

# **Introdução à Ciência de Dados para o Bacharelado em Administração**

Washington Santos da Silva

19 de janeiro de 2026

# Índice

<b>Prefácio</b>	<b>4</b>
<b>1 Introdução</b>	<b>6</b>
<b>2 Visão Geral de Ciência de Dados</b>	<b>7</b>
2.1 Introdução e contextualização . . . . .	7
2.2 A economia de dados: um breve histórico . . . . .	7
2.3 Big Data e a explosão do volume de dados . . . . .	8
2.4 O que é Ciência de Dados? . . . . .	11
2.5 Metodologias e processos: CRISP-DM . . . . .	11
2.6 Papéis profissionais na área de dados . . . . .	13
Analista de Dados . . . . .	13
Cientista de Dados (Iniciante) . . . . .	14
Engenheiro de Dados . . . . .	15
2.7 Áreas de aplicação . . . . .	15
2.8 Habilidades interpessoais e analíticas . . . . .	15
<b>3 Metodologia CRISP-DM</b>	<b>17</b>
3.1 O problema do “código antes do plano” . . . . .	17
3.2 O papel do administrador em projetos de dados . . . . .	17
3.3 Ferramentas são meios, não fins . . . . .	17
3.4 CRISP-DM como estrutura orientadora . . . . .	18
3.5 Fase 1: Compreensão do Negócio ( <i>Business Understanding</i> ) . . . . .	18
3.5.1 Checklist da Fase 1 . . . . .	19
3.6 Fase 2: Compreensão dos Dados ( <i>Data Understanding</i> ) . . . . .	19
3.6.1 Checklist da Fase 2 . . . . .	19
3.7 Fase 3: Preparação dos Dados ( <i>Data Preparation</i> ) . . . . .	19
3.7.1 Checklist da Fase 3 . . . . .	20
3.8 Fase 4: Modelagem ( <i>Modeling</i> ) . . . . .	20
3.8.1 Checklist da Fase 4 . . . . .	20
3.9 Fase 5: Avaliação ( <i>Evaluation</i> ) . . . . .	20
3.9.1 Checklist da Fase 5 . . . . .	20
3.10 Fase 6: Implementação ( <i>Deployment</i> ) . . . . .	21
3.10.1 Checklist da Fase 6 . . . . .	21

<b>4 Tipos de análise e Pertuntas de Negócio</b>	<b>22</b>
4.1 Visão geral dos tipos de análise . . . . .	22
4.2 Análise descritiva . . . . .	22
4.3 Análise diagnóstica . . . . .	23
4.4 Análise preditiva . . . . .	23
4.5 Análise prescritiva . . . . .	24
<b>Referências</b>	<b>25</b>

# Prefácio

Este livro resulta da experiência acumulada nos anos de 2024 e 2025 na oferta da disciplina *Introdução à Ciência de Dados* para o curso noturno de Bacharelado em Administração do IFMG – Campus Formiga. Ele organiza e aprofunda os materiais utilizados em sala de aula, com o objetivo de oferecer aos estudantes um recurso didático progressivo, acessível e alinhado à realidade do ensino público brasileiro.

O livro acompanha a estrutura da disciplina e funciona tanto como material de apoio às aulas quanto como um registro sistematizado dos conteúdos trabalhados ao longo do curso. Ao longo dos capítulos, o leitor encontrará referências a materiais complementares, ao site da disciplina e à recursos online, que podem ser utilizadas para estudo autônomo após o término das aulas.

Embora pensado prioritariamente para os alunos do curso de Administração do IFMG – Campus Formiga, o texto foi escrito de modo a poder ser útil a estudantes de Administração e áreas afins em outras instituições, especialmente àqueles que estão iniciando o estudo de disciplinas que envolvam análise de dados e estatística. Trata-se, portanto, de um material aberto e em desenvolvimento constante.

O leitor a quem este livro se dirige costuma ter algum contato prévio com programação, geralmente por meio de disciplinas introdutórias, mas frequentemente apresenta dificuldades com lógica, matemática básica e conceitos estatísticos. Além disso, é comum a dependência quase exclusiva de planilhas eletrônicas, o que limita a reproduzibilidade, a transparência e a escalabilidade das análises. Este livro parte dessa realidade concreta e não pressupõe fluência prévia em programação ou estatística.

A proposta é deliberada: as dificuldades inerentes à Ciência de Dados não são evitadas, mas introduzidas de forma gradual e contextualizada. Conceitos, técnicas e ferramentas são apresentados passo a passo, sempre que possível ancorados em exemplos ligados à área de Administração. O objetivo não é ensinar apenas comandos, mas contribuir para a formação de um modo de pensar analítico, crítico e orientado por dados.

A linguagem de programação adotada é **R**, escolhida pelo fato de considerá-la a melhor linguagem de programação especializada em análise de dados. O livro também introduz noções básicas de **SQL**, reconhecendo sua importância no trabalho com bases de dados reais. Ferramentas como **RStudio**, **Git** e **GitHub**, bem como o uso do **terminal**, são apresentadas como parte de um ambiente de trabalho coerente, voltado à organização, documentação e controle de versões. O sistema **Quarto** é utilizado extensivamente, reforçando a importância

da reproduzibilidade, transparência e importância de documentação dos procedimentos de análise.

Este projeto também responde a um grave problema estrutural em minha visão: a escassez de referências atualizadas em português na área de Ciência de Dados e dificuldade de acesso a livros importados em instituições públicas. Nesse contexto, a produção de materiais abertos e adaptados ao contexto local torna-se necessária.

**Washington Santos da Silva**

Professor do IFMG – Campus Formiga

# **1 Introdução**

TODO.

## 2 Visão Geral de Ciência de Dados

### 2.1 Introdução e contextualização

Nas últimas décadas, dados passaram a ocupar um papel central nas decisões econômicas, organizacionais e governamentais. Atividades cotidianas como compras online, interações em redes sociais ou o uso do GPS em smartphones geram continuamente grandes volumes de dados, que alimentam o que hoje se convencionou chamar de *economia de dados*.

Esse fenômeno não surgiu de forma repentina. Ele é resultado de um processo histórico no qual a coleta, o armazenamento e a análise de dados foram progressivamente incorporados às práticas de gestão, pesquisa e tomada de decisão. Compreender essa trajetória é fundamental para entender por que a Ciência de Dados se tornou uma área central no mundo contemporâneo.

### 2.2 A economia de dados: um breve histórico

As raízes da economia de dados remontam ao século XIX, quando jornais norte-americanos passaram a coletar informações sistemáticas sobre seus leitores e a realizar levantamentos para antecipar resultados eleitorais. Já nesse período inicial, dados eram utilizados como instrumento para reduzir incertezas e orientar decisões editoriais e comerciais, ainda que de forma incipiente e pouco padronizada (Harkness, 2021a).

No início do século XX, com a consolidação do marketing como área organizacional, empresas e pesquisadores passaram a estruturar departamentos dedicados ao estudo sistemático do comportamento do consumidor. A coleta de dados deixou de ser episódica e passou a integrar processos contínuos de análise, voltados à compreensão de mercados, preferências e padrões de consumo (Harkness, 2021a).

A partir da década de 1930, a introdução de métodos estatísticos de amostragem, notadamente os trabalhos associados a George Gallup, marcou uma mudança qualitativa importante. Previsões baseadas em dados passaram a apoiar-se em fundamentos estatísticos mais sólidos, substituindo abordagens baseadas em grandes volumes de respostas não controladas por técnicas cientificamente mais rigorosas (Harkness, 2021b).

Entre as décadas de 1950 e 1980, empresas como a Nielsen consolidaram sistemas de observação contínua de hábitos de consumo e audiência. Esse período reforçou uma distinção central na economia de dados: observar comportamentos reais, de forma sistemática, frequentemente

produz informações mais confiáveis do que simplesmente perguntar aos indivíduos sobre suas intenções ou opiniões Harkness (2021c).

Com a digitalização da economia a partir dos anos 1990, o uso de códigos de barras, programas de fidelidade e, posteriormente, plataformas digitais e redes sociais transformou os dados em um ativo estratégico de escala global. A capacidade de coletar, armazenar e analisar grandes volumes de informações passou a redefinir modelos de negócio e estruturas competitivas em diversos setores (Harkness, 2021c).

Mais recentemente, o avanço de modelos de inteligência artificial, como os grandes modelos de linguagem, representa o estágio mais sofisticado dessa trajetória. Esses sistemas dependem fortemente de grandes volumes de dados para seu treinamento e funcionamento, evidenciando que, na economia contemporânea, dados não são apenas um subproduto das atividades organizacionais, mas um recurso central que condiciona inovação, eficiência e poder econômico (Harkness, 2021d).

### **2.3 Big Data e a explosão do volume de dados**

O crescimento exponencial da produção de dados levou ao surgimento do conceito de *Big Data*, frequentemente caracterizado pelos chamados *cinco Vs*: volume, velocidade, variedade, veracidade e valor. Esses atributos ajudam a compreender não apenas a quantidade de dados gerados, mas também os desafios associados ao seu processamento e uso.

Esse aumento no volume e na complexidade dos dados tornou inviável o uso exclusivo de ferramentas tradicionais, como planilhas eletrônicas, e criou a necessidade de métodos, linguagens e infraestruturas mais robustas para análise. Não por acaso, publicações de referência na área de negócios passaram a destacar dados como um dos recursos mais valiosos da economia contemporânea.

Um marco simbólico importante na consolidação da chamada economia de dados foi a capa da revista *The Economist*, publicada em 2017, cujo título era *The World's Most Valuable Resource* (“O recurso mais valioso do mundo”) e cujo subtítulo anuncia *Data and the new rules of competition* (“Dados e as novas regras da competição”). A mensagem central não era apenas a comparação entre dados e petróleo, mas a ideia de que os dados haviam se tornado o principal recurso estratégico capaz de redefinir a dinâmica competitiva entre empresas e setores.

Diferentemente de recursos tradicionais, os dados não geram valor de forma automática. Seu valor emerge da capacidade de coletá-los, organizá-los, analisá-los e, sobretudo, utilizá-los de maneira sistemática na tomada de decisão. Nesse sentido, a capa da *The Economist* aponta para uma mudança mais profunda: empresas competitivas passam a ser aquelas capazes de transformar dados em conhecimento operacional e vantagem estratégica contínua.

Essa mudança implica novas regras de competição. Escala, velocidade de análise, capacidade de experimentação e aprendizado contínuo tornam-se fatores centrais. Organizações que



Figura 2.1: Os cinco Vs do Big Data.

dominam esses elementos conseguem adaptar produtos, processos e estratégias com maior rapidez, enquanto aquelas que tratam dados apenas como subprodutos operacionais tendem a perder relevância.

Ao destacar os dados como o recurso mais valioso da economia contemporânea, a revista reforça a necessidade de métodos, ferramentas e competências voltadas não apenas à análise técnica, mas à integração entre dados, estratégia e decisão. É nesse contexto que a Ciência de Dados se consolida como área essencial para a Administração, indo além do uso de tecnologias específicas e passando a influenciar diretamente a forma como as organizações competem.



Figura 2.2: Capa da revista *The Economist* (2017): The World's Most Valuable Resource — Data and the new rules of competition.

## 2.4 O que é Ciência de Dados?

Diante desse contexto, surge a Ciência de Dados como uma abordagem estruturada para extrair significado e valor de grandes volumes de dados. Embora o termo possa parecer intimidador, sua ideia central é relativamente simples: usar métodos analíticos e computacionais para transformar dados brutos em informações úteis para a tomada de decisão.

De forma sintética, a Ciência de Dados integra conhecimentos de estatística, computação e domínio do negócio, combinando técnicas dessas áreas para lidar com problemas reais.

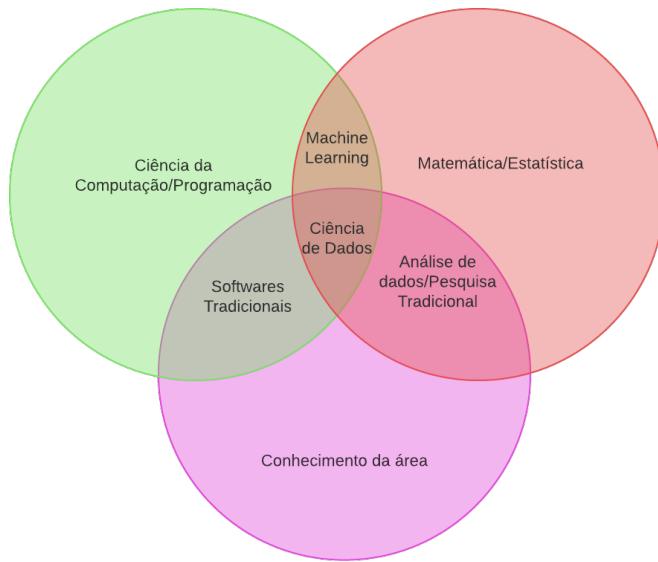


Figura 2.3: A Ciência de Dados como interseção entre estatística, computação e conhecimento do domínio.

Essa característica interdisciplinar explica tanto o potencial da área quanto a diversidade de formações presentes em equipes de dados.

## 2.5 Metodologias e processos: CRISP-DM

A prática da Ciência de Dados costuma ser organizada por meio de metodologias de projeto. Uma das mais difundidas é o CRISP-DM (*Cross Industry Standard Process for Data Mining*), que estrutura o trabalho em etapas iterativas, desde o entendimento do problema de negócio até a implantação de soluções baseadas em dados.

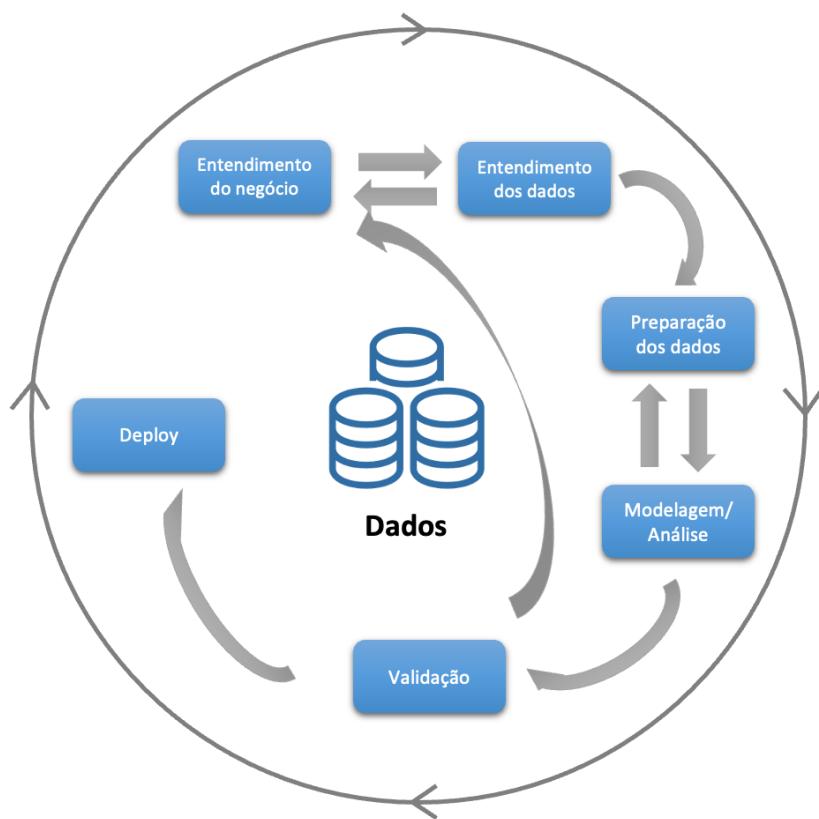


Figura 2.4: Etapas do processo CRISP-DM aplicadas a projetos de Ciência de Dados.

Esse modelo destaca que projetos de dados raramente seguem um caminho linear. É comum que análises retornem a etapas anteriores à medida que novos insights surgem ou que hipóteses iniciais precisem ser revistas.

## 2.6 Papéis profissionais na área de dados

O crescimento da área de dados deu origem a diferentes papéis profissionais, que variam conforme o foco técnico, o grau de especialização e a posição no fluxo de produção de valor a partir dos dados. Entre os papéis mais comuns estão o Analista de Dados, o Cientista de Dados e o Engenheiro de Dados.

Embora essas funções sejam conceitualmente distintas, na prática, especialmente em organizações de pequeno e médio porte, é comum que um mesmo profissional acumule responsabilidades associadas a mais de um papel. Essa sobreposição é particularmente relevante no contexto brasileiro e deve ser levada em conta ao interpretar descrições formais de cargos.

Nas subseções a seguir, apresentam-se esses papéis de forma sintética, destacando habilidades técnicas, competências analíticas e exemplos típicos de aplicação.

### Analista de Dados

O Analista de Dados atua principalmente na exploração, organização e interpretação de dados, com foco em apoiar decisões operacionais e táticas por meio de análises descritivas e diagnósticas.

#### Habilidades técnicas

- Domínio de ferramentas de visualização de dados, como Power BI ou Tableau, e conhecimento avançado em planilhas eletrônicas, como o MS Excel.
- Proficiência em linguagens de programação voltadas à análise de dados, especialmente R e/ou Python.
- Experiência prática, ainda que em nível introdutório, com bancos de dados e linguagem SQL.

#### Habilidades analíticas

- Capacidade de realizar análises estatísticas básicas e interpretar grandes volumes de dados para identificar padrões, tendências e anomalias.

#### Exemplos de aplicações práticas

- Análise de dados em setores como finanças, saúde e turismo, com o objetivo de gerar insights para melhoria de processos, gestão de recursos e atendimento ao cliente.

- Elaboração de análises **descritivas** e **diagnósticas**: a análise descritiva busca compreender *o que* ocorreu, enquanto a análise diagnóstica procura explicar *por que* ocorreu.

## Cientista de Dados (Iniciante)

O Cientista de Dados atua de forma mais aprofundada na modelagem e na construção de soluções analíticas, combinando estatística, programação e conhecimento do problema de negócios. No nível iniciante, esse papel ainda envolve forte aprendizado técnico e conceitual.

### Habilidades técnicas

- Proficiência em linguagens de programação como R e/ou Python, com uso de bibliotecas especializadas (por exemplo, tidyverse, pandas, scikit-learn, tidymodels).
- Conhecimento em SQL e em bancos de dados relacionais e não relacionais.
- Familiaridade com ferramentas de versionamento de código (Git) e ambientes de desenvolvimento.

### Habilidades analíticas

- Conhecimentos em estatística aplicada e aprendizagem de máquina.
- Capacidade de preparar, transformar e organizar conjuntos de dados para análise e modelagem.
- Capacidade de implementar algoritmos básicos de aprendizagem de máquina sob supervisão ou em projetos de escopo limitado.

### Exemplos de aplicações práticas

- Desenvolvimento de modelos de classificação e regressão para problemas como previsão, segmentação de clientes e detecção de anomalias.
- Realização de análises **preditivas** e **prescritivas**, utilizando dados históricos para antecipar comportamentos e apoiar recomendações de ação.
- Criação de provas de conceito (POCs) para validação de hipóteses de negócio baseadas em dados.
- Comunicação de resultados técnicos em formato acessível a públicos não técnicos.

## **Engenheiro de Dados**

O Engenheiro de Dados é o profissional responsável pela construção e manutenção da infraestrutura que permite o armazenamento, o processamento e o acesso eficiente aos dados utilizados pelas equipes analíticas.

- Atua no projeto, desenvolvimento e otimização de *pipelines* de dados, *data warehouses* e *data lakes*.
- Seu foco principal é garantir que os dados sejam confiáveis, consistentes, acessíveis e escaláveis, servindo como base para o trabalho de analistas e cientistas de dados.
- Em geral, esse papel envolve menor ênfase em análises estatísticas e maior concentração em aspectos de arquitetura, desempenho e integração de sistemas.

## **2.7 Áreas de aplicação**

As aplicações da Ciência de Dados são amplas e afetam diretamente o cotidiano das organizações. Em finanças, destacam-se análises de risco de crédito, detecção de fraudes e gestão de investimentos. Em marketing, técnicas de segmentação de clientes, análise de sentimentos e monitoramento de mídias sociais são amplamente utilizadas.

Esses exemplos ilustram como dados podem ser usados para compreender o passado, explicar causas e antecipar cenários futuros, correspondendo às análises descritivas, diagnósticas e preditivas.

## **2.8 Habilidades interpessoais e analíticas**

Além de competências técnicas, profissionais de dados precisam desenvolver habilidades interpessoais e analíticas que permitam transformar resultados quantitativos em decisões organizacionais concretas. Essas habilidades são essenciais para conectar análises de dados a problemas reais de negócio e para comunicar resultados de forma eficaz a diferentes públicos.

- **Pensamento analítico:** Abordar problemas de forma estruturada, formular perguntas relevantes, selecionar informações apropriadas e buscar soluções baseadas em evidências.
- **Conhecimento do negócio:** Compreender os objetivos da organização, o contexto do mercado e a forma como análises de dados se relacionam com metas estratégicas e operacionais.
- **Comunicação oral e escrita:** Capacidade de explicar resultados técnicos a pessoas sem formação técnica, utilizando linguagem clara, exemplos práticos e evitando jargões desnecessários.

- **Pensamento crítico:** Questionar suposições implícitas, avaliar a qualidade e as limitações dos dados disponíveis e considerar interpretações alternativas antes de chegar a conclusões.
- **Contar histórias com dados:** Organizar resultados e indicadores de modo a construir uma narrativa coerente, capaz de destacar os principais insights e apoiar processos de tomada de decisão.
- **Trabalho em equipe:** Colaborar com profissionais de diferentes áreas, compreender necessidades diversas e integrar perspectivas técnicas e organizacionais.
- **Gerenciamento de projetos:** Planejar etapas de trabalho, definir prioridades, estabelecer prazos realistas e comunicar o andamento das atividades às partes interessadas.
- **Adaptabilidade:** Lidar com mudanças de requisitos, ferramentas e tecnologias em um campo caracterizado por rápida evolução.
- **Curiosidade intelectual:** Demonstrar interesse contínuo em aprender, formular novas perguntas e explorar dados de forma sistemática e responsável.

 Dica

Na prática profissional em Ciência de Dados, a capacidade de **estruturar problemas, compreender o contexto do negócio e comunicar resultados** de forma clara tende a ser mais determinante do que o domínio isolado de ferramentas. As habilidades técnicas adquirem valor quando integradas a essas competências centrais.

# 3 Metodologia CRISP-DM

Este capítulo apresenta a metodologia CRISP-DM (*Cross-Industry Standard Process for Data Mining*) (IBM Corporation, 2023), amplamente utilizada para estruturar projetos de Ciência de Dados. O objetivo é oferecer uma visão clara e operacional deste método, destacando o papel do administrador na condução e no alinhamento estratégico desses projetos.

## 3.1 O problema do “código antes do plano”

Projetos de dados frequentemente falham não por limitações técnicas, mas pela ausência de uma metodologia clara. É comum que equipes iniciem o desenvolvimento de modelos ou scripts antes mesmo de compreender o problema de negócio a ser resolvido.

Esse tipo de abordagem tende a gerar desalinhamento entre soluções técnicas e necessidades organizacionais, ciclos recorrentes de retrabalho, desperdício de tempo e recursos, além de dificuldades para escalar projetos piloto para soluções corporativas.

## 3.2 O papel do administrador em projetos de dados

Nesse contexto, o administrador desempenha papel central. Cabe a ele assegurar que projetos de dados tenham início a partir de objetivos de negócio claramente definidos, com critérios mensuráveis de sucesso estabelecidos antes da implementação técnica.

Além disso, o administrador atua como elo entre equipes técnicas e a organização, traduzindo necessidades estratégicas em perguntas analíticas e garantindo que os resultados produzidos sejam relevantes para a tomada de decisão.

## 3.3 Ferramentas são meios, não fins

Ferramentas como R, Python, SQL, Quarto e Git são essenciais para projetos de Ciência de Dados, mas não constituem um fim em si mesmas. Sem uma metodologia orientadora, mesmo códigos tecnicamente sofisticados podem resolver o problema errado.

O CRISP-DM fornece contexto e direção para o uso dessas ferramentas, enfatizando que a modelagem é apenas uma das etapas de um processo mais amplo, orientado por objetivos organizacionais.

### 3.4 CRISP-DM como estrutura orientadora

O CRISP-DM é uma metodologia consolidada para organizar projetos de dados de forma sistemática e iterativa. Diferentemente de abordagens centradas em ferramentas, o processo começa com a compreensão do negócio e termina com a implementação e avaliação dos resultados no contexto organizacional.

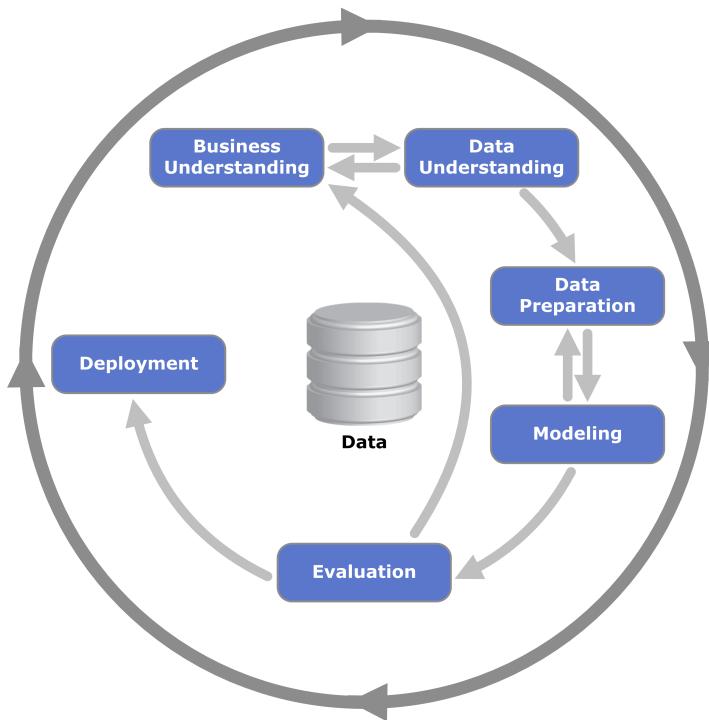


Figura 3.1: Fases da metodologia CRISP-DM e sua natureza iterativa.

As fases do CRISP-DM funcionam como pontos de verificação, permitindo avaliar o progresso, alinhar expectativas e decidir se o projeto deve avançar, ser ajustado ou interrompido.

### 3.5 Fase 1: Compreensão do Negócio (*Business Understanding*)

A primeira fase do CRISP-DM tem como foco alinhar a análise de dados aos objetivos empresariais. Antes de qualquer coleta ou modelagem, é necessário compreender claramente

qual problema se deseja resolver e por que ele é relevante.

Nessa etapa, objetivos de negócio são definidos, a situação atual é avaliada, metas analíticas são estabelecidas e um plano de projeto é elaborado.

### 3.5.1 Checklist da Fase 1

#	Tarefa	Resultados esperados
1.1	Determinar objetivos de negócio	Contexto, objetivos e critérios de sucesso
1.2	Avaliar situação	Recursos, restrições, riscos e contingências
1.3	Determinar objetivos de análise	Metas analíticas e critérios técnicos
1.4	Produzir plano do projeto	Cronograma, responsáveis e ferramentas

## 3.6 Fase 2: Compreensão dos Dados (*Data Understanding*)

Após a definição dos objetivos, inicia-se a compreensão dos dados disponíveis. O foco desta fase é conhecer profundamente as fontes de dados, suas características, limitações e problemas de qualidade.

São realizadas atividades de coleta inicial, descrição, exploração e verificação da qualidade dos dados, fornecendo base sólida para as etapas seguintes.

### 3.6.1 Checklist da Fase 2

#	Tarefa	Resultados esperados
2.1	Coletar dados iniciais	Relatório de coleta
2.2	Descrever dados	Relatório de descrição
2.3	Explorar dados	Relatório de exploração
2.4	Verificar qualidade dos dados	Relatório de qualidade

## 3.7 Fase 3: Preparação dos Dados (*Data Preparation*)

A preparação dos dados é, frequentemente, a fase mais trabalhosa do processo. Seu objetivo é transformar dados brutos em um conjunto adequado para análise e modelagem.

Inclui seleção de registros e variáveis, limpeza de inconsistências, criação de atributos derivados, integração de diferentes fontes e padronização de formatos.

### 3.7.1 Checklist da Fase 3

#	Tarefa	Resultados esperados
3.1	Selecionar dados	Justificativas de inclusão e exclusão
3.2	Limpar dados	Relatório de limpeza
3.3	Construir dados	Atributos derivados
3.4	Integrar dados	Conjuntos integrados
3.5	Formatar dados	Conjunto final documentado

## 3.8 Fase 4: Modelagem (*Modeling*)

Na fase de modelagem, técnicas analíticas e estatísticas são aplicadas aos dados preparados. São escolhidos algoritmos apropriados, definidos procedimentos de teste e ajustados parâmetros dos modelos.

É importante destacar que decisões tomadas nesta fase dependem fortemente das escolhas realizadas nas etapas anteriores.

### 3.8.1 Checklist da Fase 4

#	Tarefa	Resultados esperados
4.1	Selecionar técnicas	Técnica e pressupostos
4.2	Gerar design de teste	Estratégia de validação
4.3	Construir modelo	Modelos e parâmetros
4.4	Avaliar modelo	Métricas e ajustes

## 3.9 Fase 5: Avaliação (*Evaluation*)

A fase de avaliação verifica se os resultados obtidos geram valor para o negócio. Não se trata apenas de desempenho técnico, mas de aderência aos objetivos definidos inicialmente.

Nesta etapa, decide-se se o modelo está pronto para implementação, se ajustes adicionais são necessários ou se novas análises devem ser conduzidas.

### 3.9.1 Checklist da Fase 5

#	Tarefa	Resultados esperados
5.1	Avaliar resultados	Comparação com critérios de sucesso
5.2	Revisar processo	Lições aprendidas
5.3	Determinar próximos passos	Decisões e ações

### 3.10 Fase 6: Implementação (*Deployment*)

A implementação transforma análises e modelos em instrumentos efetivos de decisão. Envolve integração com sistemas organizacionais, definição de métricas de monitoramento e produção de relatórios ou dashboards para os gestores.

Essa fase também inclui a documentação do projeto e a consolidação de aprendizados para uso em iniciativas futuras.

#### 3.10.1 Checklist da Fase 6

#	Tarefa	Resultados esperados
6.1	Planejar implantação	Plano de implantação
6.2	Planejar monitoramento e manutenção	Plano de monitoramento
6.3	Producir relatório final	Relatório e apresentação
6.4	Revisar projeto	Documentação de experiência

## 4 Tipos de análise e Pertuntas de Negócio

Ao longo do processo CRISP-DM, diferentes tipos de análise podem ser empregados, variando em complexidade técnica e valor estratégico. Cada tipo de análise responde a uma pergunta distinta e reflete um determinado nível de maturidade analítica da organização.

As análises descritiva, diagnóstica, preditiva e prescritiva não são excludentes. Pelo contrário, costumam ser adotadas de forma progressiva e complementar, à medida que a organização desenvolve suas capacidades analíticas e passa a utilizar dados de forma mais sistemática em seus processos decisórios.

### 4.1 Visão geral dos tipos de análise

O termo *analytics* refere-se a um espectro contínuo de técnicas analíticas que evoluem em complexidade e impacto organizacional. Esse espectro vai da compreensão do que ocorreu no passado até a recomendação de ações otimizadas para o futuro.

Cada tipo de análise está associado a uma pergunta fundamental de negócio e a um conjunto específico de métodos. Em geral, as organizações avançam nessa jornada de forma gradual, consolidando competências em um nível antes de progredir para o seguinte.

A maturidade analítica de uma empresa pode ser avaliada pelo equilíbrio e pela profundidade com que esses quatro tipos de análise são incorporados às rotinas de decisão.

### 4.2 Análise descritiva

A análise descritiva representa o primeiro nível de uso sistemático de dados. Seu objetivo é resumir e organizar informações históricas de modo a tornar o passado comprehensível.

O que aconteceu?

- **Objetivo:** Sumarizar dados históricos para identificar padrões, tendências e comportamentos recorrentes.
- **Técnicas:** Estatísticas descritivas, tabelas, visualizações e dashboards.

- **Complexidade:**

No contexto da Administração, a análise descritiva é amplamente utilizada em relatórios gerenciais e sistemas de acompanhamento operacional.

**Exemplos** incluem relatórios de vendas por canal durante períodos promocionais, dashboards de monitoramento de indicadores em tempo real e análises de segmentação de clientes por região ou perfil de consumo.

### 4.3 Análise diagnóstica

A análise diagnóstica aprofunda a análise descritiva ao buscar explicações para os padrões observados. Enquanto a análise descritiva mostra *o que* aconteceu, a diagnóstica procura entender *por que* aconteceu.

Por que aconteceu?

- **Objetivo:** Investigar causas, relações e fatores associados aos resultados observados.
- **Técnicas:** Análise de correlação, segmentação, *drill-down*, análise de fatores e comparações entre grupos.
- **Complexidade:**

Em Administração, esse tipo de análise é fundamental para apoiar decisões corretivas e ajustes de estratégia, como investigar as causas de queda nas vendas após um reajuste de preços ou identificar fatores associados ao aumento do turnover em determinadas unidades da empresa.

### 4.4 Análise preditiva

A análise preditiva utiliza dados históricos para estimar comportamentos futuros. Nesse nível, modelos estatísticos e de aprendizagem de máquina passam a desempenhar papel central.

O que provavelmente acontecerá?

- **Objetivo:** Prever tendências futuras e resultados prováveis com base em padrões históricos.
- **Técnicas:** modelos de regressão, análise de séries temporais, algoritmos para classificação.
- **Complexidade:**

Exemplos típicos em Administração incluem a previsão de demanda para produtos sazonais, modelos de propensão à inadimplência em instituições financeiras e estimativas de giro de estoque para apoiar decisões de compras e logística.

## 4.5 Análise prescritiva

A análise prescritiva representa o nível mais avançado da jornada analítica. Seu foco não está apenas em prever resultados, mas em recomendar ações que maximizem objetivos organizacionais, considerando restrições e trade-offs.

O que devemos fazer?

- **Objetivo:** Recomendar decisões e ações otimizadas com base em análises descritivas, diagnósticas e preditivas.
- **Técnicas:** Otimização, simulação, algoritmos de decisão e sistemas de recomendação.
- **Complexidade:**

No contexto empresarial, a análise prescritiva é utilizada em problemas como otimização do mix de produtos por loja, recomendação personalizada em plataformas de e-commerce e definição automática de rotas logísticas em ambientes urbanos complexos.

Em geral, esse tipo de análise exige não apenas maturidade técnica, mas também processos decisórios bem estruturados e integração com sistemas operacionais.

# Referências

- HARKNESS, T. [The history of the data economy: Part I: The birth of customer insight.](#) **Significance**, v. 18, n. 2, p. 12–15, a2021.
- \_\_\_\_\_. [The history of the data economy: Part II: Analytics arrives.](#) **Significance**, v. 18, n. 4, p. 16–19, b2021.
- \_\_\_\_\_. [The history of the data economy: Part III: The new kings and queens of data.](#) **Significance**, v. 18, n. 5, p. 16–19, c2021.
- \_\_\_\_\_. [The history of the data economy: Part IV: The future.](#) **Significance**, v. 18, n. 6, p. 12–15, d2021.
- IBM CORPORATION. **Guia do IBM SPSS Modeler CRISP-DM.** [s.l.] IBM Corporation, 2023.