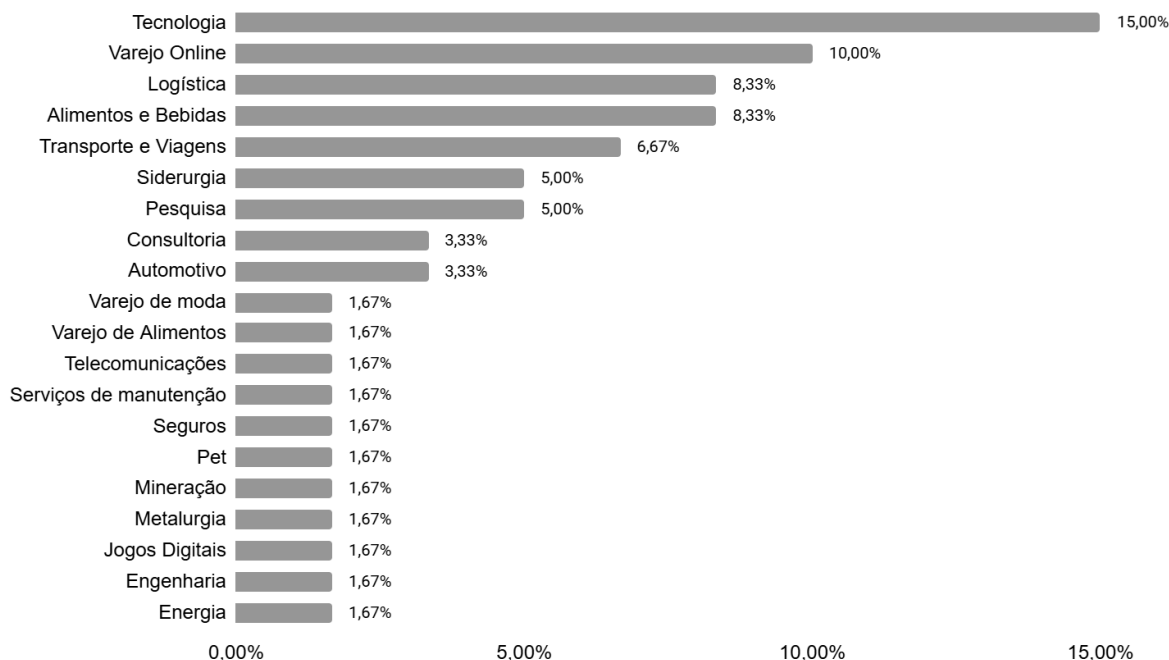
4 APRESENTAÇÃO E ANÁLISE DOS DADOS

Este capítulo tem como objetivo apresentar a aplicação e análise dos dados resultantes da pesquisa. Primeiramente, foi descrita a condução do teste piloto e a aplicação do questionário oficial. Por conseguinte, foi feita uma análise descritiva dos dados coletados, na qual são evidenciados os resultados da pesquisa e suas interpretações.

4.1 ANÁLISE DO PERFIL DOS RESPONDENTES

O presente estudo considerou 60 empresas de diversas regiões do país que se adequam às condições de seleção da amostra. De forma a melhor compreender o perfil das organizações participantes, destaca-se através do gráfico 3 o segmento de mercado dos respondentes.

Gráfico 3 - Segmento de mercado das empresas participantes



Fonte: elaborado pelo autor.

Entre os setores destacados, observa-se uma maior participação de empresas do ramo de Tecnologia, Varejo online e Logística. Os setores Automotivos, Varejo de

Moda, Serviços de Manutenção, Seguros, Telecomunicações, Mineração, Pet, Jogos Digitais, Engenharia e Energia, apesar da menor participação, também agregam valor à pesquisa, ampliando a aplicabilidade dos resultados obtidos. Com relação ao critério de seleção por amostragem intencional, o segmento de mercado não foi uma condicionante estrita.

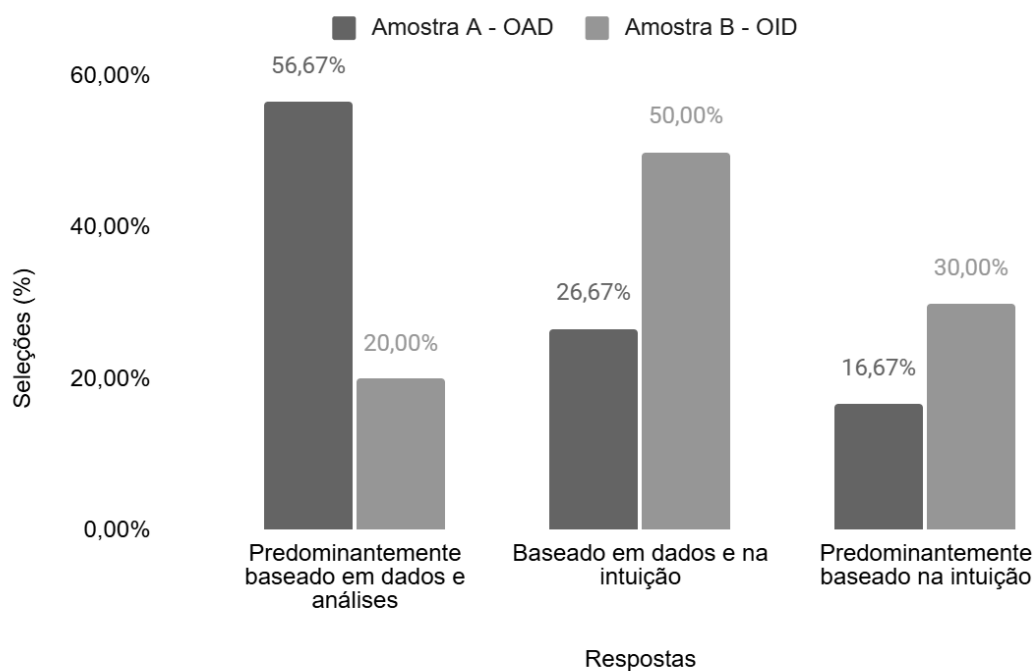
4.2 ANÁLISE EXPLORATÓRIA DOS RESULTADOS

Nesta seção, realiza-se a análise exploratória dos dados obtidos a partir do questionário aplicado, com o objetivo de interpretar e compreender os resultados. Para isso, são apresentados gráficos e tabelas que auxiliam na visualização das informações e no reconhecimento de padrões relevantes.

4.2.1 Processo decisório

A análise comparativa entre os dois grupos, referente ao uso de dados no processo decisório, é apresentada no gráfico 4 e na tabela 1, que ilustram as respostas das organizações à pergunta: “O processo de tomada de decisão na sua empresa baseia-se em dados ou na intuição?”. As respostas foram classificadas em três categorias: (i) predominantemente baseada em dados e análises, (ii) baseada em uma combinação de dados e intuição, e (iii) predominantemente baseada na intuição.

Gráfico 4: Diferenças na tomada de decisão entre OADs e OIDs



Fonte: elaborado pelo autor.

Tabela 1: Classificação do processo decisório pelas OAD e OID

Tomada de Decisão			
Respostas	Amostra A - OAD	Amostra B - OID	Total geral
Predominantemente baseado em dados e análises	17	6	23
Baseado em dados e na intuição	8	15	23
Predominantemente baseado na intuição	5	9	14
Total geral	30	30	60

Fonte: elaborado pelo autor.

Conforme ilustrado no gráfico 4, o grupo das Organizações Avançadas em Dados, possui uma predominância para a tomada de decisões baseada em dados, com 56,6% (n=17) das organizações relatando que suas decisões são fundamentadas em dados e análises. Uma proporção menor, 26,6% (n=8), indica uma abordagem híbrida, combinando dados e intuição, enquanto 16,67% (n=5) relatam uma dependência predominante da intuição no processo decisório. As frequências das respostas são indicadas na tabela 1.

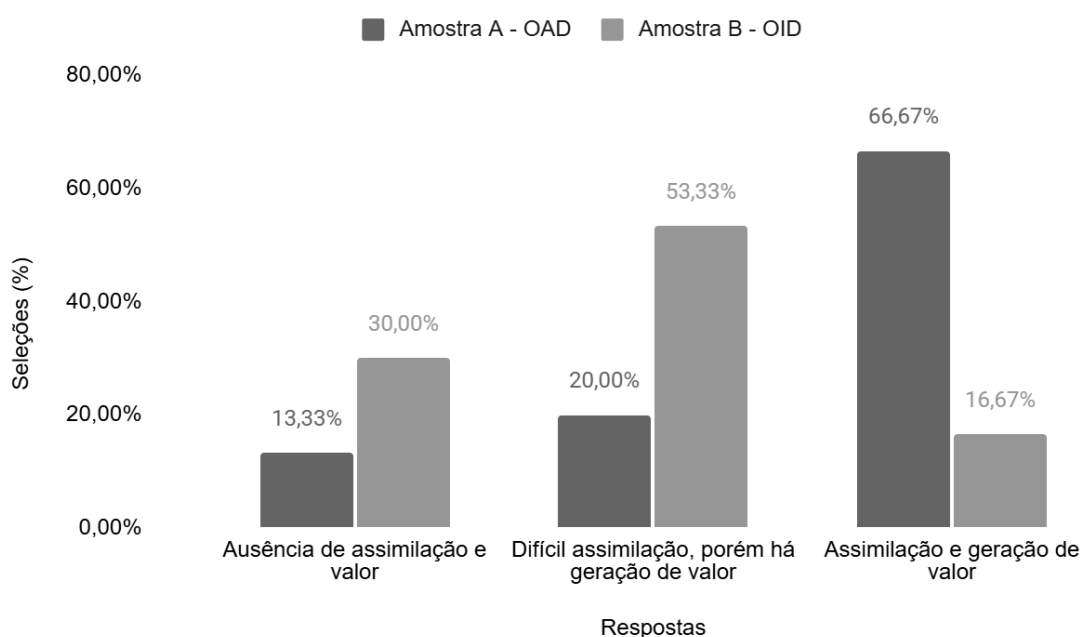
Em contraste, o grupo das OID apresenta um perfil decisório diferente. Neste grupo, apenas 20% (n=6) das organizações reportam a tomada de decisão predominantemente baseada em dados e análises. A maioria, 50% (n=15), adota uma abordagem mista, integrando dados e intuição, enquanto 30% (n=9) ainda dependem principalmente da intuição para orientar suas decisões.

Estes resultados sugerem que as OAD possuem uma maior capacidade de utilizar dados e análises como principal fundamento para suas decisões estratégicas e operacionais. Por outro lado, as OID, ainda em processo de desenvolvimento de analítico, tendem a adotar uma abordagem fundamentada entre dados e intuição.

4.2.2 Capacidade de assimilação do *big data*

A segunda pergunta do questionário visa avaliar o grau de assimilação das organizações em relação às variáveis volume, variedade e velocidade no contexto do BD. Nesse contexto, o gráfico 5 fornece uma visão comparativa da capacidade de assimilação de BD e a geração de valor das OAD e OID. A frequência de respostas dos participantes são detalhados na tabela 2. Esses elementos ilustram os resultados obtidos à indagação: “Sob a perspectiva do alto volume, variedade e velocidade dos dados no contexto do *big data*, escolha a opção que melhor representa a percepção da sua empresa.” As respostas foram agrupadas em três categorias: (i) Não conseguimos assimilar o *big data* e nem extrair valor dos dados., (ii) Temos dificuldade de assimilar o *big data* mas conseguimos extrair valor dos dados., e (iii) Não temos dificuldade de assimilar o *big data* e conseguimos extrair valor dos dados.

Gráfico 5: Diferenças de assimilação e geração de valor com dados



Fonte: elaborado pelo autor.

Tabela 2: Classificação da assimilação do *big data* pelas OAD e OID

Assimilação do <i>Big Data</i> e suas principais variáveis			
Respostas	Amostra A - OAD	Amostra B - OID	Total geral
Não conseguimos assimilar o <i>big data</i> e nem extrair valor dos dados.	4	9	13
Não temos dificuldade de assimilar o <i>big data</i> e conseguimos extrair valor dos dados.	20	5	25
Temos dificuldade de assimilar o <i>big data</i> mas conseguimos extrair valor dos dados.	6	16	22
Total geral	30	30	60

Fonte: elaborado pelo autor.

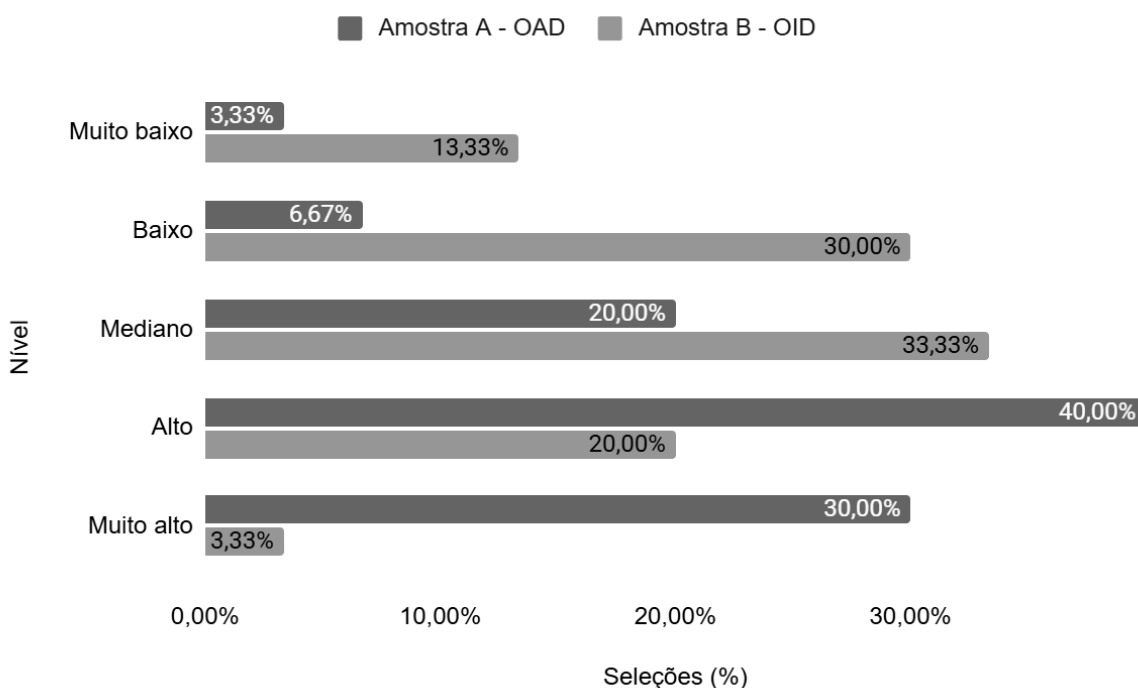
No grupo das OAD, a maioria significativa (66,7%, n=20) reporta que consegue assimilar o BD e extrair valor dos dados. Um percentual menor (20%, n=6) afirma ter dificuldade na assimilação, mas ainda consegue gerar valor a partir dos dados, enquanto 13,3% (n=4) indicam que não conseguem assimilar o BD e tampouco extrair valor.

Em contraste, o grupo das OID apresentam comportamento diferente. Mais da metade das empresas (53,3%, n=16) relatam dificuldades na assimilação dos dados, mas conseguem extrair algum valor. Aproximadamente 30% (n=9) das OID afirmam que não conseguem assimilar nem gerar valor dos dados, e apenas 16,7% (n=5) conseguem assimilar o BD e gerar valor de maneira eficaz.

4.2.3 Maturidade técnica para lidar com *big data*

A terceira pergunta do questionário busca avaliar a maturidade técnica e analítica das organizações no contexto do BD. O Gráfico 6 apresenta uma análise comparativa entre as OAD e as OID, enquanto a Tabela 3 detalha a frequência de respostas dos participantes. A questão “Como você classifica o nível de maturidade técnica e analítica da sua organização para lidar com *big data*?” utiliza escala de Likert: (i) Muito baixo, (ii) Baixo, (iii) Mediano, (iv) Alto e (v) Muito alto. Há uma indicação no questionário para que se leve em conta a qualidade dos estágios de coleta, processamento, armazenamento e análise de dados.

Gráfico 6: Comparativo do nível de maturidade para lidar com *big data*



Fonte: elaborado pelo autor.

Tabela 3: Classificação da maturidade técnica e analítica pelas OAD e OID

Maturidade Técnica e Analítica			
Respostas	Amostra A - OAD	Amostra B - OID	Total geral
Muito baixo	1	4	5
Baixo	2	9	11
Mediano	6	10	16
Alto	12	6	18
Muito alto	9	1	10
Total geral	30	30	60

Fonte: elaborado pelo autor.

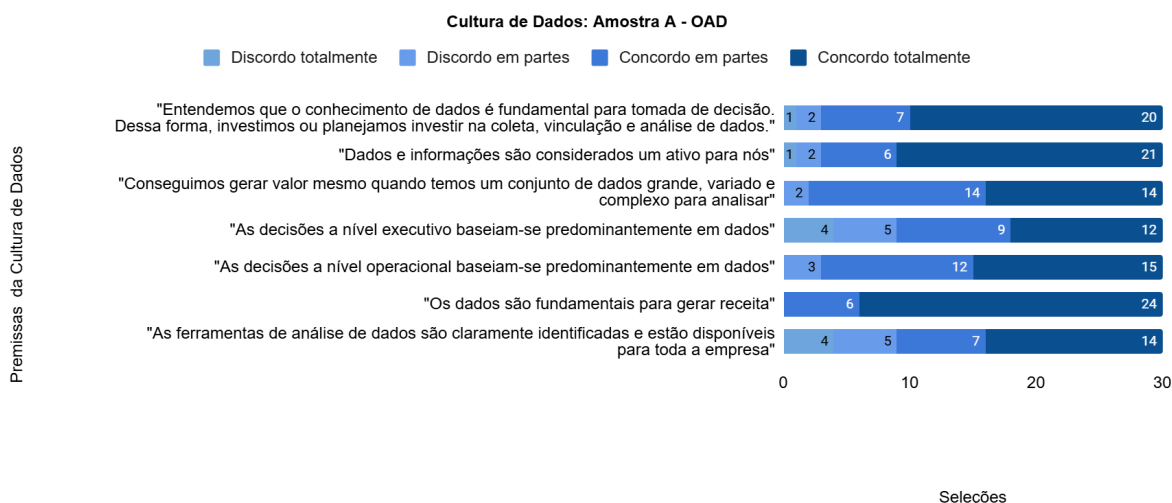
A maioria no grupo das OAD, representando 70% do total, classifica sua maturidade como Alto (40%, n=12) ou Muito alto (30%, n=9). Esses valores indicam uma capacidade bem desenvolvida para gerenciar as etapas principais de arquitetura de dados. Uma fração de 10% (n=3) das OAD se encontra nas categorias de maturidade Muito baixo e Baixo.

O cenário das OID revela uma incidência de processos com qualidade mediana (33%, n=10). Combinando as categorias Muito baixo (13,33%, n=4) e Baixo (30%, n=9), observamos que 43,33% das OID reconhecem que suas capacidades estão aquém do ideal. 20% (n=6) conseguem se posicionar como Alto. A análise indica que as OAD estão em um estágio avançado de maturidade técnica e analítica, permitindo-lhes lidar com os desafios do BD de maneira eficaz. Em contrapartida, as OID reconhecem deficiências em seus processos voltados a dados.

4.2.4 Concordância com as premissas da cultura de dados

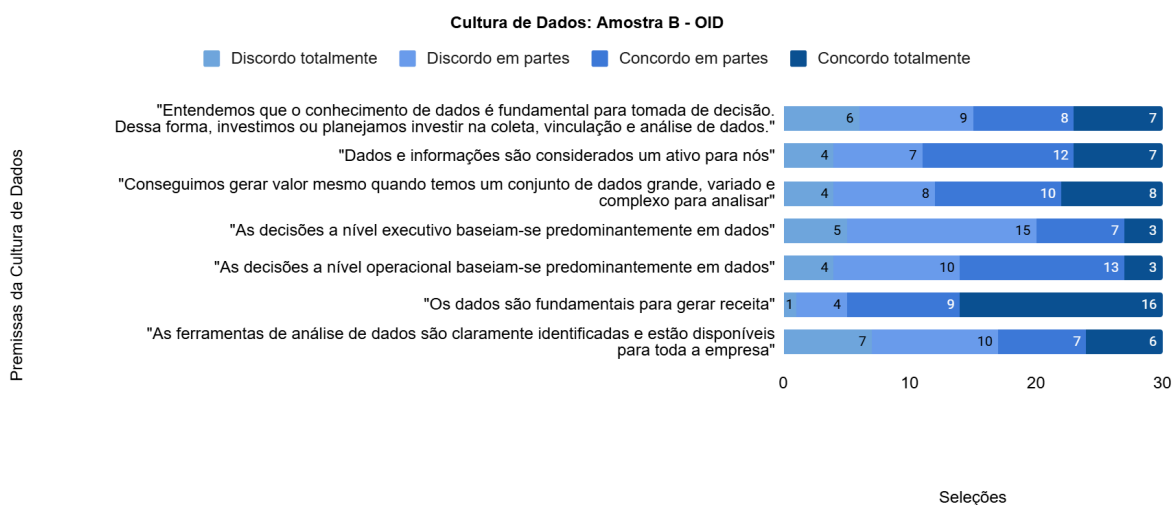
Nesta seção analisa-se os resultados da quarta pergunta: "Até que ponto você concorda com as afirmações a seguir sobre a cultura de dados na sua empresa?". Nesse tópico, a empresa selecionou o nível de importância atribuída a algumas sentenças fundamentais da cultura de dados. Os gráficos 7 e 8 elencam o nível de importância dado pelas OAD e OID, respectivamente.

Gráfico 7: Concordância das OAD com as premissas da cultura de dados



Fonte: elaborado pelo autor.

Gráfico 8: Concordância das OID com as premissas da cultura de dados



Fonte: elaborado pelo autor.

Sobre a primeira afirmação - "Entendemos que o conhecimento de dados é fundamental para a tomada de decisão. Dessa forma, investimos ou planejamos investir na coleta, vinculação e análise de dados" — observa-se pelo gráfico 7 que o grupo OAD se destaca com 20 seleções de "concordo totalmente" e 7 de "concordo em partes", enquanto o grupo OID, apresentado pelo gráfico 8, mostra uma

distribuição de 7 “concordando totalmente” e 9 “em partes”. Há aqui uma diferença notável com relação à percepção de valor para retorno de investimentos.

Na premissa "Os dados são fundamentais para gerar receita", as OAD apresentam maior concordância, com 24 seleções de "concordo totalmente". Em contrapartida, o grupo OID possui 16 respostas neste mesmo nível. Essa lacuna pode indicar que as empresas iniciantes ainda não estão maximizando o potencial dos dados para impulsionar sua rentabilidade, o que dialoga com uma possível falta de clareza do retorno financeiro.

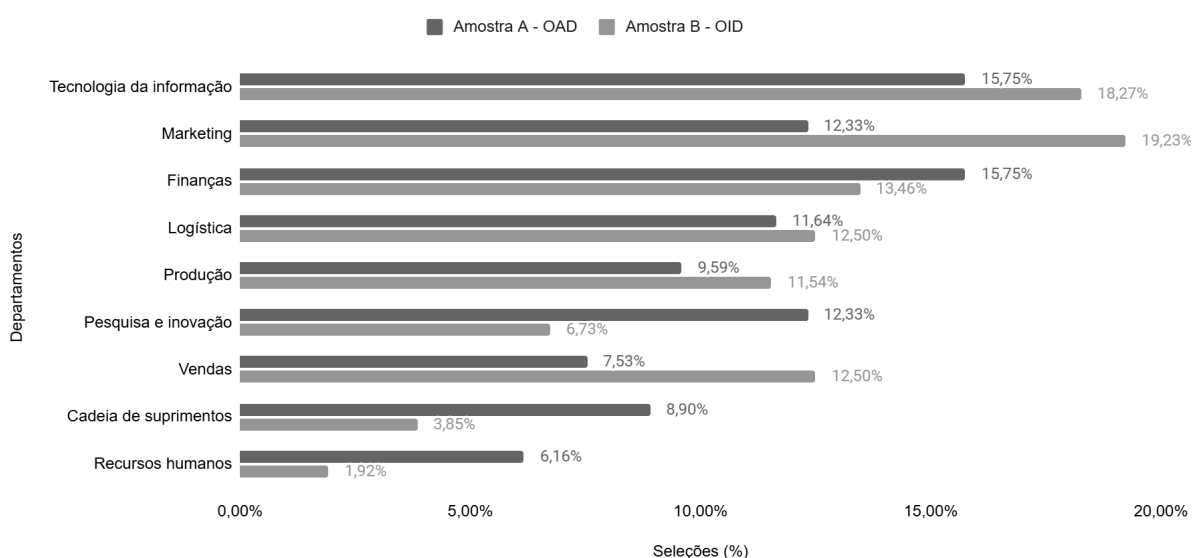
Sobre a afirmação "As ferramentas de análise de dados são claramente identificadas e estão disponíveis para toda a empresa", as OAD mostram uma concordância total de 14 seleções e 7 em partes. Já no grupo OID, apenas 6 participantes concordaram totalmente e 7 em partes, o que indica um obstáculo considerável para essas organizações na democratização do acesso às informações.

Ao analisar as decisões operacionais e executivas fundamentadas em dados, verifica-se uma diferença moderada, em ambas as amostras, na fundamentação das decisões em dados por parte da gestão e da operação. Nos dois casos, a operação consegue embasar melhor suas decisões em dados quando comparadas à gestão. As decisões operacionais tendem a ser mais imediatas e relacionadas ao dia a dia da empresa, o que pode indicar o uso em maior escala dos dados no processo decisório, através de indicadores de desempenho e métricas operacionais. As decisões executivas, que envolvem planejamento estratégico e diretrizes de longo prazo, podem demandar análises mais profundas e complexas dos dados, envolvendo incertezas e variáveis que não são tão facilmente mensuráveis (Sharda *et al.*, 2019). Para essas decisões, a integração dos dados pode exigir maior sofisticação nas análises e uma cultura de dados mais consolidada, o que, nas OID, ainda não está completamente implementado, com 15 discordâncias em partes e 5 totais.

4.2.5 Utilização de dados por departamento

A quinta pergunta do questionário: "Quais departamentos da sua empresa mais utilizam dados em suas tomadas de decisão e processos operacionais?" procura adentrar no aspecto da utilização de dados por divisão setorial das organizações.

Gráfico 9: Comparativo da utilização de dados por departamento



Fonte: elaborado pelo autor.

A partir do gráfico 9, os resultados indicam nas OAD um grau mínimo de aplicabilidade dos dados em suas rotinas. Essa percepção é notável quando avalia-se a distribuição percentual entre as colunas. Finanças e Tecnologia da Informação são os departamentos que se destacam, com 23 respostas em OAD (frequência dos resultados apresentados na tabela 4), seguidos por Marketing e Pesquisa e Inovação, que também têm uma presença significativa. Ainda para as OAD, mesmo os departamentos que recebem menos ênfase, como Recursos Humanos e Cadeia de Suprimentos, demonstram considerável índice de utilização, embora em menor escala.

Para as OID, a utilização de dados tende a ser mais restrita a alguns departamentos, evidenciando uma fundamentação menor em dados na tomada de decisões em setores como Cadeia de Suprimentos e Recursos Humanos, com 4 e 2 respostas

respectivamente. Ainda no gráfico 9, é possível observar uma diferença maior entre as respostas para esse grupo, com variações percentuais mais significativas entre as seleções. No aspecto geral, os setores de Marketing e Tecnologia da Informação se destacam, superando o terceiro colocado Finanças com quase cinco pontos percentuais.

Tabela 4: Ranking dos setores com maior utilização de dados

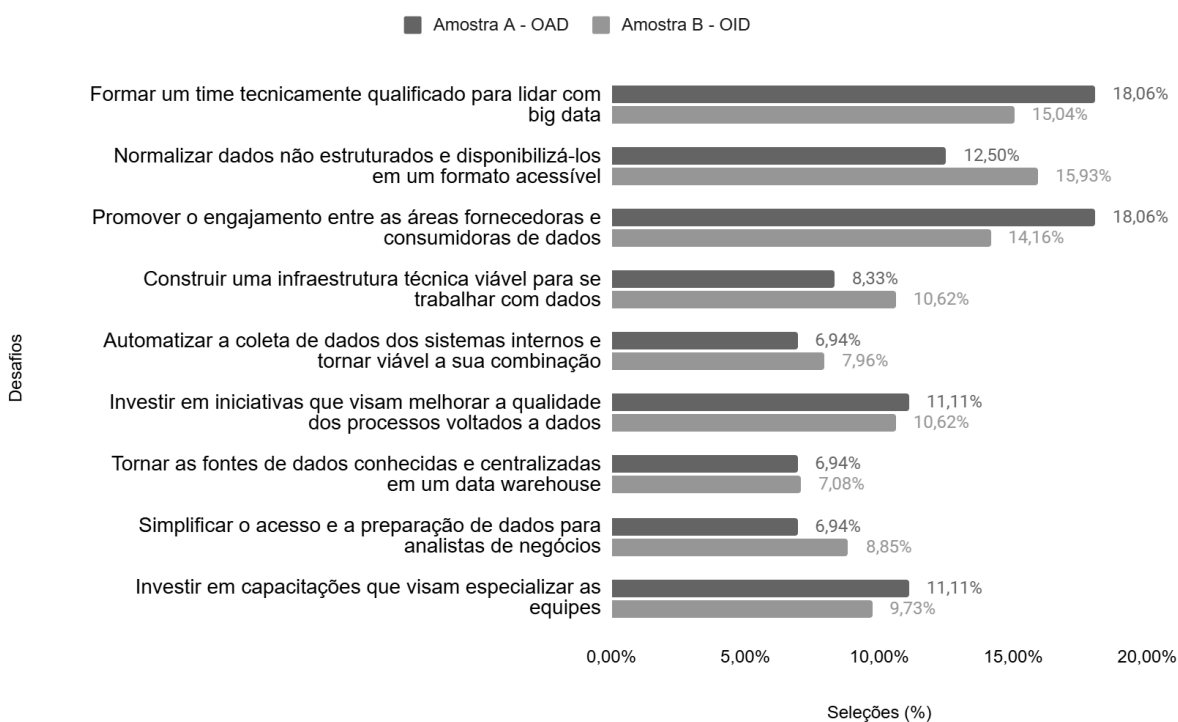
Análise de Utilização e Fundamento em Dados por Setor			
Respostas	Amostra A - OAD	Amostra B - OID	Total geral
Tecnologia da informação	23	19	42
Marketing	18	20	38
Finanças	23	14	37
Logística	17	13	30
Produção	14	12	26
Pesquisa e inovação	18	7	25
Vendas	11	13	24
Cadeia de suprimentos	13	4	17
Recursos humanos	9	2	11
Total geral	146	104	250

Fonte: elaborado pelo autor.

4.2.6 Desafios na geração de valor com dados

Os resultados da sexta pergunta: "Quais são os maiores desafios na geração de valor com os dados na sua empresa?" indicam os obstáculos enfrentados pelas empresas para extrair retorno dos dados. Nesse aspecto, é possível observar alguns pontos em comum entre as OAD e OID a partir do gráfico 10. Entre eles, um dos principais desafios identificados por ambos os grupos é a formação de equipes qualificadas para atuar com BD. Essa dificuldade é levantada por Mirarab, Mirtaheri e Asghari (2019), que discutem as barreiras que as empresas enfrentam para compor equipes especializadas em BD, especialmente devido às insuficiências de mercado. Muitas vezes, as organizações buscam superar esse obstáculo por meio de treinamentos internos.

Gráfico 10: Comparativo dos desafios da geração de valor com dados



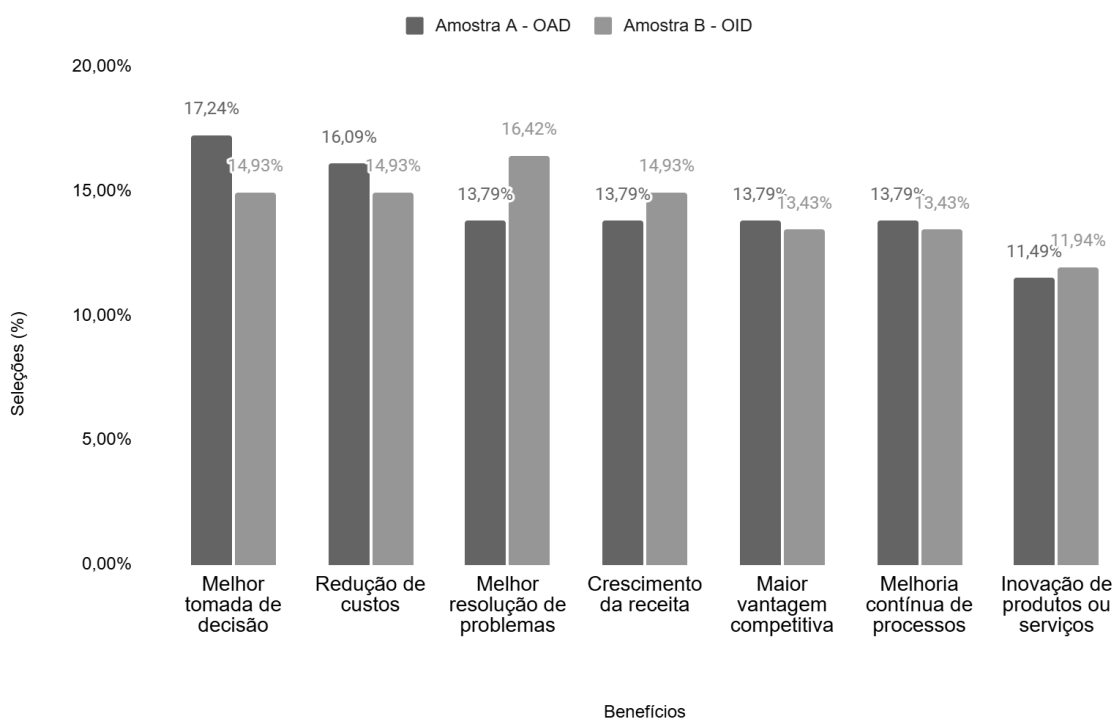
Fonte: elaborado pelo autor.

Destaca-se também a necessidade de promover o engajamento entre as áreas fornecedoras e consumidoras de dados como uma problemática comum entre os dois grupos, com 18,06% de seleção para as Organizações Avançadas em Dados (OAD) e 14,16% para as Organizações Iniciais em Dados (OID). Em relação às diferenças enfrentadas, observa-se que as OID possuem mais dificuldades técnicas, especialmente em aspectos como a normalização de dados não estruturados e a automação da coleta de dados. Por outro lado, as OAD tendem a destacar suas principais dificuldades em questões de gestão, engajamento e profissionalização das equipes.

4.2.7 Benefícios obtidos através do data-driven

Os resultados da sétima pergunta: "Quais benefícios sua empresa já obteve através do uso de dados?" indicam os retornos obtidos pelas empresas ao utilizarem dados como ferramenta estratégica.

Gráfico 11: Comparativo dos benefícios obtidos com dados



Fonte: elaborado pelo autor.

Tabela 5: Ranking dos benefícios do DD

Benefícios Obtidos através dos Dados			
Respostas	Amostra A - OAD	Amostra B - OID	Total geral
Melhor tomada de decisão	15	10	25
Redução de custos	14	10	24
Melhor resolução de problemas	12	11	23
Crescimento da receita	12	10	22
Maior vantagem competitiva	12	9	21
Melhoria contínua de processos	12	9	21
Inovação de produtos ou serviços	10	8	18
Total geral	87	67	154

Fonte: elaborado pelo autor.

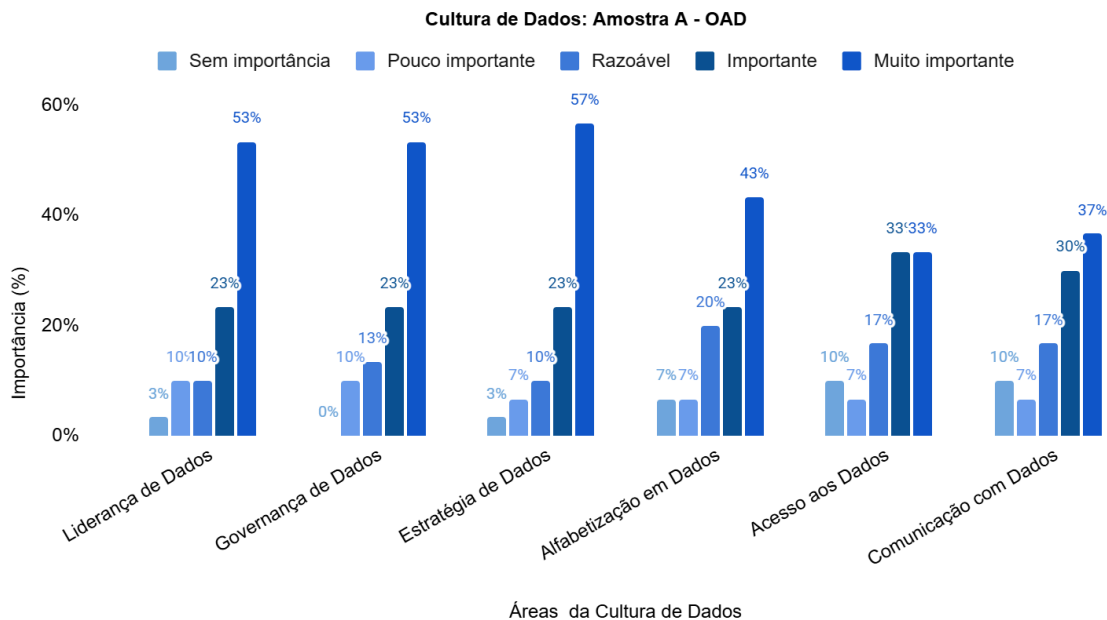
Analisa-se, pelo gráfico 11, que ambas as amostras destacam o crescimento da receita como um dos principais benefícios, com 14,93% para as OAD ($n = 12$) e 13,79% para as OID ($n = 10$). As frequências de respostas são detalhadas na tabela 5. Isso sugere que, independentemente do nível de maturidade em dados, as

empresas já atingiram em algum momento um aumento de receita proporcionado pelo uso dos dados. Além disso, a melhor tomada de decisão é valorizada por 17,24% das OID (n = 15) e 14,93% das OAD (n = 10). Um ponto relevante é a redução de custos, onde as OAD apresentam 16,09% (n = 87) em comparação a 14,93% (n = 67) das OID. Isso pode sugerir que as OAD conseguem identificar fatores que impactam nos custos da organização de forma mais assertiva, além de apresentarem uma maior faixa para resolução de problemas.

4.2.8 Cultura de Dados

O tópico 8 do questionário “Indique o nível de importância que sua empresa atualmente atribui às seguintes frentes de cultura de dados.” foi utilizado para levantar o nível de importância atribuído pelas organizações para cada frente da estrutura da cultura de dados, apresentado como um *framework* pela BARC.

Gráfico 12: Importância atribuída pelas OAD aos pilares da cultura de dados

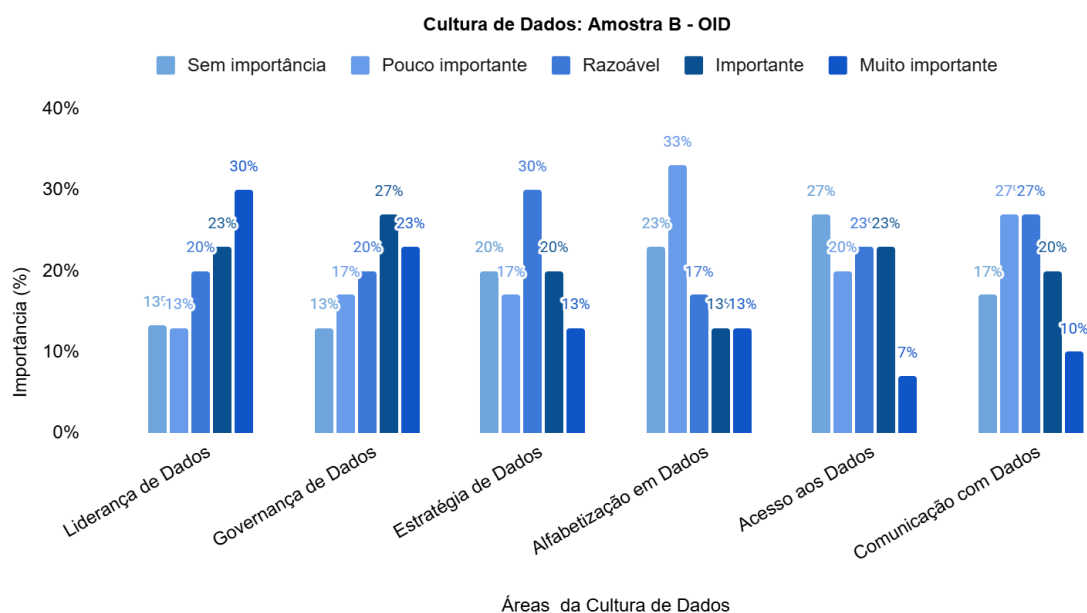


Fonte: elaborado pelo autor.

O gráfico 12 representa a percepção de relevância das OAD para cada pilar da cultura de dados. Verifica-se:

- Liderança de Dados: 53% das OAD consideram muito importante, e 23% importante, totalizando 76% de alta relevância. Isso indica que a maioria das OAD reconhece o papel da liderança na condução de iniciativas baseadas em dados.
- Governança de Dados: 53% veem como muito importante, e 23% como importante (76% no total). Isso sugere um forte reconhecimento da necessidade de políticas e práticas para gerenciar dados efetivamente.
- Estratégia de Dados: 57% consideram muito importante, e 23% importante (80% no total). Esta é a área com a maior percepção de importância, indicando que as OAD priorizam o alinhamento estratégico das iniciativas de dados com os objetivos de negócio.
- Alfabetização em Dados: 43% veem como muito importante, e 23% como importante (66% no total). Embora ainda seja considerada relevante, há uma diminuição considerável na percepção de importância comparada às áreas previamente mostradas.
- Acesso aos Dados: 33.33% consideram muito importante, e outro 33.33% importante (66.66% no total). Isso sugere que, embora valorizado, o acesso aos dados é visto como menos imperativo em comparação com outras áreas.
- Comunicação com Dados: 37% veem como muito importante, e 30% como importante (67% no total). Indica também um reconhecimento baixo a moderado da importância de comunicar efetivamente insights baseados em dados.

Gráfico 13: Importância atribuída pelas OAD aos pilares da cultura de dados



Fonte: elaborado pelo autor.

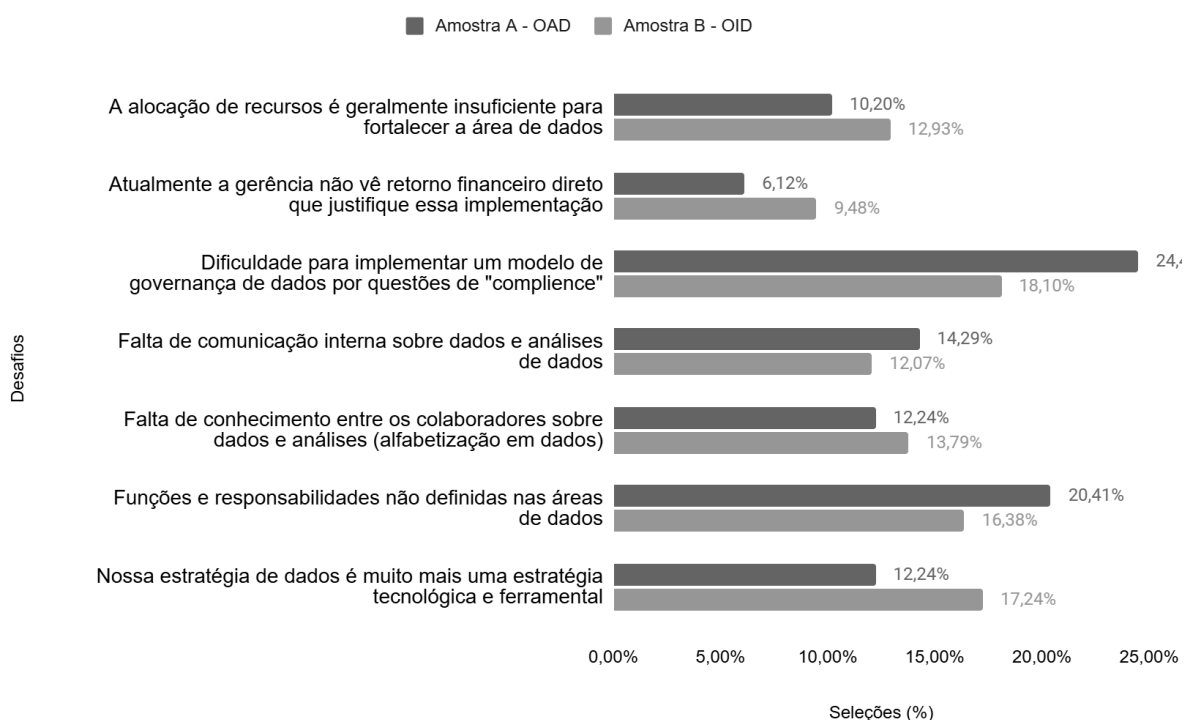
O gráfico 13 representa a percepção das OID ao mesmo tópico. Verifica-se:

- Liderança de Dados: 30% consideram muito importante, e 20% importante, totalizando 50% de alta relevância. Comparado aos outros pilares, sua relevância foi considerável.
- Governança de Dados: 23% veem como muito importante, e 27% como importante (50% no total). No contexto das OID, indica uma compreensão elevada da importância da governança de dados.
- Estratégia de Dados: 13% consideram muito importante, e 20% importante (33% no total). Esta é a área com menor percepção de importância entre as OID, contrastando fortemente com as OAD.
- Alfabetização em Dados: 13% veem como muito importante, e 17% como importante (30% no total). A baixa porcentagem sugere que muitas OID ainda não reconhecem plenamente a importância da alfabetização em dados.
- Acesso aos Dados: 23% consideram muito importante, e 7% importante (30% no total). Isso indica uma baixa valorização do acesso aos dados.
- Comunicação com Dados: 10% veem como muito importante, e 20% como importante (30% no total). Esta baixa porcentagem sugere que muitas OID ainda não priorizam a comunicação efetiva de insights baseados em dados.

4.2.9 Obstáculos para implementação da cultura de dados

O tópico 9 do questionário “Quais são ou foram os principais obstáculos à implementação ou estabelecimento da cultura de dados na sua empresa?” aborda possíveis desafios enfrentados pelas empresas na adoção da cultura de dados.

Gráfico 14: Comparativo dos desafios para implementar a cultura de dados



Fonte: elaborado pelo autor.

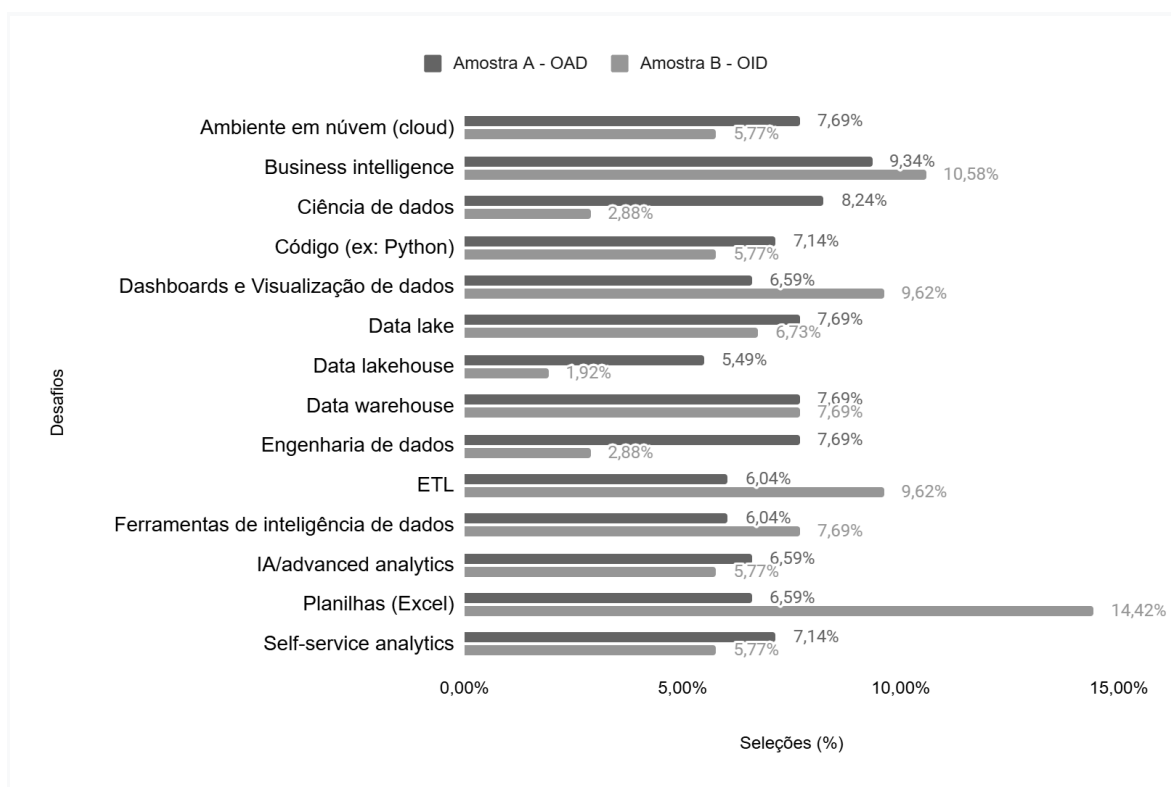
Através da análise comparativa do gráfico 14, ambos os grupos identificam dificuldades para implementação de um modelo de governança de dados por questões de "compliance" como o maior desafio, sendo mais acentuado nas OAD. Entretanto, as OAD dão maior ênfase à definição de funções e responsabilidades nas áreas de dados, enquanto as OID destacam questões tecnológicas e de alocação de recursos com maior frequência. Isso sugere que, à medida que as organizações avançam em sua maturidade de dados, o foco tende a se deslocar de aspectos técnicos para questões mais estratégicas e organizacionais.

As OID demonstram uma visão mais limitada do papel estratégico dos dados, considerando-os principalmente sob a abordagem ferramental. Além disso, enfrentam maiores dificuldades em justificar investimentos e identificar retornos financeiros diretos das iniciativas de dados. Em contraste, as OAD parecem ter uma compreensão mais ampla do valor estratégico dos dados, enfrentando menos obstáculos relacionados a recursos e retorno financeiro. Isso indica que, com o amadurecimento da cultura de dados, as organizações tendem a desenvolver uma visão mais holística do papel dos dados em seus negócios.

4.2.10 Ferramentas mais utilizadas

Os resultados da última pergunta do questionário “Quais aspectos tecnológicos e abordagens são comumente utilizados para lidar com dados na rotina da sua empresa?” elencam as principais ferramentas e competências adotadas rotineiramente pelas OAD e OID.

Gráfico 15: Principais ferramentas e abordagens no trabalho com dados



Fonte: elaborado pelo autor.

A análise do gráfico 15 revela diferenças significativas sobre as tecnologias e práticas adotadas pelas Organizações Avançadas em Dados (OAD) e Organizações Iniciantes em Dados (OID) em seus processos: As OAD demonstram uma adoção mais ampla de tecnologias avançadas. Elas lideram em áreas como ciência de dados (8,24% vs 2,88%), engenharia de dados (7,69% vs 2,88%), e data *lakehouse* (5,49% vs 1,92%). Isso sugere uma abordagem mais madura e sofisticada no uso de dados. As OID mostram uma forte dependência de planilhas Excel (14,42% vs 6,59% nas OAD) e uma maior ênfase em dashboards e visualização de dados (9,62% vs 6,59%). Em contraste, as OAD apresentam um uso superior de tecnologias e competências avançadas como IA/advanced analytics, engenharia e ciência de dados.

4.3 ANÁLISE INFERENCIAL DOS DADOS

Nesta etapa busca-se investigar se existe relação estatisticamente significativa entre algumas variáveis mostradas na análise descritiva. Para isso, aplicaram-se métodos de análise inferencial que permitem testar hipóteses e avaliar padrões nos dados.

4.3.1 Teste Qui-quadrado

O teste χ^2 (Qui-quadrado) proposto pelo estatístico Karl Pearson em 1900 é amplamente utilizado para verificar se existem diferenças significativas entre duas distribuições. Sendo um dos principais testes para avaliar associações, ele calcula o total de desvios entre o número de ocorrências observadas e o de esperadas, determinando a probabilidade de ocorrência desses desvios com base em uma distribuição χ^2 , cujo grau de liberdade são obtidos a partir de uma tabela de contingência, calculado pela fórmula (Pereira, 2000):

$$gl = (l - 1) * (c - 1)$$

Onde:

- l = número de linhas
- c = número de colunas

Dessa forma, para testes de hipóteses, adota-se:

H_0 = Não existe associação entre as categorias

H_1 = Existe alguma associação entre as categorias

De acordo com Pereira (2000), a fórmula estatística do teste Qui-quadrado é dada por:

$$\chi^2 = \sum_{i=1}^l \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Onde:

- O_{ij} = número de casos observados na linha i da coluna j;
- E_{ij} = número de casos esperados para H_0 na linha i da coluna j

Sob a hipótese, a estatística acima segue uma distribuição Qui-quadrado com graus de liberdade representado pela primeira fórmula. A frequência esperada E_{ij} é dada por:

$$E_{ij} = \frac{(n_i * n_j)}{n}$$

Onde:

- n_i = somatório de casos observados na linha i;
- n_j = somatório de casos observados na coluna j;
- n = somatório total de casos da tabela;

4.3.1.1 Teste exato de Fisher

O teste exato de Fisher utiliza tabelas de contingência 2x2 para comparar dois grupos e é indicado para amostras pequenas. Ele calcula a probabilidade exata de ocorrência de uma frequência observada ou de valores mais extremos (Pereira, 2000). Para utilizar o teste exato de Fisher, é necessário que as amostras sejam aleatórias e independentes, componham duas classes mutuamente exclusivas e

incluam, no mínimo, variáveis qualitativas. As frequências observadas devem seguir uma distribuição hipergeométrica. A probabilidade de ocorrência será calculada como:

$$P = \frac{((a+b)!(c+d)!(a+c)!(b+d)!)}{N!a!b!c!d!}$$

Onde:

- *a*: número de ocorrências na célula da primeira linha e primeira coluna.
- *b*: número de ocorrências na célula da primeira linha e segunda coluna.
- *c*: número de ocorrências na célula da segunda linha e primeira coluna.
- *d*: número de ocorrências na célula da segunda linha e segunda coluna.
- *N*: soma total de todas as observações, ou seja, $N=a+b+c+d$

Neste estudo, foram utilizados tanto o teste Qui-quadrado quanto o teste exato de Fisher para assegurar a confiabilidade dos resultados obtidos. O teste Qui-quadrado foi aplicado nos casos em que ao menos 80% das células apresentaram frequências esperadas maiores que 5 e todas as frequências foram superiores a 1 (Pereira, 2000). O teste de Fisher foi incluído como uma validação adicional no primeiro caso, já no segundo por considerar que o tamanho total das amostras não é suficientemente grande.

4.3.2 Tipo de organização e o processo decisório

A fim de validar se empresas que apresentam uma cultura *data-driven* (representado pelo grupo das OAD) são mais eficientes no processo decisório baseado em dados, faz-se necessário investigar até que ponto as diferenças obtidas nos resultados da pesquisa se devem ao acaso ou indicam associação entre a cultura de dados e a eficácia na tomada de decisão. Nesse contexto, foram estabelecidas as seguintes hipóteses:

Hipótese nula (H_0): Não existe associação significativa entre o tipo de organização (OAD ou OID) e a forma como o processo de tomada de decisão é embasado. Ou seja, a forma como ocorre o processo decisório independe do tipo de organização.

Hipótese alternativa (H_1): Existe associação significativa entre o tipo de organização (OAD ou OID) e a forma como o processo de tomada de decisão é embasado. Ou seja, o *data-driven* influência no processo decisório.

Para iniciar as análises no software R foi necessário importar a base de dados para o programa e então construir uma tabela de contingência dos resultados referentes à primeira pergunta do questionário. Os dados são demonstrados na Tabela 1 da seção 4.2.1. Nas análises, utilizou-se o R *commander* da biblioteca “Rcmdr”, com auxílio da biblioteca “MASS” (Fox, 2005). Os resultados do teste Qui-quadrado e teste exato de fisher são demonstrados nas tabelas 6 e 7:

Tabela 6: R - Frequência e percentual das linhas para a primeira associação

Tomada de Decisão	OAD	% Linha (OAD)	OID	% Linha (OID)
Baseado em dados e na intuição	8	34.8%	15	65.2%
Predominantemente baseado em dados e análises	17	73.9%	6	26.1%
Predominantemente baseado na intuição	5	35.7%	9	64.3%

Fonte: elaborado pelo autor.

Tabela 7: R - Estatística de teste para a primeira associação

Teste Estatístico	Estatística de Teste	Graus de Liberdade (df)	p-valor
Teste Qui-Quadrado (Pearson)	$X^2 = 8.5342$	2	0.01402
Teste Exato de Fisher	Não aplicável	Não aplicável	0.01755

Fonte: elaborado pelo autor.

De acordo com os resultados apresentados, a estatística do teste Qui-quadrado calculada foi $\chi^2=8.5342$, com 2 graus de liberdade (df), e valor-p correspondente de 0,01402. Como o valor-p é menor que o nível de significância usual de 0.05, rejeita-se a hipótese nula. A aplicação da hipótese alternativa indica que existe uma

associação estatisticamente significativa entre o tipo de organização (OAD ou OID) e a forma como as decisões são tomadas. O valor-p obtido para o teste exato de Fisher foi 0,01755. Como este valor também é inferior a 0.05, rejeita-se a hipótese nula, o que corrobora a conclusão do teste Qui-quadrado.

Ambos os testes indicam associação significativa entre o tipo de organização e a forma como as decisões são fundamentadas. Especificamente, as Organizações Avançadas em Dados (OAD) apresentam maior tendência a basear suas decisões em dados e análises, enquanto as Organizações Iniciais em Dados (OID) tendem a depender mais da intuição durante o processo decisório. Dessa forma, os resultados sugerem que a cultura *data-driven*, representada pelas OAD, exerce uma influência positiva no processo de tomada de decisão, favorecendo a utilização de dados e análises.

4.3.3 Nível de maturidade e capacidade de assimilação

A fim de validar se empresas que apresentam um nível maior de maturidade técnica e analítica relacionada ao BD possuem maior capacidade de assimilação do BD e geração de valor, foram realizados os testes acima para a segunda e terceira pergunta do questionário:

2. “Sob a perspectiva do alto volume, variedade e velocidade dos dados no contexto do *big data*, escolha a opção que melhor representa a percepção da sua empresa.”
3. “Como você classifica o nível de maturidade técnica e analítica da sua organização para lidar com *big data*?”.

Nesse contexto, a diferenciação entre OAD e OID não foi utilizada para que se tornasse possível o relacionamento dessas duas variáveis sob o formato de tabela de contingência. Assim, foram estabelecidas as seguintes hipóteses:

Hipótese nula (H_0): Não existe associação significativa entre o nível de maturidade técnica e analítica da empresa e a sua capacidade de assimilação do BD, bem como sua capacidade de geração de valor.

Hipótese alternativa (H_1): Existe associação significativa entre o nível de maturidade técnica e analítica da empresa e a sua capacidade de assimilação do BD, bem como sua capacidade de geração de valor.

Para iniciar as análises no software R foi necessário importar a base de dados para o programa e então construir uma tabela de contingência dos resultados referentes às perguntas 2 e 3 do questionário. Os dados totais são demonstrados na Tabela 2 da seção 4.2.2 e Tabela 3 da seção 4.2.3. Nas análises, utilizou-se o R *Commander* da biblioteca “Rcmdr”, com auxílio da biblioteca “MASS” (Fox, 2005). Os resultados do teste Qui-quadrado e teste exato de fisher são demonstrados nas tabela 8 e 9:

Tabela 8: R - Percentual das linhas para a segunda associação

Assimilação do BD	Muito Alto	Alto	Mediano	Baixo	Muito Baixo
Não conseguimos assimilar o BD e nem extrair valor dos dados	0.0%	7.7%	7.7%	53.8%	30.8%
Temos dificuldade de assimilar o BD mas conseguimos extrair valor dos dados	0.0%	27.3%	50.0%	18.2%	4.5%
Não temos dificuldade de assimilar o BD e conseguimos extrair valor dos dados	40.0%	44.0%	16.0%	0.0%	0.0%

Fonte: elaborado pelo autor.

Tabela 9: R - Estatística de teste para a segunda associação

Teste Estatístico	Estatística de Teste	Graus de Liberdade (df)	p-valor
Teste Qui-Quadrado (Pearson)	$\chi^2 = 48.999$	8	0.00000006361
Teste Exato de Fisher	Não aplicável	Não aplicável	0.00000001518

Fonte: elaborado pelo autor.

De acordo com os resultados apresentados, a estatística do teste Qui-quadrado pode não ser indicada para essa proposição, devido à presença de frequências

esperadas menores que 5. Nesse caso, será utilizado somente o teste exato de Fisher, que apresenta valor de 0.00000001518. Como o valor-p é inferior a 0.05, rejeita-se a hipótese nula.

Ao analisar os percentuais por linha na tabela 8, para as organizações que não conseguem assimilar nem extrair valor dos dados, a maioria (53%) apresenta nível de maturidade baixo, e 30.8% muito baixo.

Para as organizações que conseguem assimilar o BD e extrair valor dos dados, apresentam nível alto (44%) e muito alto (41%). Organizações que têm dificuldade de assimilar o BD, mas conseguem extrair valor, os níveis mais frequentes foram mediano (50%) e alto (27%).

O teste exato de Fisher indica que há uma associação significativa entre a maturidade técnica da organização e sua capacidade de lidar com o BD. Isso sugere que a percepção sobre a habilidade de lidar com BD está fortemente relacionada à proficiência interna em coletar, processar e utilizar dados de maneira eficaz.

5 CONSIDERAÇÕES FINAIS

Este trabalho teve como objetivo analisar como as organizações lidam com o desafio de extrair valor do amplo volume de dados disponíveis e verificar se a cultura de dados é um fator de sucesso para uma empresa atingir ganhos e melhorar sua tomada de decisão. Pode-se concluir que o objetivo geral do estudo foi concluído, uma vez que as questões elencadas no objetivo geral foram investigadas através de um survey aplicado a um grupo de grandes organizações e foram minuciosamente analisadas para se obterem conclusões fundamentadas.

Para alcançar os objetivos específicos, os principais conceitos relacionados aos objetos de estudo desta pesquisa, como *big data*, tomada de decisão e cultura de dados foram amplamente analisados com base na literatura. A cultura de dados, em particular, por se tratar de um termo relativamente novo e em ascensão no ambiente organizacional, foi investigada a partir das contribuições publicadas por especialistas de mercado e pesquisas relevantes sobre o tema. Para entender se as organizações utilizam dados para a tomada de decisão, os benefícios obtidos a partir do *data-driven*, e como enfrentam os desafios de extrair valor do *big data*, foram definidas e selecionadas amostras específicas, as quais separam-se os contextos em que elas estão inseridas para uma análise comparativa. Realizou-se, então, uma análise exploratória dos dados coletados na pesquisa, cujos resultados foram apresentados de forma gráfica e visual. De forma a validar as hipóteses previamente levantadas nesta pesquisa, utilizou-se a análise inferencial baseada em testes de proporção na investigação dos casos mencionados.

A pesquisa demonstrou que o crescimento do *big data* nos ambientes organizacionais pode ser um dos principais impulsionadores da transformação empresarial. Em muitas empresas, os dados representam ativos estratégicos que favorecem o crescimento da receita, o posicionamento competitivo no mercado e facilitam a resolução de problemas complexos. No entanto, o simples acúmulo de informações em grande escala não garante vantagem estratégica; torna-se necessário dispor de uma infraestrutura tecnológica robusta para lidar com a dimensão complexa das variáveis do *big data*. A adoção de uma cultura analítica, que prioriza decisões embasadas em dados e evita suposições infundadas,

assegura um nível de governança e estratégia que esteja igualmente dominado por toda a estrutura da empresa.

As Organizações Avançadas em Dados, que já incorporaram a cultura *data-driven*, mostraram maior eficiência em relação ao *big data*, melhor profundidade técnica e maiores chances de extrair benefícios de seus dados quando comparadas às Organizações Iniciais em Dados. Essas empresas, no geral, apresentaram maior concordância com as premissas da cultura de dados e atribuíram mais importância aos seus pilares, com ênfase em liderança, governança e estratégia. De acordo com as análises, tratam-se de empresas que possuem maior facilidade para justificar investimentos na área e enfrentam desafios diferentes das OI no que tange à geração de valor com dados. Tais empresas também empregam ferramentas mais avançadas em sua rotina.

Os resultados inferenciais confirmaram que as OAD são mais eficazes na tomada de decisão e que o nível de maturidade técnica é um fator determinante para a capacidade de assimilação e geração de valor a partir do big data. Esses achados reforçam a relevância da consolidação de uma cultura de dados como um diferencial estratégico em um cenário de crescente complexidade e volume de informações, evidenciando a capacidade das OAD de utilizarem seus ativos de dados para obter vantagens competitivas e melhorar sua eficiência operacional.

No entanto, algumas limitações do estudo devem ser reconhecidas. Primeiramente, o uso de variáveis predominantemente qualitativas no survey restringiu a abrangência de algumas análises, especialmente no que diz respeito à avaliação de impactos financeiros mais objetivos, como o retorno sobre investimento (ROI) e o crescimento de receita diretamente atribuíveis à cultura de dados. Além disso, a amostra limitada a 60 organizações, ainda que intencionalmente selecionada, pode não refletir a diversidade de realidades enfrentadas por empresas de diferentes setores e portes, o que limita a generalização dos resultados.

Com base nessas limitações, sugere-se que estudos futuros incorporem métricas financeiras e indicadores de desempenho organizacional, como faturamento e ROI, para explorar em maior profundidade as relações entre a maturidade analítica, a

cultura de dados e os resultados financeiros. Ampliar a amostra para incluir organizações de variados setores e diferentes níveis de complexidade tecnológica também pode contribuir para uma visão mais representativa das realidades empresariais.

Outra recomendação para futuras pesquisas é a análise detalhada das diferenças entre decisões operacionais e executivas no contexto da cultura de dados. Este estudo identificou nuances nessas diferenças, mas não explorou seu impacto específico no processo de implementação e consolidação da cultura *data-driven*. Investigação adicional sobre como esses dois níveis de decisão interagem com a maturidade técnica e os resultados organizacionais pode oferecer insights para estratégias de adoção e melhoria contínua.

Assim, este trabalho reforça a importância da cultura de dados como um elemento estratégico central e oferece uma base sólida para novas investigações, destacando que avanços na aplicação dessa abordagem podem ser fundamentais para o crescimento sustentável e competitivo das organizações no cenário contemporâneo.

REFERÊNCIAS

- ACHARYA, A. et al. Big data, knowledge co-creation and decision making in fashion industry. **International Journal of Information Management**, v. 42, p. 90–101, out. 2018.
- AAKER, D.; KUMAR, V.; DAY, G. **Marketing research**. 5. ed. New York: John Wiley & Sons, 1995.
- ALVAREZ, F. J.; SANCHEZ, F. S.; BELLIDO, R. Demographic and dwelling models by artificial intelligence: urban renewal opportunities in Spanish coast. **International Journal of Sustainable Development and Planning**, v. 13, n. 07, p. 941–953, 30 set. 2018.
- AMARA, B. S.; BARROS, J. C. **Pesquisa de marketing: conceitos e metodologia**. 4. ed. São Paulo: Pearson, 2007. 273 p.
- AMAZON. **O que é estratégia de dados?**. 2024. Disponível em: <https://aws.amazon.com/pt/what-is/data-strategy/>. Acesso em: 6 jun. 2024.
- ALMQUIST, A. **How to transform to a data-driven organization?**. 2024. Disponível em: www.itera.com/datadriven-2021. Acesso em: 8 jul. 2024.
- AKTER, S. et al. How to improve firm performance using big data analytics capability and business strategy alignment? **International Journal of Production Economics**, v. 182, p. 113–131, dez. 2016.
- AWAN, U. et al. Big data analytics capability and decision-making: The role of data-driven insight on circular economy performance. **Technological Forecasting and Social Change**, v. 168, p. 120766, jul. 2021.
- BANGE, C; GROSSER, T. **BARC Data Culture Survey 23: How to liberalize data access to empower data users**. 2023. Disponível em: <https://barc.com/research/barc-data-culture-survey-23/>. Acesso em: 15 mai. 2024.
- BANSAL, S. K.; KAGEMANN, S. Integrating Big Data: A Semantic Extract-Transform-Load Framework. **Computer**, v. 48, n. 3, p. 42–50, mar. 2015.
- BARC, **What is data communication?**. 2024. Disponível em: <https://barc.com/de/glossar/data-analytics/data-communication/>. Acesso em: 28 Jun. 2024.
- BERNARDO, M. R. **Os agentes de software inteligentes no processo de tomada de decisão**. 2006. Tese (Doutorado em Gestão) – Instituto Superior de Economia e Gestão, Universidade Técnica de Lisboa, Lisboa, 2006.
- BERISHA, B.; MĚZIU, E.; SHABANI, I. Big data analytics in Cloud computing: an overview. **Journal of Cloud Computing**, v. 11, n. 1, p. 24, 6 ago. 2022.

CAPPA, F. et al. Big Data for Creating and Capturing Value in the Digitalized Environment: Unpacking the Effects of Volume, Variety, and Veracity on Firm Performance*. **Journal of Product Innovation Management**, v. 38, n. 1, p. 49–67, 29 jan. 2021.

CHAUHAN, P.; SOOD, M. Big Data: Present and Future. **Computer**, v. 54, n. 4, p. 59–65, abr. 2021.

CERTO, S. **Tomada de decisões**. Em: Administração moderna. 9. ed. São Paulo: Pearson, 2005. Capítulo 7, p. 123–145.

CONFEA. **Resolução nº 235, de 9 de outubro de 1975**. Disponível em: <https://normativos.confea.org.br/Ementas/>. Acesso em: 28 mar. 2024.

COTE, C. **What is diagnostic analytics? 4 examples**. 2021. Disponível em: <https://online.hbs.edu/blog/post/types-of-data-analysis>. Acesso em: 14 abril. 2024.

COX, M.; ELLSWORTH, D. **Application-controlled demand paging for out-of-core visualization**. Proceedings. Visualization '97 (Cat. No. 97CB36155). **Anais...IEEE**, [s.d.].

DEVORE, J. L. **Probabilidade e estatística para engenharia e ciências**. 1. ed. São Paulo: Cengage, 2019. 656 p. ISBN 978-85-221-2803-7.

DE SOUSA, I. D.; ALTURAS, B. **Herbert Simon's Decision-Making Model with a Meta-Decision and its Impact on Business Intelligence Systems**. 2019 14th Iberian Conference on Information Systems and Technologies (CISTI). **Anais...IEEE**, jun. 2019.

DWIVEDI, Y. K. et al. Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. **International Journal of Information Management**, v. 57, p. 101994, abr. 2021.

FONSECA, J. J. S. **Metodologia da pesquisa científica**. Fortaleza: UEC, 2002.

FOX, J. The R Commander: A Basic Statistics Graphical User Interface to R. **Journal of Statistical Software**, v. 14, n. 9, p. 1–42, 2005. Disponível em: <https://www.jstatsoft.org/article/view/v014i09>. Acesso em: 16 nov. 2024.

WAMBA, S.; MISHRA, D. Big data integration with business processes: a literature review. **Business Process Management Journal**, v. 23, n. 3, p. 477–492, 5 jun. 2017.

GAO, J.; SARWAR, Z. How do firms create business value and dynamic capabilities by leveraging big data analytics management capability? **Information Technology and Management**, 16 out. 2022.

GIL, A. C. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2007

GIL, A. C. **Como elaborar projetos de pesquisa**. 5. ed. São Paulo: Atlas, 2010. 184 p..

GOOGLE. **O que é governança de dados?**. 2024. Disponível em: <https://cloud.google.com/learn/what-is-data-governance?>. Acesso em: 11 jul. 2024.

GRANT, R. M. Toward a knowledge-based theory of the firm. **Strategic Management Journal**, v. 17, n. S2, p. 109–122, 20 dez. 1996.

GUALTIERI, M. **Hadoop Is Data's Darling For A Reason**. Forrester, 16 de janeiro de 2016. Disponível em: <https://www.forrester.com/blogs/hadoop-is-datas-darling-for-a-reason/>. Acesso em: 20 de mar. de 2024.

GUNTER, H. **Como elaborar um questionário**. (Série: Planejamento de Pesquisa nas Ciências Sociais, Nº 1). Brasília, DF: UnB, Laboratório de Psicologia Ambiental, 2003. Disponível em: <http://beco-do-bosque.net/XTextos/01Questionario.pdf>. Acesso em: 29 jun. 2013.

GUPTA, S. Scalable Efficient Big Data Pipeline Architecture. ML4DEVs, 2024. Disponível em: <https://www.ml4devs.com/articles/scalable-efficient-big-data-analytics-machine-learning-pipeline-architecture-on-cloud>. Acesso em: 06 mai. 2024.

HIGGINS, T. P. et al. Flat choroidal melanoma masquerading as central serous chorioretinopathy. **Oman journal of ophthalmology**, v. 9, n. 3, p. 174–176, 2016.

INFINITSCALE. **Data-driven success in scaling an online business**. 2024. Disponível em: <https://infinityscale.com/data-driven-success-in-scaling-an-online-business/>. Acesso em: 8 jul. 2024.

KISH, L. **Survey sampling**. New York: John Wiley & Sons, Inc., 1965.

MCDANIEL JR., C. D.; GATES, R. **Fundamentos de pesquisa de marketing**. 4. ed. Rio de Janeiro: LTC, 2005. 312 p.

MALHOTRA, N. K. **Pesquisa de marketing: uma orientação aplicada**. 6. ed. Porto Alegre: Bookman, 2012. 735 p.

MANZONI, L. **Netflix: Faturamento, Lucro e Curiosidades**. Disponível em: <https://br.investing.com/academy/statistics/faturamento-netflix/>. Acesso em: 7 jul. 2024.

MATOS, B. M. R. **Os agentes de software inteligentes no processo de tomada de decisão**. 2006. Tese (Doutorado em Gestão) – Instituto Superior de Economia e Gestão, Universidade Técnica de Lisboa, Lisboa, 2006.

MATTAR, F. N. **Pesquisa de marketing: metodologia, planejamento, execução e análise**. 2. ed., v. 2. São Paulo: Atlas, 1994.

MATTAR, F. **Pesquisa de marketing**. São Paulo: Atlas, 1996.

MATTAR, F. N. **Pesquisa de marketing**. 3. ed. compacta. São Paulo: Atlas, 2001. 275 p.

MAXIMIANO, A. C. A. Introdução à administração. Ed. **compacta**. São Paulo: Atlas, 2009. 294 p.

MICROSOFT. **Relational vs. NoSQL data**. 2022. Disponível em: <https://learn.microsoft.com/en-us/dotnet/architecture/cloud-native/relational-vs-nosql-data>. Acesso em: 6 jun. 2024.

MINTZBERG, H.; RAISINGHANI, D.; THEORET, A. The structure of “unstructured” decision processes. **Administrative science quarterly**, v. 21, n. 2, p. 246, 1976.

MIRARAB, A.; MIRTAHERI, S. L.; ASGHARI, S. A. Value creation with big data analytics for enterprises: a survey. **TELKOMNIKA (Telecommunication Computing Electronics and Control)**, v. 17, n. 6, p. 2790, 1 dez. 2019.

MISRA, N. N. et al. IoT, Big Data, and Artificial Intelligence in Agriculture and Food Industry. **IEEE Internet of Things Journal**, v. 9, n. 9, p. 6305–6324, 1 maio 2022.

MONTGOMERY, D. C.; RUNGER, G. C. **Applied statistics and probability for engineers**. 7. ed. Nashville, TN, USA: John Wiley & Sons, 2018.

MYLER, L. **Better data quality equals higher marketing ROI**. Disponível em: <https://www.forbes.com/sites/larrymyler/2017/07/11/better-data-quality-equals-higher-marketing-roi>. Acesso em: 13 abr. 2024.

NASCIMENTO, P. T.; LIMA, A. C.; RUSSO, R. F.; SOUSA, W. **Tomada de decisão nas organizações**. Colaboração de YU, A. S. O. São Paulo: Saraiva, 2011.

NETFLIX. **Research areas analytics: Driving insights from data**. 2024. Disponível em: <https://br.investing.com/academy/statistics/faturamento-netflix/>. Acesso em: 7 jul. 2024.

NUTT, P. C. Types of Organizational Decision Processes. **Administrative Science Quarterly**, v. 29, n. 3, p. 414, set. 1984.

OUSSOUS, A. et al. Big Data technologies: A survey. **Journal of King Saud University - Computer and Information Sciences**, v. 30, n. 4, p. 431–448, out. 2018.

PARADZA, D.; DARAMOLA, O. Business Intelligence and Business Value in Organisations: A Systematic Literature Review. **Sustainability**, v. 13, n. 20, p. 11382, 15 out. 2021.

PEREIRA, J. C. R. **Análise de dados qualitativos: estratégias metodológicas para as ciências da saúde, humanas e sociais**. 3. ed. São Paulo: EDUSP, 2000.

PUCPR. **Do dado à decisão: o que é e como construir uma cultura data-driven**. 2024. Disponível em: <https://posdigital.pucpr.br/blog/cultura-data-driven>. Acesso em: 6 jul. 2024.

RANJAN, J.; FOROPON, C. Big Data Analytics in Building the Competitive Intelligence of Organizations. **International Journal of Information Management**, v. 56, p. 102231, fev. 2021.

REHMAN, M. H.-U.; CHANG, V.; BATOOL, A.; WAH, T. Y. Big data reduction framework for value creation in sustainable enterprises. **International Journal of Information Management**, v. 36, n. 6, p. 917–928, 2016.

RIALTI, R. et al. Big data and dynamic capabilities: a bibliometric analysis and systematic literature review. **Management Decision**, v. 57, n. 8, p. 2052–2068, 12 set. 2019.

SANDHU, A. K. Big data with cloud computing: Discussions and challenges. **Big Data Mining and Analytics**, v. 5, n. 1, p. 32–40, mar. 2022.

SANTOS, A. R. **Metodologia científica: a construção do conhecimento**. Rio de Janeiro: DP&A, 1999.

SAMARA, B. S.; BARROS, J. C. **Pesquisa de marketing: conceitos e metodologia**. 4. ed. São Paulo: Pearson, 2007. 273 p.

SHAH, B. et al. **Security Issues and Recommendations for the Lifecycle of Data in Cloud Computing**. Department of Computer Engineering and Information Technology, 2010.

SHARDA, R.; DELEN, D.; TURBAN, E. **Business Intelligence e Análise de Dados para Gestão do Negócio**. 4^a ed. Porto Alegre: Bookman, 2019. 584 p.

SHIMIZU, T. **Decisão nas organizações: introdução aos problemas de decisão encontrados nas organizações e nos sistemas de apoio à decisão**. São Paulo: Atlas, 2001. 313 p.

SIMON, H. A. **Comportamento administrativo: estudo dos processos decisórios nas organizações administrativas**. 2. ed. Rio de Janeiro: Fundação Getúlio Vargas, 1965.

SIMON, H. A. **The new science of management decision**. Old Tappan, NJ, USA: Prentice Hall, 1977.

TABLEAU. **A Guide To Data Driven Decision Making: What It Is, Its Importance, & How To Implement It**. 2024. Disponível em:

<https://www.tableau.com/learn/articles/data-driven-decision-making>. Acesso em: 8 jul. 2024.

TABLEAU. **Data culture drives data and AI success**. 2024. Disponível em: <https://www.tableau.com/why-tableau/data-culture>. Acesso em: 15 jul. 2024.

TARGIT. **5 benefícios do Analytics Self-Service**. 2024. Disponível em: <https://targitbrasil.com/5-beneficios-do-analytics-self-service/>. Acesso em: 17 jul. 2024.

TERA. **Liderança de Dados: definindo a expertise que sua organização precisa**. 2024. Disponível em: <https://medium.com/somos-tera/lideran%C3%A7a-de-dados-definindo-a-expertise-que-e-sua-organiza%C3%A7%C3%A3o-precisa-b1dcce17f59>. Acesso em: 8 jul. 2024.

THOMAS H. DAVENPORT; RANDY BEAN. **Data and AI Leadership Executive Survey 2022**. Boston: [s.n.].

VENABLES, W. N.; SMITH, M. **An introduction to R**. 2014. Disponível em: <http://cran.r-project.org/doc/manuals/R-intro.pdf>. Acesso em: 17 out. 2024.

WANG, J. et al. Big Data Service Architecture: A Survey. **School of Computer & Communication Engineering**, v. 21, 2020.

ZAFAR, S. Utilizing big data for health care automation: Obligations, fitness and challenges. Em: **Intelligent Systems Reference Library**. Cham: Springer International Publishing, 2019. p. 263–278.

APÊNDICE

APÊNDICE A

Questionário Aplicado - Geração de Valor com Big Data: Uma pesquisa sobre a cultura de dados dentro das empresas.



Geração de Valor com Big Data: Uma pesquisa sobre a cultura de dados dentro das empresas.

O objetivo deste questionário é fazer um levantamento do cenário atual de como as grandes organizações estão lidando com o Big Data e a adoção dos conceitos "cultura de dados" e "data driven", termos cada vez mais empregados para descrever a utilização estratégica de dados na tomada de decisões.

As respostas fornecidas serão tratadas com total confidencialidade e os dados coletados não serão compartilhados com terceiros. As informações aqui coletadas serão utilizadas exclusivamente para fins acadêmicos.

1. O processo de tomada de decisão na sua empresa baseia-se em dados ou na intuição?

- Predominantemente baseado na intuição
- Baseado em dados e na intuição
- Predominantemente baseado em dados e análises

2. Sob a perspectiva do alto volume, variedade e velocidade dos dados no contexto do big data, escolha a opção que melhor representa a percepção da sua empresa.

- Não conseguimos assimilar o bigdata e nem extrair valor dos dados.
- Temos dificuldade de assimilar o bigdata mas conseguimos extrair valor dos dados.
- Não temos dificuldade de assimilar o bigdata e conseguimos extrair valor dos dados.

3. Como você classifica o nível de maturidade técnica e analítica da sua organização para lidar com big data?

Leve em consideração a padronização e qualidade dos estágios principais de uma arquitetura de dados: coleta, processamento, armazenamento e análise de dados.

- Muito baixo
- Baixo
- Mediano
- Alto
- Muito alto

4. Até que ponto você concorda com as afirmações a seguir sobre a cultura de dados na sua empresa?

"Entendemos que o conhecimento de dados é fundamental para tomada de decisão. Dessa forma, investimos ou planejamos investir na coleta, vinculação e análise de dados."

- Discordo totalmente
- Discordo em partes
- Concordo em partes
- Concordo totalmente

"Dados e informações são considerados um ativo para nós"

- Discordo totalmente
- Discordo em partes
- Concordo em partes
- Concordo totalmente

"As decisões a nível operacional baseiam-se predominantemente em dados"

- Discordo totalmente
- Discordo em partes
- Concordo em partes
- Concordo totalmente

"Os dados são fundamentais para gerar receita"

- Discordo totalmente
- Discordo em partes
- Concordo em partes
- Concordo totalmente

"As ferramentas de análise de dados são claramente identificadas e estão disponíveis para toda a empresa"

- Discordo totalmente
- Discordo em partes
- Concordo em partes
- Concordo totalmente

"Conseguimos gerar valor mesmo quando temos um conjunto de dados grande, variado e complexo para analisar"

- Discordo totalmente
- Discordo em partes
- Concordo em partes
- Concordo totalmente

"As decisões a nível executivo baseiam-se predominantemente em dados"

- Discordo totalmente
- Discordo em partes
- Concordo em partes
- Concordo totalmente

5. Quais departamentos da sua empresa mais utilizam dados em suas tomadas de decisão e processos operacionais?

- Finanças
- Vendas
- Marketing
- Produção
- Logística
- Cadeia de suprimentos
- Pesquisa e inovação
- Recursos humanos
- Tecnologia da informação

6. Quais são os maiores desafios na geração de valor com os dados na sua empresa?

- Tornar as fontes de dados conhecidas e centralizadas em um data warehouse
- Investir em capacitações que visam especializar as equipes
- Simplificar o acesso e a preparação de dados para analistas de negócios
- Automatizar a coleta de dados dos sistemas internos e tornar viável a sua combinação
- Investir em iniciativas que visam melhorar a qualidade dos processos voltados a dados
- Promover o engajamento entre as áreas fornecedoras e consumidoras de dados
- Normalizar dados não estruturados e disponibilizá-los em um formato acessível
- Construir uma infraestrutura técnica viável para se trabalhar com dados
- Formar um time tecnicamente qualificado para lidar com big data

7. Quais benefícios sua empresa já obteve através do uso de dados? *

- Melhor tomada de decisão
- Melhoria contínua de processos
- Redução de custos
- Crescimento da receita
- Maior vantagem competitiva
- Melhor resolução de problemas
- Inovação de produtos ou serviços

8. A seguir, indique o nível de importância que sua empresa atualmente atribui às seguintes frentes de cultura de dados.

Liderança de dados

- Sem importância
- Pouco importante
- Razoável
- Importante
- Muito importante

Estratégia de dados

- Sem importância
- Pouco importante
- Razoável
- Importante
- Muito importante

Governança de dados

- Sem importância
- Pouco importante
- Razoável
- Importante
- Muito importante

Alfabetização em dados

- Sem importância
- Pouco importante
- Razoável
- Importante
- Muito importante

Democratização de acesso aos dados

- Sem importância
- Pouco importante
- Razoável
- Importante
- Muito importante

Comunicação com dados

- Sem importância
- Pouco importante
- Razoável
- Importante

9. Quais são ou foram os principais obstáculos à implementação ou estabelecimento de uma cultura de dados na sua empresa?

- A alocação de recursos é geralmente insuficiente para fortalecer a área de dados
- Falta de conhecimento entre os colaboradores sobre dados e análises (alfabetização em dados)
- Funções e responsabilidades não definidas nas áreas de dados
- Falta de comunicação interna sobre dados e análises de dados
- Nossa estratégia de dados é muito mais uma estratégia tecnológica e ferramental
- Dificuldade para implementar um modelo de governança de dados por questões de "compliance"
- Atualmente a gerência não vê retorno financeiro direto que justifique essa implementação

10. Quais aspectos tecnológicos e abordagens são comumente utilizados para lidar com dados na rotina da sua empresa?

- Data warehouse
- Business intelligence
- Planilhas (Excel)
- Self-service analytics
- Data lake
- Código (ex: Python)
- ETL
- IA/advanced analytics
- Ferramentas de inteligência de dados
- Dashboards e Visualização de dados
- Data lakehouse
- Ciência de dados
- Engenharia de dados
- Ambiente em nuvem (cloud)

