

# SEJA UM ANALISTA DE DADOS

COMO USAR A ANÁLISE PARA  
TRANSFORMAR DADOS EM VALOR

**JORDAN MORROW**



ALTA BOOKS  
GRUPO EDITORIAL  
Rio de Janeiro, 2024

# SUMÁRIO

*Sobre o autor* 1

*Agradecimentos* 3

*Prefácio* 5

## **Introdução** 7

A analogia do médico 7

Detalhamento de conteúdo 10

Resumo 11

Nota 12

## **PARTE UM** 13

Dados e análise 13

### 1 DEFININDO DADOS E ANÁLISE 15

O exemplo da mineração na montanha 16

Habilidades de dados e análise — alfabetização de dados 19

Baseando-se em dados 22

PMV — proficiência mínima viável 24

Resumo do capítulo 26

Notas 27

### 2 DEFININDO OS QUATRO NÍVEIS DE ANÁLISE 29

Nível analítico 1 — descritivo 29

Nível analítico 2 — diagnóstico 32

Nível analítico 3 — preditivo 34

Nível analítico 4 — prescritivo 37

Resumo do capítulo 38

Notas 40

3	<b>O PODER DA ANÁLISE NA TOMADA DE DECISÃO</b>	41
	Dados	42
	Análise	43
	Análise descritiva	44
	Análise diagnóstica	45
	Análise preditiva	47
	Análise prescritiva	48
	Estrutura, decisão, storytelling com dados	49
	Resumo do capítulo	50
	Notas	51
	<b>PARTE DOIS</b>	53
	Os quatro níveis de análise: definir, potencializar, entender e aprender	53
4	<b>ANÁLISE DESCRITIVA</b>	55
	O que é a análise descritiva?	55
	Funções	59
	Ferramentas e tecnologias	63
	Resumo do capítulo	63
5	<b>COMO A ANÁLISE DESCRITIVA É USADA HOJE?</b>	65
	Democratização de dados	65
	Democratização da análise descritiva — ferramentas e tecnologia	67
	Exemplos de mercado	73
	Ética em dados e análise descritiva	75
	Resumo do capítulo	75
	Notas	76
6	<b>COMO PESSOAS E ORGANIZAÇÕES PODEM SE APERFEIÇOAR NA ANÁLISE DESCRITIVA</b>	77
	Análise descritiva e solução de problemas baseada em dados	78
	Análise descritiva e tomada de decisão baseada em dados	84
	Análise descritiva e execução baseada em dados	86
	Resumo do capítulo	88
	Notas	89

- 7 **ANÁLISE DIAGNÓSTICA** 91  
 O que é análise diagnóstica? 91  
 A crise imobiliária 91  
 A pergunta “por que” 92  
 Um exemplo pessoal 93  
 Análise diagnóstica e funções organizacionais 94  
 Ferramentas e tecnologias 99  
 Resumo do capítulo 100  
 Notas 101
- 8 **COMO A ANÁLISE DIAGNÓSTICA É USADA HOJE?** 103  
 Democratização de dados — análise diagnóstica 103  
 Democratização da análise diagnóstica — ferramentas e tecnologia 104  
 Análise diagnóstica — visualização de dados 105  
 Análise diagnóstica — programação 106  
 Análise diagnóstica — estatística 107  
 Exemplos de mercado — continuação do Capítulo 5 107  
 Ética em dados e análise diagnóstica 111  
 Resumo do capítulo 111  
 Nota 112
- 9 **COMO PESSOAS E ORGANIZAÇÕES PODEM SE  
 APERFEIÇOAR NA ANÁLISE DIAGNÓSTICA** 113  
 Mentalidade de dados e análise — pessoas 113  
 Mentalidade de dados e análise — organizações 114  
 Análise diagnóstica e o tridata 116  
 Análise diagnóstica e solução de problemas baseada em dados 116  
 Análise diagnóstica e tomada de decisão baseada em dados 120  
 Análise diagnóstica e execução baseada em dados 121  
 Uma observação sobre o aprendizado 122  
 Resumo do capítulo 123  
 Notas 124
- 10 **ANÁLISE PREDITIVA** 125  
 O que é análise preditiva? 125  
 Funções 129

- Ferramentas e tecnologias 133
- Resumo do capítulo 134
- Notas 135
  
- 11 **COMO A ANÁLISE PREDITIVA É USADA HOJE?** 137
  - Democratização de dados — análise preditiva 137
  - Democratização da análise preditiva — ferramentas e tecnologia 140
  - Visualização de dados, storytelling com dados e mais 142
  - Exemplos de mercado — continuação do Capítulo 8 145
  - Ética em dados e análise preditiva 147
  - Resumo do capítulo 148
  - Notas 148
  
- 12 **COMO PESSOAS E ORGANIZAÇÕES PODEM SE APERFEIÇOAR NA ANÁLISE PREDITIVA** 149
  - Mentalidade de dados e análise — análise preditiva 149
  - A mentalidade importa 150
  - Análise preditiva e o tridata 152
  - Análise preditiva e a solução de problemas baseada em dados 152
  - Análise preditiva e a tomada de decisão baseada em dados 155
  - Análise preditiva e a execução baseada em dados 156
  - Resumo do capítulo 157
  
- 13 **ANÁLISE PRESCRITIVA** 159
  - O que é a análise prescritiva? 159
  - Funções 165
  - Ferramentas e tecnologias 166
  - Resumo do capítulo 166
  - Notas 167
  
- 14 **COMO A ANÁLISE PRESCRITIVA É USADA HOJE?** 169
  - Democratização de dados — análise prescritiva 169
  - Democratização da análise prescritiva — ferramentas e tecnologia 171
  - Estratégia de dados 172
  - Storytelling com dados 173
  - Exemplos de mercado 173
  - Ética em dados e análise prescritiva 176

- Resumo do capítulo 176
- Notas 177
- 15 **COMO PESSOAS E ORGANIZAÇÕES PODEM SE APERFEIÇOAR NA ANÁLISE PRESCRITIVA** 179
- Mentalidade de dados, análise e alfabetização de dados — análise prescritiva 179
- Análise prescritiva e o tridata 182
- Solução de problemas baseada em dados e análise prescritiva 182
- Tomada de decisão baseada em dados e análise prescritiva 183
- Execução baseada em dados e análise prescritiva 185
- Resumo do capítulo 186
- Nota 186
- PARTE TRÊS** 187
- Juntando tudo 187
- 16 **USANDO OS QUATRO NÍVEIS DE ANÁLISE PARA CAPACITAR A TOMADA DE DECISÃO** 189
- Seis etapas da progressão analítica 189
- Tomar uma decisão com os quatro níveis de análise 196
- Resumo do capítulo 199
- Conclusão** 201
- Ida ao médico 201
- Mentalidade iterativa 201
- O tridata 202
- Os quatro níveis de análise — um por um 202
- O propósito dos dados e da análise 203
- Juntando tudo isso em seus negócios e em sua vida pessoal — principais conclusões 204
- Notas 205
- 17 **ÍNDICE** 207

# PARTE UM

## Dados e análise



Já que está lendo este livro e, possivelmente, já leu meus outros dois livros, suponho que tenha alguma compreensão do que são dados e análise. Caso contrário, esta seção estabelecerá uma base e o capacitará com conhecimento sobre esses dois tópicos. Dito isso, não gastarei uma quantidade exorbitante de tempo para defini-los, embora forneça o que poderia chamar de definições de livros didáticos. Em vez disso, quero discutir dados e análise por meio do que chamarei de “mina de ouro” de dados e análise, ou, neste caso, uma mina de dados e análise. Para ajudar a discutir essa mina, entraremos no vasto mundo dos dados.



# Definindo dados e análise

Qual a finalidade dos dados? É algo mágico que proporcionará às pessoas e organizações mais sucesso do que jamais sonharam? Bem, isso pode ser uma hipérbole, mas a realidade é que pessoas e organizações podem se fortalecer muito com dados e poder analítico. Para ajudar nisso, daremos uma olhada na grande quantidade de dados que existe no mundo e, neste caso, na quantidade que existirá. Pode ser muito legal pensar na quantidade de dados que existe no mundo. Os números que podem ser lançados com a quantidade de dados podem ser vistos como algo insondável.

Há muitas maneiras divertidas de analisar os dados, mas daremos uma olhada na quantidade total. De acordo com o Departamento de Pesquisa do Statista, a previsão é que até o ano de 2025 a criação de dados em todo o mundo será de mais de 180 zettabytes. Você sabe quantos bytes existem em um zettabyte? Você <sup>[1]</sup> sequer sabia que esse termo existia? Deixe-me escrever esse número para você, e saiba que há 21 zeros em um zettabyte: 180.000.000.000.000.000.000. Não é irreal? Estamos produzindo muitos dados. Lembre-se de que nem todos os dados são valiosos ou impactantes para nós, mas, por enquanto, nos deleitaremos com a quantidade de dados que estamos criando.

Outra maneira de ver isso vem do FirstsiteGuide, que resume dizendo que levaríamos 181 milhões de anos para baixar todos os dados da internet hoje. Esse <sup>[2]</sup> artigo foi escrito ou atualizado em janeiro de 2022. São tantos dados — 181 milhões de anos para fazer isso! Não sei quanto a você, mas eu tenho outras coisas para fazer além de me preocupar com o download de todos eles, inclusive, não viverei tanto tempo. Para se ter uma ideia, a ciência diz que há cerca de 180 milhões de anos foi quando a Pangeia se dividiu. Por q<sup>[3]</sup>ue estamos produzindo tantos dados? Por que podemos armazenar tanto?

A realidade é que a digitalização do mundo está permitindo a produção de mais e mais dados. Com tantos dados sendo produzidos e a barata disponibilidade de armazenamento de dados, por que não armazenar cada

vez mais? Agora, este não será um livro ou capítulo sobre tecnologia de armazenamento em nuvem, porém é mais fácil armazenar dados em nossa era atual. Com dados e análise, ter mais não é necessariamente o que queremos. Vou lhe dar um exemplo. Talvez eu goste de sorvete, mas, quando alguém pega o pote de sorvete maior, o sorvete acaba perdendo sua utilidade (um professor usou essa analogia em meu curso de graduação na universidade). A última colherada não tem um sabor tão fantástico quanto a primeira. O mesmo pode acontecer com todos os dados que produzimos e armazenamos. Primeiro, o excesso de dados pode dificultar a limpeza, o gerenciamento e que se arquitete bem os dados. Em segundo lugar, pode ser difícil encontrar os dados certos de que você precisa para responder às suas perguntas. Terceiro, isso pode custar à organização, não em custos de armazenamento em nuvem, e sim no custo potencial e/ou de oportunidade encontrado na pesquisa e na utilização de dados em excesso para responder às perguntas.

Com todos esses dados, pode-se dizer: “E daí?” Pode-se dizer: “Quem se importa?” Alguém pode até perguntar: “Você não acha que todos esses dados são exagerados?” Podemos nos deparar com histórias incríveis de como os dados foram usados ou estão sendo usados, mas será que dados sustentam todo o exagero? A resposta é sim, o exagero é real e justificado. Sim, estamos produzindo e podendo armazenar grandes quantidades de dados, mas eles ficam ali. É preciso trabalhar para lhes dar vida, para torná-los analisáveis, para potencializá-los a nós mesmos e a nossas organizações. Precisamos limpar, organizar, projetar, arquitetar e estabelecer processos. Precisamos fazer muito com eles. Uma das principais coisas que precisamos fazer é dar vida aos dados. Eles têm valor inerente, e é necessário trabalhar a fim de prepará-los para o trabalho de front-end, mas dados sem análise são apenas dados. No entanto, quero deixar claro que a análise sem bons dados pode ser uma coisa muito ruim.

## O EXEMPLO DA MINERAÇÃO NA MONTANHA

Para dar vida a isso, usarei uma analogia. Minerais valiosos estão presentes na terra. A mineração é um processo de extração desses minerais e de obtenção de valor com eles. Quero que imagine que descobriu uma montanha com ouro e diamantes — bem incrível. Você espera extrair muitos minerais valiosos, mas não tem certeza do valor exato dos minerais

encontrados nela. De qualquer forma, você decide investir nisso, esperando receber um grande retorno sobre seu investimento. O que você faz em seguida? Bem, decide construir uma mina e uma empresa para obter esse valor. Para essa mina, você compra o terreno necessário, estabelece a infraestrutura certa, compra as ferramentas e os equipamentos certos e contrata o pessoal certo. Então você pensa e desenvolve os processos de engenharia para extrair o ouro ou os minerais valiosos que encontrar. O interessante disso é que a mina em si tem muito valor real e potencial. Pode-se, de fato, dizer que a mina pode valer bilhões e bilhões de dólares, dependendo da quantidade de ouro ou de diamantes nela contida. Mas será que o fato de saber que há valor significa que se está explorando todo o potencial da mina? Não, é claro que não. É preciso uma estratégia, é necessário o pessoal certo, não apenas aquele que você acha que deve contratar, são necessários processos, eficácia, governança e controle. O que é realmente necessário é um processo e um plano estratégicos, nos quais se faça as coisas de forma eficaz e eficiente, para que se possa obter o valor buscado. Em essência, você precisa extrair o ouro e os minerais e, em seguida, vendê-los ou transformá-los em valor. Não é tão simples, certo? Pode não ser simples, mas você precisa fazer isso de forma eficaz para obter um forte retorno sobre seu investimento.

Uma empresa que utilizou dados para ajudá-la a aumentar o valor é a Netflix. Eu arriscaria dizer que muitas empresas de streaming usam dados para otimizar seus serviços. No caso da Netflix, a empresa usa dados para aprimorar sua organização e criar um benefício para seus clientes.<sup>[4]</sup>

Para aprofundar nossa analogia com a mina, uma vez que tenha a engenharia e a infraestrutura instaladas, você poderá extrair mais o valor da mina e retirar os diamantes ou o ouro, encontrando, assim, o valor ou o “insight” (sim, estou associando isso agora à mina de dados e análise de que estamos falando) dentro da mina. A extração do mineral é um processo que pode ser simples, talvez apenas uma simples pá seja necessária; ou complexo, quando grandes máquinas ou processos são necessários para se extrair o valor da mina, mas o tempo, o custo e a energia usada para isso valem a pena.

Por fim, depois de comprar e estabelecer sua mina, e depois de se esforçar para explorá-la a fim de obter seu valor, você precisará ter um propósito ou tomar uma decisão com o ouro e os diamantes em mãos. Seja derreter o ouro ou vender os diamantes para joias, o que quer que seja, você precisa tomar uma decisão sobre o que fará com os minerais extraídos.

Como você já deve ter adivinhado, a analogia da mineração é uma correlação com a mina de dados e análise que quero que construa em sua própria mente, para utilizar em sua vida pessoal e em seu trabalho/carreira. Essa também é uma analogia relacionada à estratégia de dados de uma organização. As organizações estão armazenando grandes quantidades de dados. Há ouro ou diamantes nessa grande quantidade de dados, e a análise é o equipamento necessário para extraí-los. Vamos explicar o que quero dizer.

As organizações têm essa capacidade de construir suas próprias “minas” de dados, neste caso, arquiteturas de dados em larga escala, modelagem, engenharia, e assim por diante. Principalmente estabelecendo a base necessária para o sucesso dos dados e análise, a organização realizou a primeira etapa de sua mina de dados e análise organizacional. A base de dados, a arquitetura, a modelagem e o armazenamento de dados representam a construção da estrutura e a apropriação da mina. Em sua mina de dados, há valor. Os dados têm valor. Você está armazenando muito potencial, mas precisa de algo para extrair esse valor. Aqui está o nosso mundo da análise.

*Analytics* é definido pelo dicionário Merriam-Webster como análise... bem, o que isso significa?<sup>[5]</sup> Análise é definida como um olhar profundo em algo que pode ser complexo para que se entenda, como um todo ou considerando suas partes isoladas.<sup>[6]</sup> A mesma definição de análise do Merriam-Webster basicamente diz que estamos observando todas as partes que compõem algo em seu todo.<sup>[7]</sup> Na análise, examinamos os dados e as informações — em alguns casos, pode ser uma visão maciça e holística dos dados — e os dividimos em partes. Estamos encontrando seus componentes. Essa é uma maneira fantástica e surpreendente de ver as coisas. Nesse caso, estamos desmembrando o terreno ao redor da mina para encontrar o ouro e os diamantes que estão lá dentro.

A análise, com os dados, transforma dados e informações em recursos úteis e significativos e nos ajuda a entender sua natureza. Com os dados, temos uma ferramenta poderosa para nos ajudar a tentar descobrir o “por que” das coisas que estamos vendo em uma organização, talvez por meio de painéis, visualizações de dados ou KPIs. A análise tem o poder de nos levar de grandes somas de dados para o valor. No caso da mina, a extração trouxe à tona o valor da enorme quantidade de minerais. Esse valor estava na forma de ouro e diamantes.

Vivemos em um mundo de sinal e ruído. Você já ouviu esses termos antes? Sinal e ruído nos vêm do uso do rádio. Nos rádios, você girava o botão para passar pelo “ruído” da estática e encontrar o “sinal” do rádio. Hoje em dia, é mais fácil ouvir nossa música com clareza, mas antes era assim que se fazia. Essa analogia de girar o botão de dados para encontrar o ruído é poderosa. Em nosso caso, vamos pensar na análise como o botão que estamos girando para encontrar nosso insight, o sinal. A realidade é que há muito ruído nos dados (ou seja, partes que não têm valor), e nosso trabalho é encontrar o sinal. Esses são todos os nossos dados, pessoais ou organizacionais. Essa discussão sobre sinal e ruído está de acordo com nossa analogia da mineração. Muito similar a uma mina de ouro, onde é preciso peneirar a sujeira para encontrar o valor. Como uma organização trabalha para encontrar o insight nos dados, é importante entender essa analogia entre o sinal e o ruído.

## HABILIDADES DE DADOS E ANÁLISE — ALFABETIZAÇÃO DE DADOS

No mundo dos dados e da análise, é necessário que existam as competências e habilidades certas em uma organização. Você não extrairá minerais de sua mina sem as pessoas e as ferramentas certas, e o mesmo se aplica aos dados e à análise. No caso dos dados e da análise, esse é o mundo dos dados e da alfabetização, sobre o qual escrevi em meu primeiro livro, *Be Data Literate*. A alfabetização de dados deve capacitar as pessoas com as competências e habilidades certas para dar vida aos dados. Lembre-se da Introdução deste livro. Na verdade, existem quatro níveis de análise no mundo atual, mas a maioria das pessoas passará seu tempo nos dois primeiros níveis: descritivo e diagnóstico. Lembre-se da analogia que compartilhei sobre estar doente. Sabemos que há pepitas de sabedoria e conhecimento no que diz respeito à sua melhora — o que está errado e por quê, e o que fazer para ajudar na sua recuperação. O diagnóstico dos dados é o sinal, mas adivinhe só? Há muitas coisas diferentes que um médico pode escolher. Sim, algumas seriam absurdas com doenças diferentes, mas esperamos que, por meio dos dados e das informações fornecidas, o sinal seja encontrado e todo o ruído seja eliminado.

Esse é o mesmo princípio no trabalho de dados e análise. Os indivíduos de uma organização devem ter as habilidades necessárias para decifrar todo o ruído e encontrar o sinal, que se espera que se transforme em crescimento,

retorno e decisões incríveis para a empresa. Vamos ver um exemplo de dados e análise no trabalho usando uma de minhas visualizações de dados favoritas: a marcha de Napoleão em Moscou. Veja a Figura 1.1.

Dê uma olhada nessa imagem, atribuída ao engenheiro civil Charles Joseph Minard. Essa visualização mostra o que foi gerado a partir de algo em uma época na qual talvez tenha havido muitos dados. Podemos ver, à esquerda, Napoleão marchando em direção a Moscou, movendo-se da esquerda para a direita. À medida que vemos esse movimento, percebemos que a linha mais clara fica mais fina onde o exército de Napoleão se torna menor, e a viagem de volta em preto mostra que isso ocorre ainda mais. Que visualização maravilhosa! Mas será que isso nos mostra por que isso ocorreu? Não, não exatamente. Aqui encontramos, talvez, algum sinal dentro do ruído, mas, de modo geral, seria necessário fazer mais.

Geralmente, estamos montando uma imagem poderosa dos dados e da análise. Vamos dar uma olhada em mais exemplos de dados e análise, para que nossa mente fique bem posicionada sobre o que cada um é, o que nos capacitará nos próximos capítulos.

#### Exemplos de dados:

- As informações que recebemos para nos informar sobre a situação econômica atual.
- A American Express é uma organização que tem muitos dados. Toda vez que uma pessoa usa um de seus cartões de crédito, são gerados dados para a American Express.
- Seu aniversário, seu número de identificação pessoal (equivalente ao CPF no Brasil), sua pontuação de crédito, possíveis informações fornecidas a empresas de mídia social e tudo o mais que compõe as informações pessoais.
- O número de vendas de carros por fabricante.
- E muitas outras formas de dados.

#### Exemplos de análise:

- Descobrir que há sazonalidade nas tendências de vendas de carros.
- Insights fornecidos por aprendizado de máquina e inteligência artificial, geralmente na forma de análises preditiva e prescritiva.
- Previsões de vendas no varejo.
- Painéis de controle de uma organização, geralmente na forma de análise descritiva.
- E muitas, muitas outras formas de análise.



## BASEANDO-SE EM DADOS

Não é preciso dizer que, para que uma pessoa e/ou uma organização seja baseada em dados, os dados e a análise devem desempenhar um papel fundamental. Para ser baseada em dados, a organização deve utilizar dados em seus processos de tomada de decisão. Faz sentido que, para que uma organização seja baseada em dados, os dados e a análise devam ser utilizados como parte vital da organização. Certo, este pode ser um momento “dã”, mas a realidade é que o fluxo de dados e análise em uma organização é feito de forma inadequada no mundo atual. Em alguns casos, de forma muito ruim. De acordo com minha experiência e com os convites contínuos que recebo de empresas para falar sobre alfabetização de dados ou outros tópicos, essas empresas estão investindo mais em sua alfabetização de dados ou não estão onde querem estar de uma perspectiva baseada em dados. Um exemplo de uma organização que investe pesadamente e da maneira correta é o Miami Dade College, que se esforça para ser baseado em dados. Não se trata de uma organização pobre em dados, mas de uma organização que está se esforçando para ser uma “faculdade baseada em dados”, como diz a senhora presidente. A escola investe muito no painel de liderança, e tive a grande oportunidade de participar de um workshop sobre dados. Que exemplo de como fazer as coisas da maneira certa! A senhora presidente está responsabilizando os líderes e ela “percebe” isso, o que significa que ela entende a necessidade de ser baseada em dados. Com esse bom exemplo, por que alguns estão se saindo mal?

Na seção anterior, falamos sobre alfabetização de dados. A falta de alfabetização de dados no mundo de hoje, sendo o ano de 2022 enquanto escrevo este texto, é impressionante. No estudo da Qlik sobre alfabetização de dados, lançado em 2022, descobriu-se que apenas 11% das pessoas estão totalmente confiantes em suas habilidades de alfabetização de dados.<sup>[8]</sup> Com um nível tão baixo de alfabetização de dados, não é de admirar que as organizações não estejam conseguindo ser baseadas em dados, não é mesmo?

Outro problema que encontramos na falta de habilidades para se basear em dados é a maneira como os dados e a análise estão dispersos em uma organização. Sua organização tem uma estratégia de dados holística bem comunicada em toda a organização? Você sabe quem tem o quê? Talvez ainda mais importante, você sabe “por que” eles têm o que têm?

Os dados e a análise devem ser dispersos e permeados por toda a organização. Quando pensamos na analogia dos dados e da análise sob a perspectiva de um médico, podemos ver como funciona o processo ou o fluxo de pensamento