

Uma visão geral da inteligência de negócios, análise de dados e ciência de dados

OBJETIVOS DIDÁTICOS

- Compreender a necessidade de apoio computadorizado para se tomar decisões gerenciais.
- Reconhecer a evolução de tal apoio computadorizado no ambiente atual – análise/ciência de dados.
- Descrever a metodologia e os conceitos da inteligência de negócios (BI – *business intelligence*).
- Aprender sobre os diferentes tipos de análise de dados e examinar aplicações selecionadas.
- Compreender o ecossistema de análise de dados a fim de identificar os diversos agentes relevantes e as oportunidades de carreira profissional.

O ambiente empresarial está em constante evolução, tornando-se cada vez mais complexo. Organizações, tanto públicas quanto privadas, veem-se pressionadas a reagir rapidamente a tal evolução e a inovarem seu *modus operandi*. Isso exige que as organizações sejam ágeis e tomem decisões operacionais frequentes, rápidas, estratégicas e táticas, algumas das quais são bastante complexas. Para que tais decisões sejam tomadas, pode ser preciso quantidades consideráveis de dados, informações e conhecimentos relevantes. E seu processamento, à luz das decisões necessárias, deve ser feito com rapidez, muitas vezes em tempo real, o que geralmente exige suporte computadorizado.

Este livro trata da análise de negócios na forma de suporte computadorizado para a tomada de decisões gerenciais. Seu foco recai nas bases teóricas e conceituais do apoio a decisões, bem como nas ferramentas e técnicas comerciais que se encontram disponíveis. Nele, são apresentados os fundamentos das técnicas e o modo como tais sistemas são construídos e usados. Seguimos uma abordagem EEE ao introduzirmos esses tópicos: **Exposição, Experiência e Exploração**. O livro, antes de mais nada, expõe diversas técnicas de análise de dados e suas aplicações. A ideia é que o estudante sinta-se inspirado ao aprender como outras organizações vêm empregando a análise de dados para tomar decisões ou obter vantagens competitivas. Acreditamos que tal **exposição** às recentes aplicações da análise de dados, e a como

isso pode ser replicado, é um componente didático básico para aprender sobre o tema. Ao descrever as técnicas, também apresentamos exemplos de ferramentas de software que podem ser usadas para desenvolver tais aplicações. O livro não se atém a uma única ferramenta de software, para que os estudantes possam **experimentar** essas técnicas usando diversas delas disponíveis no mercado. Esperamos que essa exposição e experiência habilite e motive os leitores a explorarem o potencial de tais técnicas em seus próprios ramos de atuação. Para facilitar tal **exploração**, incluímos exercícios que direcionam o leitor para a Teradata University Network (TUN) e para outros sites que incluem exercícios em grupo onde apropriado.

1.1 VINHETA DE ABERTURA: Análise de dados esportivos – uma empolgante fronteira para aprender e entender as aplicações da análise de dados

A aplicação da análise de dados a problemas empresariais é uma habilidade-chave, sobre a qual você aprenderá neste livro. Atualmente, muitas dessas técnicas estão sendo aplicadas para aprimorar processos decisórios em todos os aspectos dos esportes, uma área em franca ebulição denominada análise de dados esportivos (*sports analytics*). A análise de dados esportivos é a arte e a ciência de reunir dados sobre atletas e equipes para melhor embasar decisões esportivas, como decidir quais jogadores devem ser recrutados, quanto pagar de salário a cada atleta, quem colocar para jogar, quem escolher como treinador, como evitar lesões e quando atletas devem ser negociados ou aposentados. Para equipes, isso envolve decisões empresariais como o preço de ingressos, bem como decisões de formação de plantel, análise de pontos fortes e fracos de equipes adversárias e muitas decisões no calor do jogo.

Na verdade, a análise de dados esportivos está se tornando uma especialidade dentro da análise de dados. Trata-se de uma área importante, pois os esportes representam uma grande indústria, gerando cerca de U\$145 bilhões em receitas anuais, sem contar os U\$100 bilhões adicionais em apostas legais e outros U\$300 bilhões em ilegais, segundo a PriceWaterhouse.¹ Em 2014, apenas U\$125 milhões foram gastos em análise de dados (menos de 0,1% das receitas). Calcula-se que este valor deve crescer em ritmo acelerado até alcançar U\$4,7 bilhões em 2021.²

O uso de análise de dados nos esportes foi popularizado pelo livro *Moneyball*, de Michael Lewis, de 2003, e pelo filme estrelado por Brad Pitt em 2011. Ambos retratam como Billy Beane, diretor executivo dos Oakland A's, utilizou análise de dados para fazer um time perdedor começar a ganhar. Em particular, ele contratou um analista que empregava análise de dados para draftar jogadores capazes de avançar de base, em oposição a jogadores que se destacavam em parâmetros mais tradicionais, como corridas impulsionadas e bases roubadas. Esses *insights* lhes permitiram draftar talentos renegados por outras equipes a salários iniciais razoáveis. E funcionou – eles chegaram aos *playoffs* em 2002 e 2003.

¹“Changing the Game: Outlook for the Global Sports Market to 2015,” PriceWaterhouseCoopers Report, em <https://www.pwc.com/gx/en/hospitality-leisure/pdf/changing-the-game-outlook-for-the-global-sports-marketto-2015.pdf>. Dados sobre apostas de <https://www.capcredit.com/how-much-americansspend-on-sports-each-year/>.

²“Sports Analytics Market Worth \$4.7B by 2021,” Wintergreen Research Press Release, noticiado por PR Newswire em <http://www.prnewswire.com/news-releases/sports-analytics-market-worth-47-billion-by-2021-509869871.html>, June 25, 2015.

Hoje, a análise de dados está sendo usada em todos os aspectos dos esportes. Sua aplicação pode ser dividida em departamento administrativo e departamento esportivo. Uma boa descrição com 30 exemplos aparece no artigo de levantamento de Tom Davenport.³ Quando empregada por departamentos administrativos, a análise de dados envolve a análise comportamental de torcedores, incluindo desde modelos preditivos de renovação de ingressos de temporada e venda regular de ingressos até a contabilização de tuítes de torcedores a respeito do time, dos atletas, da equipe técnica e dos proprietários. Isso é muito similar à tradicional gestão de relacionamento com clientes. A análise financeira também é uma área-chave, em que tetos salariais (no caso dos profissionais) e limites de bolsas de estudo (no caso de programas universitários) fazem parte da equação.

Já as aplicações por parte do departamento esportivo envolvem tanto atletas individuais quanto desempenho em equipe. Para atletas individuais, há um foco em modelos analíticos de recrutamento, análise de dados de força, condicionamento e desenvolvimento, além de modelos preditivos (MPs) para evitar treinamento excessivo e lesões. Pesquisas na área de concussões estão em alta. A análise de dados de equipes como um todo inclui estratégias e táticas, avaliação de adversários e formação do plantel ideal sob várias situações dentro e fora de campo.

Os exemplos a seguir ilustram como três organizações esportivas utilizam análise de dados para aprimorar operações esportivas, da mesma forma que a análise de dados aprimorou processos decisórios em setores tradicionais.

Exemplo 1: departamento administrativo

Dave Ward trabalha como analista de negócios para uma equipe de beisebol, focado em receitas. Ele analisa venda de ingressos, tanto na forma de pacotes para a temporada inteira quanto entradas avulsas para jogos únicos. Dentre as perguntas que sua área de responsabilidade tenta responder estão: por que compradores de ingressos de temporada renovam (ou deixam de renovar) seus pacotes, e quais fatores motivam compras avulsas de última hora. Outra pergunta é: quanto cobrar pelos ingressos.

Algumas das técnicas analíticas que Dave emprega incluem estatísticas simples sobre o comportamento de torcedores, como dados de públicos totais e enquetes sobre a propensão a compras futuras. No entanto, as respostas dos torcedores nem sempre correspondem a suas ações. Dave conduziu uma enquete junto aos torcedores por local de assento (“nível”) e perguntou sobre suas propensões a renovarem seu pacote de temporada. Mas quando comparou o que eles disseram com o que de fato fizeram, encontrou grandes diferenças (veja a Figura 1.1). Ele descobriu que 69% dos torcedores que responderam à enquete afirmando que “provavelmente não renovariam” na verdade o fizeram. Essa é uma descoberta útil com repercussões práticas – na Figura 1.1, clientes nas células verdes são os mais propensos a renovar seus pacotes, e por isso exigem menos investimentos em marketing para serem convertidos quando comparados, por exemplo, aos clientes nas células azuis.

Muitos fatores, porém, influenciam o comportamento de compra de ingressos por parte dos torcedores, o que motiva estatísticas e análises de dados mais sofisticadas. Para ambas áreas, mas sobretudo para ingressos avulsos, Dave está estimulando o uso de precificação dinâmica – o que substituiria o modelo de simples

³Thomas Davenport, “Analytics in Sports: The New Science of Winning,” International Institute for Analytics White paper, patrocinado por SAS, fevereiro de 2014. No site da SAS em: http://www.sas.com/content/dam/SAS/en_us/doc/whitepaper2/iia-analytics-in-sports-106993.pdf. (Acessado em julho de 2016.)

Nível	Bastante provável	Provável	Talvez	Provavelmente não	Certamente não
1	92	88	75	69	45
2	88	81	70	65	38
3	80	76	68	55	36
4	77	72	65	45	25
5	75	70	60	35	25

FIGURA 1.1 Renovações de ingressos de temporada – resultados de enquete.

precificação estática por nível de assento por uma precificação diariamente flutuante de assentos individuais. Essa é uma área fértil de pesquisa para muitas equipes esportivas, com um potencial imenso de aumento de receitas. Seu novo modelo de preços leva em consideração, por exemplo, o retrospecto atual da equipe, o adversário a ser enfrentado, as datas e os horários dos jogos, quais estrelas atuam em cada time, o histórico de cada torcedor em renovação de ingressos de temporada ou compra de ingressos avulsos, bem como fatores como localização de assento, quantidade de assentos e informações em tempo real sobre congestionamentos no horário do jogo e até mesmo as condições meteorológicas. Veja a Figura 1.2.

Quais desses fatores são importantes? Em que ordem? Com sua extensa experiência em estatística, Dave constrói modelos de regressão para isolar os fatores que melhor determinam para esses históricos comportamentais e para criar MPs identificando como gastar recursos de marketing para aumentar as receitas. Ele desenvolve modelos de fidelidade para criar segmentos de clientes que renovarão seus pacotes de temporada, que não os renovarão ou que ficarão indecisos, o que acaba possibilitando campanhas de marketing mais refinadas.

Além disso, ele afere o sentimento dos torcedores conforme comentários em redes sociais, ajudando a segmentá-los em patamares de fidelidade. Outros estudos sobre motivadores de comparecimento a jogos avulsos ajudam o departamento de marketing a entender o impacto de distribuição de brindes, como bonecos e camisas, ou de anúncios publicitários na TV.

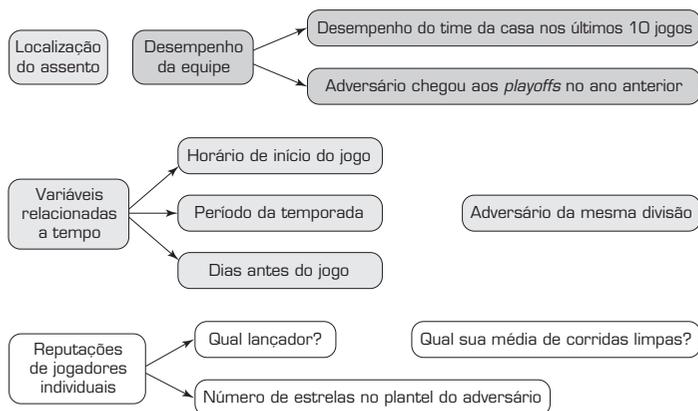


FIGURA 1.2 Trabalho anterior de precificação dinâmica – Major League Baseball. *Fonte:* adaptado de C. Kemper and C. Breuer, “How Efficient is Dynamic Pricing for Sports Events? Designing a Dynamic Pricing Model for Bayern Munich”, *Intl. Journal of Sports Finance*, 11, pp. 4-25, 2016.

Além das receitas, há muitas outras áreas analíticas abordadas pela equipe de Dave, incluindo merchandising, receitas de transmissão por rádio e TV, dicas ao diretor executivo quanto a negociações salariais, análise de recrutamentos tendo em vista o teto de folha salarial, nível de eficiência de canais publicitários e conscientização de marca, além de análise de parcerias. Ele é um cara bastante atarefado!

Exemplo 2: o treinador

Bob Breedlove é o treinador de uma importante equipe universitária de futebol americano. A única coisa que importa para ele é vencer os jogos. Suas áreas de foco são o recrutamento de jogadores do ensino médio, a adaptação deles a seus sistemas de ataque e defesa e a obtenção de seu máximo esforço em dias de jogos. Eis uma amostra dos questionamentos em sua área de responsabilidade: Quem devemos recrutar? Que tipo de treinamento pode desenvolver suas habilidades? O quanto devo exigir fisicamente de nossos atletas? Quais são os pontos fortes e fracos dos adversários, e como identificamos suas tendências táticas?

Felizmente, sua equipe acaba de contratar uma nova assessora técnica, Dar Beranek, especializada em ajudar treinadores a tomar decisões táticas. Ela está trabalhando com uma equipe de estagiários dedicada a criar análises de adversários. Eles usaram o VT de um jogo com comentários do treinador para desenvolver um modelo de árvore de decisão em cascata (Figura 1.3) a fim de prever se a próxima jogada será uma corrida ou um passe. Para o coordenador defensivo, eles construíram mapas de calor (Figura 1.4) dos ataques via passes do adversário, ilustrando suas tendências de lançar para a direita ou para a esquerda dependendo da formação defensiva enfrentada. Por fim, desenvolveram uma análise de séries temporais (Figura 1.5) para jogadas explosivas (definidas como um ganho superior a 16 jardas para jogadas via passe ou superior a 12 jardas para jogadas via corrida). Para cada jogada, eles comparam o resultado final com suas próprias formações defensivas e com as formações ofensivas da outra equipe, o que ajuda o treinador Breedlove a reagir mais depressa a trocas de formação durante um jogo. Nos Capítulos 2 a 5 e no Capítulo 7, explicaremos em mais detalhes as técnicas analíticas que geraram essas figuras.

Um novo trabalho que Dar está fomentando envolve a aprimoramento dos modelos de recrutamento de atletas do ensino médio. A cada ano, a equipe concede,

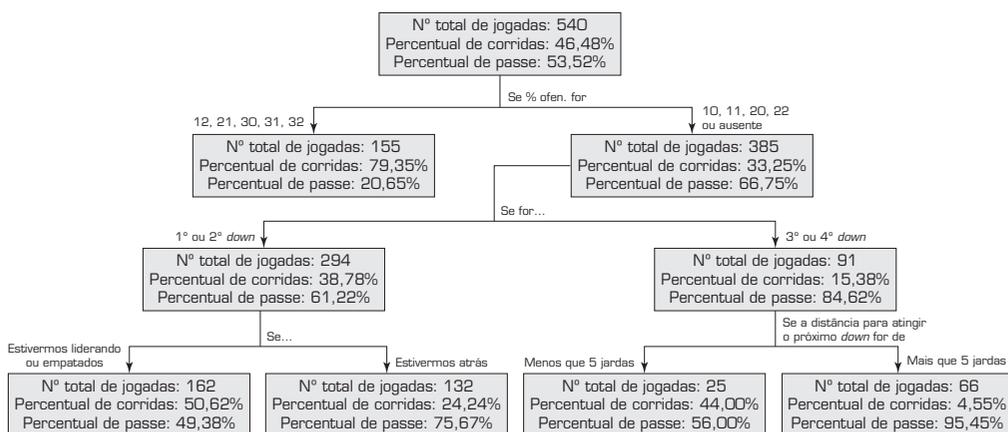


FIGURA 1.3 Árvore de decisão em cascata para jogadas de corrida ou de passe.

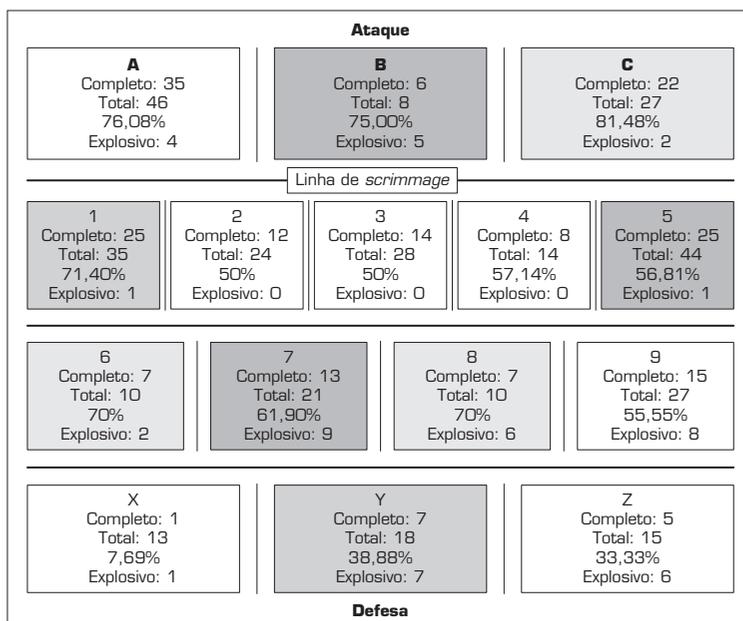


FIGURA 1.4 Análise de zonas por mapa de calor para jogadas via passe.

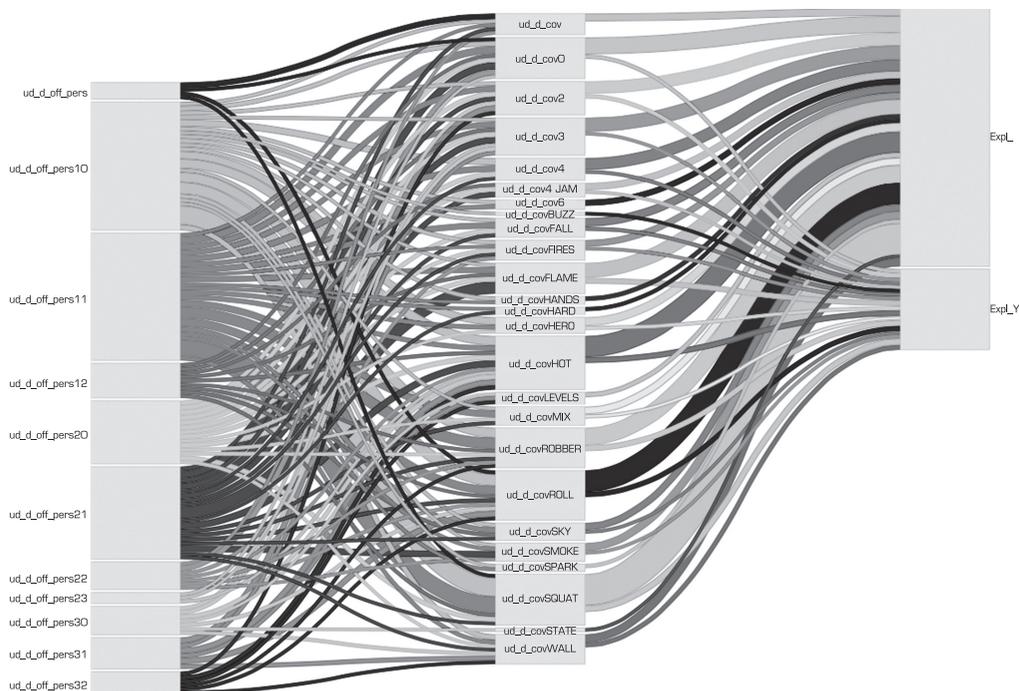


FIGURA 1.5 Análise por série temporal de jogadas explosivas.

por exemplo, três bolsas-atleta para estudantes que atuam na posição de *wide receiver*. Para escolher os melhores jogadores, Dar considera necessário ir além de dados como velocidade de corrida, impulsão ou comprimento dos braços; para ela, novos critérios devem ser incluídos, como velocidade de rotação da cabeça para receber um passe, tempo de reação a diversos estímulos e precisão em trajetos de corrida para recepção de passe. Algumas de suas ideias ilustrando esses conceitos aparecem no site da TUN; procure por BSI Case of Precision Football.⁴

Exemplo 3: o preparador físico

O Dr. Dan Johnson é o preparador físico de uma equipe universitária de futebol feminino. Seu trabalho é ajudar as jogadoras a evitarem lesões e aconselhar a equipe técnica quanto à carga de treinamento das atletas. Ele também tem um interesse no bem-estar das jogadoras, incluindo suas horas de sono e de descanso entre sessões leves e pesadas de treino. A meta é garantir que as atletas estejam preparadas para os dias de jogo com a máxima eficiência.

E com o advento dos dispositivos portáteis de rastreamento, há muito mais dados para o Dr. Dan analisar. Suas jogadoras treinam usando coletes que contêm sensores capazes de medir dados internos como frequência cardíaca e respiratória e temperatura corporal. Os coletes também incluem acelerômetros que medem dados externos como distância percorrida e velocidade, bem como acelerações e desacelerações. Ele sabe quais jogadoras estão dando seu máximo durante os treinos e quais não estão.

Atualmente, seu foco são as pesquisas que preveem ou previnem lesões em atletas (Figura 1.6). Algumas tarefas simples, como o Teste de Permanência em Agachamento numa Única Perna – ficar parado num pé só, depois no outro –, com diferenciais de pontuação superiores a 10% podem fornecer importantes informações sobre forças e fraquezas no tronco (Figura 1.7). Se uma atleta sofre uma forte pancada durante uma partida, o preparador físico pode realizar um teste do lado de fora do campo para conferir a reação a um estímulo num dispositivo móvel, o que complementa protocolos tradicionais de detecção de concussão. Sensores de sono mostram quem está obtendo repouso adequado (e quem passa a noite inteira na farrá). Ele solicita exames de ressonância magnética no laboratório universitário para saber quais atletas apresentam risco de lesão cerebral.

QUESTÕES SOBRE ESSES EXEMPLOS

1. Quais são os três fatores que podem fazer parte de um modelo preditivo (MP) para renovações de ingressos de temporada?
2. Quais são as duas técnicas que as equipes de futebol americano podem usar como análise de adversários?
3. Como os dispositivos portáteis de rastreamento previnem lesões e aumentam a segurança dos atletas? Quais novos dados analíticos os preparadores físicos podem usar?
4. Quais outras aplicações da análise de dados você consegue vislumbrar nos esportes?

⁴Business Scenario Investigation BSI: The Case of Precision Football (video). (Fall 2015). Aparece em <http://www.teradatauniversitynetwork.com/About-Us/Whats-New/BSI-Sports-Analytics-Precision-Football//Fall 2015>. (Acessado em setembro de 2016.)



FIGURA 1.6 Modelos de lesões no futebol.⁵

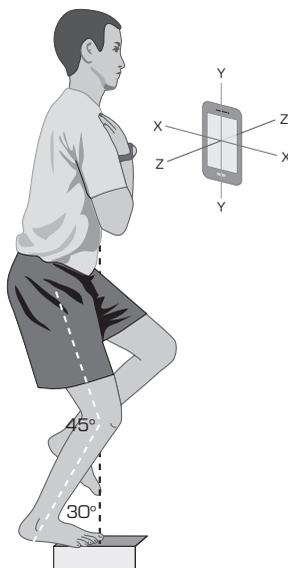


FIGURA 1.7 Teste de Permanência em Agachamento numa Única Perna – teste de força corporal no tronco. (Fonte: figura adaptada de Gary Wilkerson e Ashish Gupta.)

⁵“Women’s Soccer Injuries,” National Center for Catastrophic Sports Injury Research Report, NCAA. As compilações de dados factuais sobre lesões da NCAA Sport Injury são produzidas pela Datalys Center for Sports Injury Research and Prevention em colaboração com a National Collegiate Athletic Association e com a STOP Sports Injuries. Aparece em https://www.ncaa.org/sites/default/files/NCAA_W_Soccer_Injuries_WEB.pdf. (Acessado em novembro de 2016.)

O que podemos aprender com essas vinhetas?

Além dos analistas de departamentos administrativos, dos treinadores, preparadores físicos e dos especialistas em desempenho, há muitos outros envolvidos no esporte que utilizam dados, desde zeladores de campos de golfe que aferem as condições do solo e da grama para torneios do PGA, até árbitros de beisebol e basquete que são avaliados de acordo com os erros e acertos que cometem. Na verdade, é difícil achar uma área esportiva que *não* esteja sendo afetada pela disponibilidade de mais dados, sobretudo aqueles advindos de sensores.

As habilidades que você aprenderá neste livro em análise de negócios também se aplicarão aos esportes. Caso queira se aprofundar nessa área, sugerimos que consulte a seção Sports Analytics da Teradata University Network (TUN), um recurso gratuito para docentes e discentes. Nesse site, você encontrará descrições do que ler para aprender mais a respeito de análise de dados esportivos, compilações de locais onde poderá encontrar conjuntos de dados publicamente disponíveis para análise, bem como exemplos de projetos estudantis em análise de dados esportivos e entrevistas com profissionais da área que usam análise de dados em seus trabalhos. Boa sorte em seus estudos de análise de dados!

Fonte e créditos: com a colaboração do Dr. Dave Schrader, que se aposentou após 24 anos em desenvolvimento avançado e marketing na Teradata. Ele permanece no Conselho Diretor da Teradata University Network, onde desfruta de sua aposentadoria ajudando estudantes e professores a aprenderem mais sobre análise de dados esportivos. As imagens de futebol americano (Figuras 1.3 – 1.5) foram construídas por Peter Liang e Jacob Pearson, pós-graduandos da Oklahoma State University, como parte de um projeto estudantil no primeiro semestre de 2016. As gravuras de treinamento (Figuras 1.6 e 1.7) foram adaptadas de imagens fornecidas pelo prof. Gary Wilkerson, da University of Tennessee, em Chattanooga, e pelo prof. Ashish Gupta, da Auburn University.

1.2 Rumos dos ambientes de negócios e novas exigências para apoio à decisão e análise de dados

A vinheta de abertura ilustra como um setor inteiro pode aplicar análise de dados para desenvolver relatórios sobre o que está acontecendo, prever o que deve acontecer e então tomar decisões para tirar o melhor proveito da situação que se apresenta. Tais etapas exigem uma organização para reunir e analisar vastas coleções de dados. A partir de suas aplicações tradicionais em funções de folha de pagamento e contabilidade, sistemas computadorizados já penetraram em áreas gerenciais complexas, desde projeto e gerenciamento de fábricas automatizadas até a aplicação de métodos analíticos para a avaliação de propostas de fusões e aquisições. Praticamente todos os executivos sabem que as tecnologias da informação são vitais para seus negócios e já as adotam de forma extensiva.

Aplicações computadorizadas passaram de atividades de processamento e monitoramento de transações para tarefas de análise e solução de problemas, e boa parte disso por meio de tecnologias na nuvem, em muitos casos acessadas via dispositivos móveis. Ferramentas de análise de dados e BI como armazenamento de dados, mineração de dados, processamento analítico online (OLAP – *online analytical processing*), *dashboards* e uso de sistemas baseados na nuvem para apoio a decisões são os pilares da gestão moderna. Os gestores precisam dispor de redes de sistemas

informatizados em alta velocidade (via cabo ou sem fio) para auxiliá-los em sua tarefa mais importante: tomar decisões. Em muitos casos, tais decisões estão sendo rotineiramente automatizadas, eliminando-se a necessidade de qualquer intervenção gerencial.

Além dos óbvios avanços em hardware, software e capacidades de rede, alguns desenvolvimentos contribuíram claramente para aprimorar o apoio a decisões e a análise de dados de diversas maneiras, incluindo:

- **Comunicação e colaboração em grupo.** Atualmente, muitas decisões são tomadas por grupos cujos membros estão geograficamente dispersos. Os membros podem colaborar entre si e se comunicar prontamente usando ferramentas colaborativas, além dos onipresentes *smartphones*. A colaboração é especialmente importante ao longo da cadeia de suprimento, na qual parceiros – desde fornecedores até clientes – precisam compartilhar informações. A reunião de um grupo de tomadores de decisão, sobretudo de especialistas, em um mesmo local pode ser dispendiosa. Sistemas de informação podem aprimorar o processo colaborativo de um grupo e permitir que seus membros se situem em locais diferentes (economizando custos de deslocamento). Acima de tudo, tal colaboração ao longo da cadeia de suprimento permite que os fabricantes fiquem a par de novos padrões de demanda quase em tempo real, levando-os a reagir mais depressa a alterações no mercado.
- **Avanços no gerenciamento de dados.** Muitas decisões envolvem computações complexas. Os dados para isso podem ser armazenados em diferentes bases de dados em qualquer lugar da organização, ou mesmo fora dela. Os dados podem incluir texto, áudio, imagens e vídeo, e podem estar em diferentes idiomas. Muitas vezes, é preciso transmitir dados rapidamente a partir de locais distantes. Os sistemas atuais são capazes de pesquisar, armazenar e transmitir os dados necessários de forma ágil, econômica, segura e transparente.
- **Gerenciamento de gigantescos *data warehouses* e Big Data.** Grandes *data warehouses*, como aqueles operados pela Walmart, contêm quantidades desconhecidas de dados. Métodos especiais, incluindo computação paralela, Hadoop/Spark e assim por diante, estão disponíveis para organização, pesquisa e mineração de dados. Os custos relativos a armazenamento e mineração de dados estão em franco declínio. As tecnologias incluídas na ampla categoria de Big Data possibilitam a transmissão de uma profusão de dados provenientes de várias fontes e em diversas formas, o que permite um cenário bastante diferente de desempenho organizacional que não era possível no passado.
- **Suporte analítico.** Com mais tecnologias de dados e análise, mais alternativas podem ser avaliadas, previsões podem ser aprimoradas, análises de risco podem ser aceleradas e as opiniões de especialistas (que podem estar em locais remotos) podem ser coletadas com agilidade e a um custo reduzido. Conhecimentos especializados podem ser até mesmo derivados de sistemas analíticos. Com tais ferramentas, os tomadores de decisões podem conduzir simulações complexas, conferir muitos cenários e aferir impactos diversos de forma rápida e econômica. Este, é claro, é o foco de muitos capítulos neste livro.
- **Superação de limites cognitivos no processamento e armazenamento de informações.** Segundo Simon (1977), a mente humana apresenta uma capacidade limitada para processar e armazenar informações. Devido a seus limites cognitivos, as pessoas às vezes encontram dificuldade de recordar e utilizar informações sem cometer equívocos. O termo *limites cognitivos* indica que a capacidade de

resolução de problemas de um indivíduo revela-se limitada quando uma ampla gama de informações e conhecimentos diversos se faz necessária. Sistemas computadorizados possibilitam que as pessoas superem seus limites cognitivos ao acessarem e processarem rapidamente vastas quantidades de informações armazenadas.

- **Gestão de conhecimento.** As organizações reúnem vastas coleções de informações a respeito de suas operações, seus clientes, procedimentos internos, interações com funcionários e assim por diante, por meio das comunicações estruturadas e não estruturadas que ocorrem entre as várias partes interessadas. Sistemas de gestão de conhecimento se tornaram fontes de embasamento formal e informal para decisões gerenciais, embora por vezes nem sejam chamados por esse nome. Tecnologias como análise de texto e o IBM Watson estão tornando possível a geração de valor a partir de tais coleções de conhecimento.
- **Suporte a qualquer hora, em qualquer lugar.** Usando tecnologia sem fio, gestores conseguem acessar informações a qualquer hora e em qualquer lugar, além de analisá-las, interpretá-las e comunicá-las aos envolvidos. Essa talvez tenha sido a maior mudança ocorrida nos últimos anos. A velocidade com que as informações precisam ser processadas e convertidas em decisões alterou em muito as expectativas tanto por parte dos clientes quanto das empresas. Essas e outras funcionalidades vêm motivando o uso de suporte computadorizado para decisões desde o fim dos anos 60, mas com maior vigor desde meados dos anos 90. O crescimento das tecnologias móveis, das plataformas de redes sociais e das ferramentas analíticas viabilizou um novo nível de suporte por sistemas de informação (SI) para os gestores. Esse crescimento na oferta de embasamento de dados para qualquer decisão se estende não apenas aos gestores, mas também aos consumidores. Para começar, estudaremos uma visão geral das tecnologias que vêm sendo amplamente referidas como BI. A partir daí, ampliaremos nossos horizontes para introduzir vários tipos de análise de dados.

SEÇÃO 1.2 QUESTÕES DE REVISÃO

1. Quais são algumas das tendências fundamentais baseadas em sistemas que levaram as decisões apoiadas em SI a um novo patamar?
2. Liste algumas funcionalidades dos sistemas de informação capazes de facilitar o processo decisório gerencial.
3. Como um computador pode ajudar os humanos a superar seus limites cognitivos?

1.3 Evolução do apoio computadorizado a decisões até a análise/ciência de dados

A linha do tempo na Figura 1.8 exhibe a terminologia utilizada para descrever a análise de dados desde os anos 70. Durante essa década, o foco dos sistemas de informação para apoiar decisões era o fornecimento de relatórios estruturados e periódicos que os gestores podiam usar (ou ignorar) ao tomar decisões. As empresas começaram a criar relatórios rotineiros para apoiar os tomadores de decisões (gestores) sobre o que ocorrera no período prévio (como um dia, semana, mês ou trimestre anterior). Embora fosse útil saber o que ocorrera no passado, os gestores queriam mais do que isso: precisavam de uma variedade de relatórios em diferentes níveis de granularidade

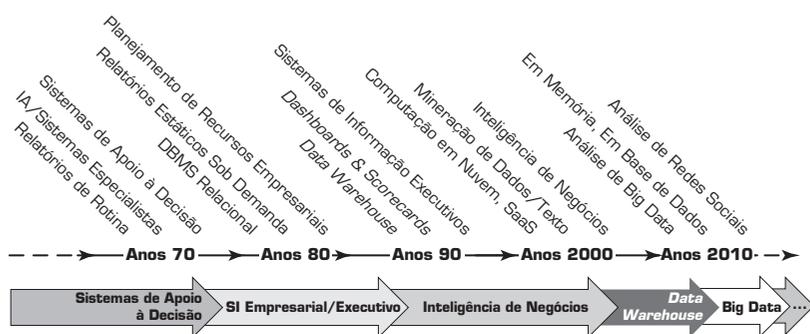


FIGURA 1.8 Evolução do apoio à decisão, inteligência de negócios e análise de dados.

para melhor entender e abordar novas exigências e desafios para a empresa. Eles costumavam ser chamados de sistemas de informação gerencial (MIS – *management information systems*). No início da década de 1970, Scott-Morton articulou pela primeira vez os principais conceitos de DSS (*decision support system*, ou sistema de apoio à decisão). Ele definiu os DSSs como “sistemas computadorizados interativos, que ajudam os tomadores de decisões a utilizar *dados* e *modelos* para solucionar problemas não estruturados” (Gorry and Scott-Morton, 1971). O trecho a seguir é outra definição clássica de DSS, fornecida por Keen e Scott-Morton (1978):

Sistemas de apoio à decisão complementam os recursos intelectuais dos indivíduos com as capacidades do computador para melhorar a qualidade das decisões. Trata-se de um sistema de apoio computadorizado para tomadores de decisões gerenciais que lidam com problemas semiestruturados.

Observe que o termo *sistema de apoio à decisão*, assim como *sistema de informação gerencial* e vários outros termos no campo de TI, é uma expressão de livre conteúdo (ou seja, significa coisas diferentes para pessoas diferentes). Sendo assim, não existe uma definição universalmente aceita de DSS.

Durante os primórdios da análise de dados, os dados costumavam provir de especialistas de cada área usando processos manuais (como entrevistas e enquetes) para construir modelos matemáticos ou baseados em conhecimento para solucionar problemas restritos de otimização. A ideia era fazer o melhor com recursos limitados. Tais modelos de apoio à decisão costumavam ser chamados de pesquisa operacional (PO). Os problemas que eram complexos demais para ser resolvidos idealmente (usando-se técnicas de programação matemática linear ou não linear) eram abordados usando-se métodos heurísticos, como modelos de simulação. (Introduziremos isso como análise de dados prescritiva mais adiante neste capítulo e em mais detalhes no Capítulo 6.)

No fim dos anos 70 e início dos 80, além dos modelos maduros de PO que estavam sendo usados em diversos ramos e sistemas governamentais, uma nova e empolgante linha de modelos surgiu: sistemas especialistas baseados em regras. Tais sistemas prometiam capturar o conhecimento de especialistas num formato que os computadores podiam processar (via uma coleção de regras tipo “se-então-senão” ou heurísticas), e poderiam ser consultados da mesma forma que consultamos especialistas em cada área para identificar um problema estruturado e prescrever a solução mais provável. Os sistemas especialistas permitiam que escassos conhecimentos especializados fossem disponibilizados onde e quando necessário, usando um DSS “inteligente”.

Os anos 80 testemunharam uma mudança significativa no modo como as organizações capturavam dados empresariais. A antiga prática era ter múltiplos sistemas de informação desmembrados e feitos sob medida para capturar dados transacionais de diferentes departamentos ou funções dentro da organização (como contabilidade, marketing e vendas, finanças, produção). Na década de 1980, tais sistemas foram integrados na organização como um todo, seguindo o modelo atualmente chamado de planejamento de recursos empresariais (ERP – *enterprise resource planning*). As antigas estruturas marcadas pela representação sequencial e não padronizada de dados foram substituídas por sistemas de gerenciamento de banco de dados relacional (RDBM – *relational database management*). Tais sistemas aprimoraram a captura e o armazenamento de dados, bem como as relações entre campos de dados organizacionais, significativamente reduzindo, ao mesmo tempo, a replicação de informações. A necessidade de sistemas RDBM e ERP surgiu quando a integridade e a consistência dos dados se tornaram um problema, entravando a eficiência das práticas empresariais. Com o ERP, todos os dados de cada canto da empresa são coletados e integrados segundo uma estrutura consistente, fazendo com que cada parte da organização tenha acesso à única versão da verdade quando e onde for necessária. Além do advento de sistemas ERP, ou talvez devido a tais sistemas, os relatórios de dados empresariais passaram a ser empregados sob demanda e conforme a necessidade. Assim, os tomadores de decisões podiam determinar quando precisavam ou queriam criar relatórios especializados para investigar problemas e oportunidades organizacionais.

Já na década de 1990, a exigência de relatórios mais versáteis levou ao desenvolvimento dos sistemas de informação executivos (*executive information systems* – EIS; DSSs projetados e desenvolvidos para executivos e suas necessidades de tomada de decisões). Tais sistemas foram projetados como painéis gráficos e planilhas de resultados, para que pudessem ser visualmente atraentes e ao mesmo tempo se concentrassem nos indicadores básicos de desempenho a serem monitorados pelos tomadores de decisões. Para viabilizar relatórios com tamanha versatilidade sem botar a perder a integridade transacional dos sistemas de informações empresariais, foi necessário criar um patamar intermediário de dados na forma de um repositório DW (*data warehouse*) a fim de suportar especificamente os relatórios empresariais e o processo decisório. De uma hora para outra, a maioria das empresas de médio e grande porte adotou os *data warehouses* como sua plataforma para decisões de âmbito geral. Os *dashboards* e *scorecards* desenvolvidos em planilhas obtinham seus dados a partir de um DW, e, dessa forma, não atravancavam a eficiência dos sistemas de transações empresariais, costumeiramente chamados de sistemas ERP.

Na virada do milênio, os DSSs baseados em DW começaram a ser chamados de sistemas de BI. Conforme aumentava a quantidade de dados longitudinais acumulados nos DWs, o mesmo ocorria com as capacidades de hardware e software para acompanhar a rápida evolução das necessidades dos tomadores de decisões. Devido ao mercado competitivo globalizado, os tomadores de decisões precisavam de informações atualizadas em um formato bastante digerido para atacarem problemas empresariais e tirarem proveito das oportunidades do mercado com agilidade. Como os dados em um DW são atualizados de tempos em tempos, eles não refletem as últimas informações. Para dar conta deste problema de latência de informações, fornecedores de DW desenvolveram um sistema para atualizar os dados com maior frequência, o que levou aos termos *armazenamento de dados em tempo real* e, mais realisticamente, *armazenamento de dados no tempo certo*, o qual difere do anterior por adotar uma política de renovação de dados baseada no nível exigido de atualização

das informações (isto é, nem todos os dados precisam ser renovados em tempo real). Os DWs são muito grandes e ricos em recursos, e tornou-se necessário “minerar” os dados corporativos a fim de “revelar” pepitas novas e úteis de conhecimento para aprimorar processos e práticas empresariais, o que deu origem aos termos *mineração de dados* e *mineração de texto*. Com volumes e variedades cada vez maiores de dados, surgiram exigências de maior armazenamento e mais poder de processamento. Se por um lado as grandes corporações dispunham dos meios para resolver este problema, as empresas de médio porte precisavam de modelos de negócio mais administráveis financeiramente. Essa exigência levou a arquiteturas voltadas para serviços e a fornecimento de análise de dados na forma de prestação de serviços. Empresas menores ganharam, portanto, acesso a recursos de análise de dados segundo um modelo sob demanda, passando a pagar apenas pelo que precisavam usar, em vez de investirem recursos financeiramente proibitivos em hardware e software.

Na década de 2010, estamos assistindo a mais uma quebra de paradigma no modo como os dados são capturados e usados. Em grande parte pela difusão do uso da Internet, novos meios de geração de dados emergiram. Dentre todas as novas fontes de dados (como etiquetas de identificação de rádio frequência [RFID], medidores de energia digital, *logs* na Web do tipo *clickstream*, aparelhos domésticos inteligentes, equipamentos portáteis de monitoramento de saúde), talvez as mais interessantes e desafiadoras sejam as redes sociais. Esses dados não estruturados são ricos em conteúdo informativo, mas a análise de tais fontes de dados impõe desafios consideráveis para sistemas computadorizados, tanto em termos de software quanto hardware. Recentemente, o termo *Big Data* foi cunhado para destacar os desafios que essas novas fontes de dados trouxeram para nós. Muitos avanços em hardware (tais como processamento paralelo massivo com enorme memória computacional e sistemas computadorizados com multiprocessador altamente paralelo) e em software/algoritmos (tais como Hadoop com MapReduce e NoSQL) foram desenvolvidos para enfrentar os desafios de Big Data.

É difícil prever o que a próxima década trará e quais serão os novos termos relacionados à análise de dados. O tempo decorrido entre novas quebras de paradigma em sistemas de informação e especialmente em análise de dados vem encolhendo, e essa tendência deve continuar no futuro próximo. Muito embora a análise de dados não seja novidade, a explosão em sua popularidade é bastante nova. Graças à recente explosão em Big Data, nos modos de coletar e armazenar esses dados e nas ferramentas de software intuitivas, diagnósticos embasados por dados estão mais acessíveis do que nunca a profissionais de negócios. Portanto, em meio a concorrência global, gestores se deparam com uma incrível oportunidade de tomarem melhores decisões usando análise de dados para aumentar receitas, diminuir custos, desenvolver produtos melhores, melhorar a experiência dos clientes, rastrear fraudes antes que aconteçam, além de aumentar o engajamento dos consumidores por meio de segmentação e customização. Hoje em dia, cada vez mais empresas estão preparando seus funcionários para aplicarem técnicas de análise de negócios a fim de melhorarem a eficiência e a eficácia de seus processos decisórios cotidianos.

A próxima seção tem como foco um sistema referencial para BI. Embora a maioria das pessoas provavelmente concorde que a BI evoluiu para se tornar análise de dados e ciência de dados, muitos fornecedores e pesquisadores ainda usam esse termo. Assim, a Seção 1.4 faz uma homenagem a essa história concentrando-se especificamente naquilo que vem sendo chamado de BI. Após a próxima seção, introduziremos a análise de dados e usaremos este termo para classificar todos os conceitos relacionados.

SEÇÃO 1.3 QUESTÕES DE REVISÃO

1. Liste três dos termos predecessores de análise de dados.
2. Qual era a principal diferença entre os sistemas denominados MIS, DSS e Sistemas de Informação Executivos?
3. O DSS evoluiu para se tornar BI ou vice-versa?

1.4 Um quadro referencial para inteligência de negócios

Os conceitos de apoio à decisão apresentados nas Seções 1.2 e 1.3 foram implementados aos poucos, sob nomes diferentes, por muitos fornecedores que criaram ferramentas e metodologias para apoiar decisões. Conforme destacado na Seção 1.3, quando os sistemas passaram a englobar empresas inteiras, os gestores puderam acessar relatórios práticos que os ajudavam a tomar decisões mais depressa. Tais sistemas, que costumavam ser chamados de EISs, passaram então a oferecer capacidades adicionais de visualização, alertas e mensurações de desempenho. Em 2006, os principais produtos comerciais já apareciam sob o termo inteligência de negócios (BI).

Definições de BI

Inteligência de negócios (BI – *business intelligence*) é um termo guarda-chuva que combina arquiteturas, ferramentas, bases de dados, ferramentas analíticas, aplicativos e metodologias.* Assim como DSS, trata-se de uma expressão de livre conteúdo, com significados diferentes de uma pessoa para outra. Parte da confusão provém da enxurrada de siglas e expressões associadas, como gestão de desempenho de negócios (BPM – *business performance management*). O principal objetivo do BI é possibilitar acesso interativo (às vezes em tempo real) a dados, permitir a manipulação de dados e oferecer a gestores empresariais e analistas a capacidade de conduzir análises apropriadas. Ao analisarem dados, situações e desempenhos históricos e atuais, os tomadores de decisões obtêm vislumbres valiosos que lhes permitem tomar decisões mais embasadas e melhores. O processo de BI baseia-se na *transformação* de dados em informações, depois em decisões e por fim em ações.

Uma breve história da BI

O termo *BI* foi cunhado pelo Gartner Group em meados dos anos 90. No entanto, conforme a história apresentada na seção anterior destacou, o conceito é bem mais antigo; suas raízes remontam aos sistemas de informação gerenciais (MIS) geradores de relatórios dos anos 70. Durante esse período, estes sistemas de informação geravam relatórios estáticos, bidimensionais e sem capacidade analítica. No início da década de 1980, o conceito de EISs emergiu. Tal conceito ampliou o suporte computadorizado aos gestores e executivos do primeiro escalão. Dentre as funcionalidades introduzidas estavam extração dinâmica de relatórios multidimensionais (*ad hoc* ou sob demanda), geração de previsões, análise de tendências, aprofundamento em detalhes, acesso a *status* e fatores cruciais de sucesso. Esses recursos apareceram em dezenas de produtos comerciais até meados dos anos 90.

*N. de R.T.: Neste livro, fazemos distinção entre o sistema e a técnica ou tecnologia BI. Assim, é utilizado “o BI” quando se tratar do sistema BI e “a BI” quando relacionado à técnica ou tecnologia.

Em seguida, as mesmas funcionalidades e algumas novas apareceram sob o nome de BI. Atualmente, um bom sistema empresarial de informação baseado em BI contém todas as informações de que os executivos precisam. Dessa forma, o conceito original de EIS foi transformado em BI. Em 2005, sistemas de BI passaram a incluir funcionalidades de inteligência artificial, bem como poderosos recursos analíticos. A Figura 1.9 ilustra as diversas ferramentas e técnicas que podem ser incluídas em um sistema BI. Ela ilustra também a evolução do BI. As ferramentas mostradas na Figura 1.9 apresentam as funcionalidades do BI. Os produtos mais sofisticados de BI incluem um maior número de tais funcionalidades; outros se especializam em apenas algumas delas.

A arquitetura do BI

Um sistema de BI apresenta quatro componentes principais: um *DW*, com seus dados-fonte; *análise de negócios*, uma coleção de ferramentas para manipular, minerar e analisar os dados no *DW*; *BPM* para monitorar e analisar desempenhos; e uma *interface do usuário* (como um *dashboard*). A relação entre esses componentes está ilustrada na Figura 1.10.

As origens e motivações do BI

De onde vieram as abordagens modernas a armazenamento de dados e BI? Quais são suas raízes, e como essas raízes afetam o modo como as organizações estão gerindo essas iniciativas atualmente? Hoje em dia, os investimentos em tecnologia da informação estão sob crescente escrutínio em termos de impacto e potencial sobre o balanço financeiro. O mesmo vale para *DW* e os aplicativos de BI que tornam essas iniciativas possíveis.

As organizações estão sendo levadas a capturar, compreender e aproveitar seus dados para embasar decisões capazes de aprimorar operações empresariais. Leis e

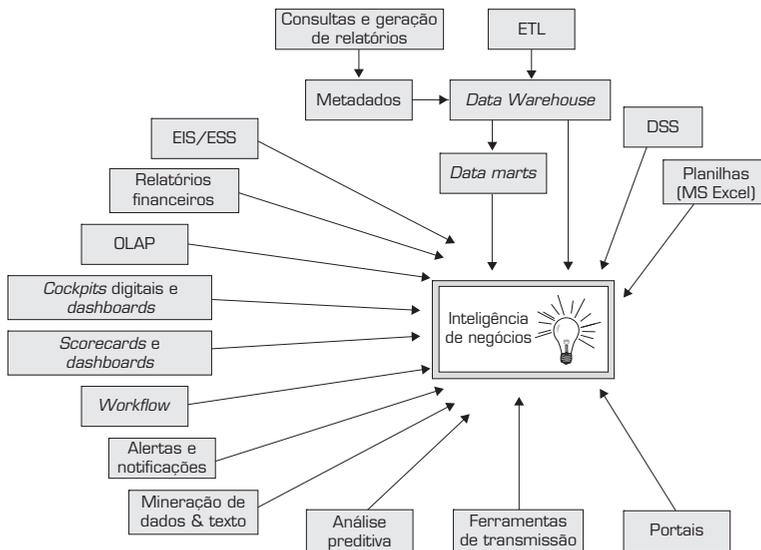


FIGURA 1.9 Evolução da inteligência de negócios (BI).

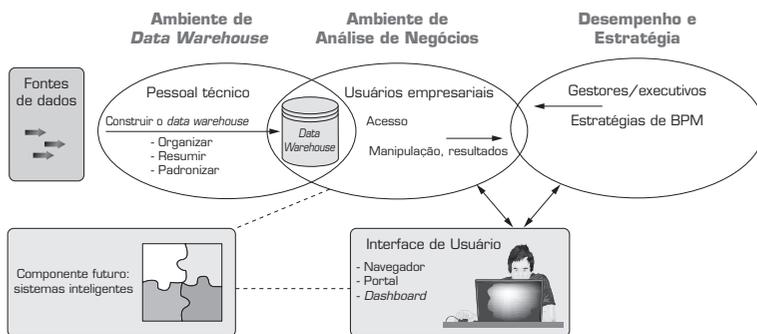


FIGURA 1.10 Uma arquitetura de BI em alto nível. (Fonte: baseado em W. Eckerson, *Smart Companies in the 21st Century: The Secrets of Creating Successful Business Intelligent Solutions*. The Data Warehousing Institute, Seattle, WA, 2003, p. 32, Illustration 5.)

regulamentações (como a Sarbanes-Oxley Act, de 2002) passaram a exigir que líderes empresariais documentem seus processos comerciais e ratifiquem a legitimidade das informações que os embasam e que são repassadas a partes interessadas. Ademais, os ciclos empresariais estão cada vez mais comprimidos; tomar decisões mais ágeis, mais embasadas e melhores é, portanto, um imperativo competitivo. Os gestores precisam das *informações certas na hora certa e no lugar certo*. Este é o mantra das abordagens modernas de BI.

As organizações precisam atuar com inteligência. Prestar bastante atenção à gestão das iniciativas de BI é um aspecto necessário dos negócios. Não chega a surpreender, então, que as organizações estejam capitaneando cada vez mais o BI, ou sua nova encarnação, a análise de dados. O Caso Aplicado 1.1 ilustra uma tal aplicação de BI que ajudou muitas companhias aéreas, além, é claro, das empresas que prestaram tais serviços a elas.

Caso aplicado 1.1

A Sabre ajuda seus clientes por meio de *dashboards* e análise de dados

A Sabre é uma das líderes mundiais no ramo de viagens, prestado tanto serviços do tipo empresa-cliente quanto empresa-empresa. Ela atende viajantes, agências de viagem, corporações e fornecedores do setor por meio de suas quatro empresas principais: Travelocity, Sabre Travel Network, Sabre Airline Solutions e Sabre Hospitality Solutions. O atual ambiente volátil da economia global impõe desafios competitivos consideráveis para o setor aeronáutico. Para se manter à frente da concorrência, a Sabre Airline Solutions reconheceu que

executivos de companhias aéreas precisavam de ferramentas mais poderosas para administrar suas decisões de negócios, eliminando o processo tradicional, manual e demorado de agregar informações financeiras e afins necessárias para iniciativas práticas. Tais ferramentas permitem um apoio à decisão em tempo real em companhias aéreas do mundo inteiro, maximizando seus retornos (e, por sua vez, também os da Sabre) e o valor oferecido aos clientes pelo uso crescente de dados de diagnóstico e direcionamento.

(Continua)

Caso aplicado 1.1 (Continuação)

A Sabre desenvolveu um *Data Warehouse* Empresarial de Viagem (ETDW – Enterprise Travel Data Warehouse) usando a Teradata para manter suas reservas descomuns de dados. O ETDW é atualizado quase em tempo real com lotes que rodam a cada 15 minutos, reunindo dados de todos os empreendimentos da Sabre. A empresa utiliza seu ETDW para criar *Dashboards* Executivos Sabre que proporcionam vislumbres quase em tempo real usando a plataforma Cognos BI com infraestruturas tecnológicas Oracle Data Integrator e Oracle Goldengate. Os *Dashboards* Executivos oferecem aos altos gestores de suas companhias aéreas clientes uma solução ágil, automatizada e fácil de usar, agregando parâmetros cruciais de desempenho de um modo sucinto e fornecendo uma visão em 360° da saúde em geral de suas empresas. Numa dessas companhias aéreas, os *Dashboards* Executivos da Sabre fornecem aos gestores seniores um resumo diário e intradiário de indicadores-chave em um único aplicativo, substituindo o processo semanal que levava 8 horas para gerar o mesmo relatório a partir de várias fontes de dados. E o uso de *dashboards* não se limita aos clientes externos; a Sabre também os utiliza em seus próprios levantamentos de desempenho operacional interno.

Os *dashboards* ajudam os clientes da Sabre a obterem uma clara compreensão dos dados por meio de suas exibições visuais que incorporam capacidades de aprofundamento. Isso substitui apresentações superficiais e possibilita uma revisão detalhada dos dados com menos esforço e em menos tempo. Ademais, facilita o diálogo em equipe ao disponibilizar os dados/parâmetros referentes a desempenho de vendas para muitos interessados, incluindo bilhetagem, assentos comprados e voados, desempenho operacional, como movimentação e rastreamento de voos, reservas, estoques e faturamento em múltiplos canais de distribuição da companhia. Os sistemas de *dashboards*

oferecem infraestrutura escalável, suporte geográfico a interface de usuário, integração de dados e agregação, para permitir que os executivos sejam mais proativos ao tomarem medidas positivas para a saúde em geral da companhia aérea.

Com seu ETDW, a Sabre também foi capaz de desenvolver outras soluções baseadas na Web para análises e extração de relatórios, tirando proveito de dados para investigar perfis de clientes e outras interações de vendas a fim de aumentar o valor oferecido pela empresa. Isso permite uma melhor segmentação de clientes e agregação de valor aos serviços.

Questões para discussão

1. Como se dá a extração tradicional de relatórios? Como ela é usada na organização?
2. Como a análise de dados pode ser usada para transformar a extração tradicional de relatórios?
3. De que forma a extração interativa de relatórios pode auxiliar as organizações em seu processo decisório?

O que podemos aprender com este caso aplicado

Este caso mostra que organizações que costumavam usar relatórios apenas para rastrear atividades empresariais internas e cumprir com normas de regulamentação estipuladas pelo governo agora estão abordando cientificamente seus dados empresariais transacionais. As possibilidades de geração de relatórios se ampliaram, já que as organizações estão tentando analisar seus dados transacionais arquivados para entender as tendências e padrões subjacentes e ocultos, permitindo-lhes diagnosticar áreas problemáticas e tomar melhores decisões para resolvê-las, além de aproveitar oportunidades presentes e futuras no mercado. A extração de relatórios evoluiu e passou a incluir interatividade online, permitindo que usuários gerem e criem relatórios instantâneos

e customizados, munidos de ferramentas de visualização e de conexão com a base de dados, com capacidades de aprofundamento em dados resumidos.

Fonte: Teradata.com, “Sabre Airline Solutions,” Terry, D. (2011), “Sabre Streamlines Decision Making,” <http://www.teradatamagazine.com/v11n04/Features/Sabre-Streamlines-Decision-Making/> (Acessado em julho de 2016.)

Um exercício multimídia em inteligência de negócios

A TUN inclui vídeos (similares ao seriado de televisão *CSI*) para ilustrar conceitos da análise de dados em diferentes ramos. Eles se chamam “Vídeos BSI” (Business Scenario Investigations, ou Investigações de Cenários de Negócios). Além de serem divertidos, também suscitam questões para discussão em sala de aula. Para começar, acesse <http://www.teradatauniversitynetwork.com/Library/Items/BSI-The-Case-of-the-Misconnecting-Passengers> ou www.youtube.com/watch?v=NXEL5F4_aKA. Assista ao vídeo no YouTube. Basicamente, você precisa assumir o papel de encarregado de uma central de serviço ao cliente. Um voo está com sua chegada atrasada, e vários passageiros podem acabar perdendo suas conexões. Há assentos em um voo de partida que podem acomodar dois dos quatro passageiros. Quais dois passageiros devem receber prioridade? Você recebe informações sobre os perfis dos clientes e suas relações com a companhia aérea. É provável que sua decisão se modifique conforme você for descobrindo mais sobre os perfis dos clientes.

Assista ao vídeo, pause quando apropriado e responda as perguntas sobre quais passageiros devem receber prioridade. Em seguida, retome o vídeo para obter mais informações. Depois de encerrado o vídeo, você pode ver detalhes relacionados a ele e como a análise foi preparada em um conjunto de slides em www.slideshare.net/teradata/bsi-how-we-did-it-the-case-of-the-misconnecting-passengers.

Essa excursão multimídia fornece um exemplo de como informações adicionais disponibilizadas por uma empresa de DW são capazes de auxiliar na tomada de uma decisão.

Embora algumas pessoas considerem DSS e BI sinônimos, esses sistemas não são, atualmente, o mesmo. É interessante observar que algumas pessoas creem que o DSS é uma parte do BI – uma de suas ferramentas analíticas. Outras pensam que o BI é um caso especial do DSS que lida com extração de relatórios, comunicação e colaboração (uma forma de DSS voltada para dados). Outra explicação (Watson, 2005) é que o BI é resultado de uma revolução contínua, e, como tal, o DSS é um dos elementos originais do BI. Além do mais, como ressaltado da próxima seção em diante, em muitos círculos, o BI foi englobado pelos novos termos *análise de dados* ou *ciência de dados*.

Processamento de transações versus processamento analítico

Para ilustrar as principais características do BI, começaremos mostrando aquilo que o BI não é – ou seja, processamento de transações. Estamos familiarizados com sistemas de informação que suportam nossas transações, como saques em caixas eletrônicos, depósitos bancários, leitura de código de barras em mercados e assim por diante. Esses sistemas de *processamento de transações* estão constantemente lidando

com atualizações daquilo que chamamos de *bases de dados operacionais*. Numa transação de saque em caixa eletrônico, por exemplo, precisamos reduzir nosso saldo bancário de acordo ao valor sacado; já um depósito bancário soma-se a uma conta; e uma compra em um mercado acaba se refletindo no total de vendas da loja naquele dia, e deve refletir-se numa redução apropriada no seu estoque do produto adquirido; e assim por diante. Esses sistemas de **processamento de transações online (OLTP – *online transaction processing*)** lidam com os negócios cotidianos de uma empresa. Em contraste, um DW geralmente representa um sistema distinto que fornece armazenamento para dados que serão usados para *análise*. O objetivo é que tal análise acabe proporcionando à gestão a capacidade de vasculhar dados atrás de informações a respeito dos negócios, que podem ser usadas para embasar e agilizar decisões táticas ou operacionais. Forneceremos uma definição mais técnica de DW no Capítulo 2, mas basta afirmar que DWs são voltados a trabalhar com dados informatizados usados em sistemas de **processamento analítico online (OLAP – *online analytical processing*)**.

A maior parte dos dados operacionais em sistemas de planejamento de recursos empresariais (ERP – *enterprise resource planning*) – e seus componentes complementares como *gestão de cadeia de suprimento (SCM – supply chain management)* ou CRM – fica armazenada em um sistema OLTP, que é um tipo de processamento computadorizado em que o computador responde imediatamente a solicitações do usuário. Cada solicitação é considerada uma *transação*, que é um registro computadorizado de um evento discreto, como um recibo de estoque ou a encomenda de um cliente. Em outras palavras, uma transação exige um conjunto de duas ou mais atualizações de base de dados que precisam ser completadas de um modo “tudo ou nada”.

O próprio design que torna o sistema OLTP eficiente para processamento de transações o torna ineficiente para relatórios, consultas e análises *ad hoc* para usuários finais. Na década de 1980, muitos usuários empresariais se referiam a seus *mainframes* como “buracos negros”, porque todas as informações entravam neles, mas nenhuma retornava jamais. Todas as solicitações de relatórios tinham de ser programadas pelo pessoal de TI, ao passo que os relatórios “pré-enlatados” podiam ser gerados com agendamento, e consultas *ad hoc* em tempo real eram praticamente impossíveis. Embora os sistemas ERP baseados em cliente/servidor dos anos 90 tenham facilitado um pouco a extração de relatórios, ainda estavam longe de oferecer praticidade a usuários finais regulares sem conhecimentos técnicos em termos de produção de relatórios operacionais, análises interativas, etc. Para dirimir esses problemas, as noções de DW e BI foram criadas.

Os DWs contêm uma ampla variedade de dados que apresentam as condições empresariais em um certo instante no tempo. A ideia era criar uma infraestrutura de base de dados que permanecesse sempre online e que contivesse todas as informações dos sistemas OLTP, incluindo dados históricos, mas reorganizados e estruturados de modo a aumentar a agilidade e a eficiência para consultas, análises e embasamento de decisões. É a separação do OLTP e da análise e embasamento de decisões que gera os benefícios do BI que foram descritos anteriormente.

Planejamento e alinhamento apropriados com a estratégia de negócios

Em primeiro lugar, os motivos fundamentais para investir em BI precisam estar alinhados com a estratégia de negócios da empresa. A BI não pode ser um mero

exercício técnico para o departamento de sistemas informatizados. Ela precisa ser um agente de mudança do modo com que a empresa conduz seus negócios, aprimorando os processos comerciais e embasando os processos decisórios com dados mais concretos. Muitos consultores e praticantes de BI envolvidos em iniciativas bem-sucedidas na área sugerem que um quadro referencial para planejamento é uma pré-condição necessária. Um desses quadros referenciais, desenvolvido pela Gartner, Inc. (2004), decompõe planejamento e execução em componentes de *negócios, organização, funcionalidade e infraestrutura*. Nos âmbitos organizacional e de negócios, objetivos estratégicos e operacionais devem ser definidos ao se avaliar as habilidades organizacionais disponíveis para cumpri-los. Questões de cultura organizacional envolvendo iniciativas de BI e a motivação para tais iniciativas e procedimentos de compartilhamento intraorganizacional de melhores práticas de BI precisam ser considerados pelo alto escalão – com planos prontos para preparar a organização para as mudanças. Um dos primeiros passos nesse processo é avaliar a área de SI, o conjunto de habilidades das classes potenciais de usuários e se a cultura é passível de mudança. A partir desse levantamento, e assumindo-se que haja justificativa e necessidade de avançar, uma empresa pode preparar um plano detalhado de ação. Outro aspecto crucial para o sucesso da implementação de BI é a integração de diversos projetos de BI (a maioria das empresas utiliza vários deles ao mesmo tempo) entre si e com outros sistemas de TI da organização e junto a seus parceiros comerciais.

Se a estratégia da empresa estiver apropriadamente alinhada com os motivos para iniciativas de DW e BI, e se a área de SI da empresa for capaz de cumprir seu papel em tal projeto, e se a comunidade potencial de usuários estiver pronta e motivada, só então é recomendado iniciar o BI e estabelecer uma Central de Competência de BI dentro da empresa. A central pode cumprir algumas ou todas as seguintes funções (Gartner, 2004):

- A central pode demonstrar como o BI está claramente vinculada à estratégia e à execução da estratégia.
- A central pode estimular a interação entre as comunidades potenciais de usuários empresariais e a área de SI.
- A central pode servir como repositório e disseminador das melhores práticas de BI entre as diferentes linhas de negócios.
- Padrões de excelência em práticas de BI podem ser capitaneados e encorajados por toda a empresa.
- A área de SI pode aprender bastante ao interagir com as comunidades de usuários, como a respeito da variedade de tipos de ferramentas analíticas necessárias.
- A comunidade de usuários empresariais e a área de SI podem compreender melhor por que a plataforma de DW precisa ser flexível para atender às exigências empresariais em constante evolução.
- A central pode ajudar atores importantes, como altos executivos, a enxergarem como o BI pode cumprir um papel importante.

Outros fator importante para o sucesso do BI é sua capacidade de facilitar um ambiente em tempo real e sob demanda, introduzido a seguir.

BI em tempo real e sob demanda é possível

A demanda por acesso instantâneo e sob demanda a informações dispersas cresceu devido à necessidade cada vez maior de diminuir a lacuna entre os dados

operacionais e os objetivos estratégicos. Como resultado, uma categoria de produtos denominados *aplicativos de BI em tempo real* acabou surgindo. A introdução de novas tecnologias de geração de dados, como RFID e outros sensores, só faz acelerar esse crescimento e a exigência subsequente de BI em tempo real. Sistemas tradicionais de BI utilizam um grande volume de dados *estáticos* que foram extraídos, limpos e abastecidos em um DW a fim de produzir relatórios e análises. No entanto, além de relatórios, os usuários também precisam de monitoramento de negócios, análises de desempenho e um vislumbre de por que as coisas estão acontecendo. Tudo isso pode ajudar os usuários, que precisam ficar a par (praticamente em tempo real) das mudanças nos dados ou da disponibilidade de relatórios, alertas e notificações relevantes envolvendo eventos e tendências emergentes em aplicativos de mídia social. Além disso, aplicativos empresariais podem ser programados para agir dependendo das descobertas desses sistemas de BI em tempo real. Um aplicativo SCM, por exemplo, pode encaminhar automaticamente uma encomenda de mais mercadorias assim que o estoque em tempo real cair abaixo de um certo patamar, ou um aplicativo CRM pode automaticamente convocar um representante de vendas e um controlador de crédito para conferir um cliente que fez uma encomenda online superior a US\$10 mil.

Uma abordagem de BI em tempo real utiliza o modelo DW de sistemas tradicionais de BI. Nesse caso, produtos de fornecedores de plataformas inovadoras de BI oferecem uma solução de prestação de serviço em tempo real que povoa o DW muito mais depressa do que as típicas atualizações noturnas de lotes em *extrair/transformar/carregar* (veja o Capítulo 3). Uma segunda abordagem, geralmente chamada de *gestão de atividades empresariais* (BAM – *business activity management*), é adotada por fornecedores de BAM puro e/ou BAM híbrido de *middleware* (como Savvion, Iteration Software, Vitria, webMethods, Quantive, Tibco ou Vineyard Software). Ela deixa o DW totalmente de lado e utiliza **serviços Web** ou outros meios de monitoramento para encontrar eventos empresariais-chave. Esses **agentes inteligentes** podem ser colocados em um servidor em separado na rede ou nas próprias bases de dados de aplicativos transacionais, e podem usar abordagens baseadas em eventos ou processos para mensurar e monitorar processos operacionais de modo proativo e inteligente.

Desenvolvimento e aquisição de sistemas BI

Atualmente, muitos fornecedores oferecem ferramentas diversificadas, algumas das quais são completamente pré-programadas (chamadas de *shells*); tudo que você precisa fazer é inserir os seus números. Tais ferramentas podem ser compradas ou arrendadas. Para uma lista de produtos, demos, artigos técnicos e informações mais recentes, confira diretórios de produtos em tdwi.org. A inscrição gratuita de usuário é obrigatória. Quase todos os aplicativos de BI são construídos com *shells* oferecidos por fornecedores que podem criar por conta própria soluções customizadas para um cliente ou então trabalhar com outro fornecedor terceirizado. A questão que resta às empresas é qual alternativa selecionar: aquisição, aluguel ou desenvolvimento. Cada uma dessas alternativas inclui diversas opções. Um dos principais critérios para tomar tal decisão é a justificativa e análise de custo/benefício.

Justificativa e análise de custo/benefício

Com a multiplicação das aplicações potenciais de BI, maior ainda a necessidade de justificá-las e priorizá-las. Isso não é tarefa fácil, devido à grande quantidade de

benefícios intangíveis. Tanto benefícios diretos quanto indiretos têm de ser identificados. Obviamente, é aí que o conhecimento sobre aplicações similares em outras organizações e casos de estudo vem bem a calhar. O Data Warehousing Institute (tdwi.org), por exemplo, fornece uma profusão de informações sobre produtos e suas aplicações e implementações inovadoras. Tais informações podem ser úteis para estimar os benefícios diretos e indiretos.

Segurança e proteção de privacidade

Esta é uma questão extremamente importante no desenvolvimento de um sistema computadorizado, sobretudo em BI que contém dados que podem possuir valor estratégico. Além disso, a privacidade de funcionários e clientes precisa ser protegida.

Integração de sistemas e aplicativos

Com a exceção de alguns itens acessórios, todos os aplicativos de BI devem ser integrados com outros sistemas como bases de dados, sistemas legados, sistemas empresariais (especialmente ERP e CRM), *e-commerce* (lado das vendas, lado das compras) e muitos mais. Além disso, aplicativos de BI costumam ter conexão com a Internet, e muitas vezes com sistemas informatizados de parceiros comerciais.

Ademais, ferramentas de BI às vezes precisam ser integradas entre si, criando sinergia. A necessidade de integração levou fornecedores de software a incluir cada vez mais funcionalidades em seus produtos. Clientes que compram um pacote de software “tudo em um” lidam com um único fornecedor e não precisam se preocupar com conectividade entre sistemas. Porém, podem abrir mão da vantagem de criar sistemas formados pelos componentes mais especializados.

SEÇÃO 1.4 QUESTÕES DE REVISÃO

1. Defina *BI*.
2. Liste e descreva os principais componentes do BI.
3. Defina *OLTP*.
4. Defina *OLAP*.
5. Liste alguns dos tópicos de implementação abordados pelo relatório da Gartner.
6. Liste alguns dos fatores de sucesso do BI.

1.5 Visão geral da análise de dados

A expressão *análise de dados* acabou em grande parte substituindo os componentes individuais anteriores de tecnologias de apoio à decisão computadorizadas que estavam disponíveis sob vários rótulos no passado. Na verdade, muitos proponentes e acadêmicos utilizam hoje a expressão *análise de dados* no lugar de BI. Embora muitos autores e consultores trabalhem com definições ligeiramente diferentes, a **análise de dados** pode ser vista como o processo de desenvolvimento de decisões ou recomendações práticas para ações baseadas em vislumbres gerados por dados históricos. De acordo com o Institute for Operations Research and Management Science (Informs), a análise de dados representa a combinação de tecnologia computadorizada, técnicas de ciência administrativa e estatística para solucionar problemas reais. Muitas outras organizações, é claro, propuseram suas próprias interpretações e motivações

para análise de dados. O SAS Institute Inc., por exemplo, propôs oito níveis de análise de dados que começam por relatórios padronizados gerados por sistema computadorizado. Tais relatórios proporcionam essencialmente uma noção do que está acontecendo com uma organização. Tecnologias adicionais permitiram criar relatórios mais customizados que podem ser gerados de maneira *ad hoc*. A próxima extensão de relatórios nos leva a consultas ao estilo OLAP, que permitem que o usuário se aprofunde e determine fontes específicas de interesse ou oportunidades. As tecnologias disponíveis nos dias de hoje também podem emitir alertas automáticos para um tomador de decisões quando patamares de desempenho são ultrapassados. No âmbito dos consumidores, vemos tais alertas na meteorologia e em outras áreas. Mas alertas similares também podem ser gerados em configurações específicas quando as vendas ficam acima ou abaixo de um determinado patamar por um certo período, ou quando o estoque de um produto específico está minguando. Todos esses aplicativos são possibilitados por meio de análises e consultas em torno de dados sendo coletados por uma organização.

O próximo nível de análise pode envolver um aprofundamento estatístico para melhor compreender padrões. Isso pode ser levado ainda mais longe pelo desenvolvimento de modelos preditivos de como os clientes podem reagir a uma campanha de marketing específica ou a ofertas de produtos/serviços. Quando uma organização tem uma boa noção do que está acontecendo e do que deve vir a acontecer, ela também pode empregar outras técnicas para tomar melhores decisões dependendo das circunstâncias. Esses oito níveis de análise de dados são descritos em mais detalhes num artigo técnico da SAS (sas.com/news/sascom/analytics_levels.pdf).

Essa ideia de olhar para todos os dados a fim de entender o que está acontecendo, o que virá a acontecer e como tirar o melhor proveito disso também foi abarcada pelo Informs, ao propor três níveis de análise de dados. Esses três níveis são identificados (informs.org/Community/Analytics) como descritivo, preditivo e prescritivo. A Figura 1.11 apresenta uma visão gráfica desses três níveis de análise de dados. Ela sugere que esses três níveis representam etapas de certa forma independentes e

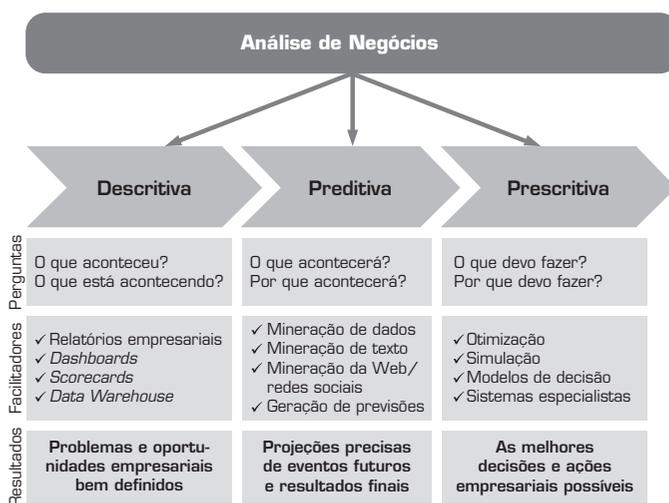


FIGURA 1.11 Três tipos de análise de dados.

que um tipo de aplicação de análise de dados leva a outro. Também sugere que há, na verdade, uma certa sobreposição entre esses três tipos de análise de dados. Seja como for, a natureza interconectada dos diferentes tipos de aplicações de análise de dados fica evidente. A seguir, introduziremos esses três níveis de análise de dados.

Análise de dados descritiva

A **análise de dados descritiva (ou de extração de relatórios)** diz respeito a conhecer o que está acontecendo na organização e entender tendências e causas subjacentes de tais ocorrências. Em primeiro lugar, isso envolve a consolidação de fontes de dados e a disponibilidade de todos os dados relevantes de um modo que permita a extração e a análise apropriadas de relatórios. Geralmente, o desenvolvimento dessa infraestrutura de dados faz parte dos DWs. A partir dessa infraestrutura de dados, podemos desenvolver relatórios, consultas, alertas e tendências apropriados usando ferramentas e técnicas de extração de relatórios.

Uma tecnologia significativa que se tornou parte fundamental dessa área é a visualização. Usando as mais recentes ferramentas de visualização do mercado, podemos agora obter retratos claros das operações de nossa organização. Os Casos Aplicados 1.2 e 1.3 destacam algumas dessas aplicações. Recursos de visualizações coloridas discutidas nessas aplicações estão disponíveis online ou no site dssbibook.com.

Caso aplicado 1.2

A Silvaris aprimora os negócios com análise visual e capacidades de geração de relatórios em tempo real

A Silvaris Corporation foi fundada no ano 2000 por uma equipe de profissionais do ramo florestal para fornecer avanços tecnológicos ao setor madeireiro e de material de construção. Ela é a primeira plataforma de comércio eletrônico nos Estados Unidos voltada especificamente para produtos florestais, e está sediada em Seattle, Washington. É líder no fornecimento em atacado de produtos industriais de madeira e materiais de construção reaproveitáveis.

A Silvaris vende seus produtos e presta serviços internacionais de logística para mais de 3.500 clientes. Para administrar os diversos processos envolvidos a cada transação, ela criou uma plataforma de comércio online para rastrear fluxos de informação relacionados a transações entre comerciantes, contabilidade, crédito e logística. Isso lhe permitiu

compartilhar suas informações em tempo real com seus clientes e parceiros. Porém, devido à forte flutuação no preço dos materiais, a Silvaris precisou obter um retrato em tempo real dos dados sem transformá-los em um formato separado para geração de relatórios.

Assim, a empresa passou a usar o Tableau, devido à sua capacidade de conexão e visualização de dados ao vivo. Aproveitando seus *dashboards* fáceis de entender e explicar, a Silvaris começou a usar o Tableau para gerar relatórios. Isso a ajudou a filtrar informações rapidamente e a identificar questões que afetavam seus negócios. Assim, ela passou a gerir encomendas online *versus* offline com sucesso usando os relatórios gerados pelo Tableau. Atualmente, a Silvaris faz um acompanhamento das encomendas online feitas por clientes, e sabe quando deve enviar lembretes

(Continua)

Caso aplicado 1.2 (Continuação)

de renovação para determinados clientes para mantê-los comprando online.

Além disso, usando o Tableau, seus analistas podem poupar tempo gerando *dashboards*, em vez de redigirem centenas de páginas de relatórios.

Questões para discussão

1. Qual foi o desafio enfrentado pela Silvaris?
2. Como a Silvaris resolveu seu problema usando visualização de dados com Tableau?

O que podemos aprender com este caso aplicado

Muitos setores precisam analisar dados em tempo real. Análises em tempo real permitem que os analistas identifiquem aspectos que afetam seus negócios. Às vezes, a visualização é a melhor maneira de começar a analisar fluxos de dados ao vivo. O Tableau é uma ferramenta de visualização de dados deste tipo, com a capacidade de analisar dados ao vivo sem precisar transformá-los em um formato separado para geração de relatórios.

Fontes: Tableau.com, “Silvaris Augments Proprietary Technology Platform with Tableau’s Real-Time Reporting Capabilities,” http://www.tableau.com/sites/default/files/case-studies/silvarisbusiness-dashboards_0.pdf (acessado em julho de 2016); Silvaris.com, “Overview,” <http://www.silvaris.com/About/> (acessado em julho de 2016).

Caso aplicado 1.3

A Siemens reduz custos com o uso de visualização de dados

A Siemens é uma empresa alemã sediada em Berlim e é uma das maiores empresas do mundo, centrada nas áreas de eletrificação, automação e digitalização. Seu faturamento anual é de 76 bilhões de euros.

O grupo de análise de dados visual da Siemens é encarregado de soluções de extração de relatórios ponta a ponta e de consultoria para todas as necessidades de BI da Siemens internacional. Esse grupo estava enfrentando o desafio de fornecer soluções de geração de relatórios para a organização como um todo, em seus diferentes departamentos, sem perder o equilíbrio entre capacidades de governança e de autoatendimento. A Siemens precisava de uma plataforma capaz de analisar múltiplas instâncias de enquetes de satisfação de clientes, processos de logística e relatórios financeiros. Tal plataforma devia ser fácil de usar para seus funcionários, para que pudessem aplicar os dados em análise e embasamento de decisões. Além disso, a plataforma devia ser facilmente integrada aos

sistemas já existentes da Siemens e proporcionar uma experiência sem percalços aos funcionários.

Eles começaram usando a Dundas BI, uma líder em fornecimento global de BI e soluções de visualização de dados. Isso permitiu que a Siemens criasse painéis altamente interativos, capazes de detectar problemas com antecedência e, assim, poupar cifras consideráveis de dinheiro. Os *dashboards* desenvolvidos pela Dundas BI ajudaram o grupo de logística global da Siemens a determinar, por exemplo, de que forma taxas diferentes de suprimento afetam a operação, ajudando, assim, a reduzir o ciclo logístico em 12% e os custos com resíduos em 25%.

Questões para discussão

1. Quais desafios foram enfrentados pelo grupo de análise de dados visual da Siemens?
2. De que forma a ferramenta de visualização de dados Dundas BI ajudou a Siemens a reduzir custos?

O que podemos aprender com este caso aplicado

Muitas organizações desejam ferramentas que possam ser usadas para analisar dados a partir de múltiplas divisões. Tais ferramentas podem

ajudar a melhorar o desempenho e a exibir dados com clareza a seus usuários, para que possam identificar problemas organizacionais facilmente.

Fontes: Dundas.com, "How Siemens Drastically Reduced Cost with Managed BI Applications," <http://www.dundas.com/resource/getcasestudy?caseStudyName=09-03-2016-Siemens%2FDundas-BI-Siemens-Case-Study.pdf> (acessado em julho de 2016); Wikipedia.org, "SIEMENS," <https://en.wikipedia.org/wiki/Siemens> (acessado em julho de 2016); Siemens.com, "About Siemens," <http://www.siemens.com/about/en/> (acessado em julho de 2016).

Análise de dados preditiva

A **análise de dados preditiva** visa determinar o que é mais provável de acontecer no futuro. Essa análise se baseia em técnicas estatísticas, bem como em outras técnicas desenvolvidas mais recentemente que recaem na categoria geral de **mineração de dados**. A meta dessas técnicas é conseguir prever se o cliente está propenso a migrar para um concorrente, o que o cliente tende a comprar a seguir e em qual quantidade, a quais promoções o cliente reagiria, qual o risco de crédito do cliente, e assim por diante. Inúmeras técnicas são usadas para desenvolver aplicativos analíticos preditivos, incluindo diversos algoritmos de classificação. Conforme descrito nos Capítulos 4 e 5, por exemplo, podemos usar técnicas de classificação como regressão logística, modelos de árvores de decisão e redes neurais para prever o nível de sucesso de bilheteria de um determinado filme a ser lançado. Também podemos usar algoritmos agregadores para segmentar clientes em diferentes grupos e preparar promoções específicas voltadas a cada um deles. Por fim, podemos usar técnicas de mineração de associação para estimar as relações entre diferentes comportamentos de compra. Ou seja, se um cliente compra um determinado produto, o que mais ele estará propenso a comprar? Tal análise pode ajudar um vendedor a recomendar ou promover produtos relacionados. Qualquer produto procurado na Amazon.com, por exemplo, leva o vendedor a sugerir outros produtos similares em que o cliente talvez esteja interessado. Estudaremos essas técnicas e suas aplicações nos Capítulos 3 a 6. O Caso Aplicado 1.4 ilustra uma dessas aplicações nos esportes.

Caso aplicado 1.4

Analisando lesões esportivas

Qualquer atividade esportiva é propensa a lesões. Se não houver um cuidado apropriado com as lesões, a equipe sairá prejudicada. O uso de análise de dados para entender as lesões pode ajudar a gerar vislumbres valiosos para que treinadores e médicos esportivos lidem com a composição da equipe, compreendam

os perfis dos atletas e, acima de tudo, tomem melhores decisões envolvendo a escalação ou não de jogadores em diferentes situações.

Em um estudo exploratório, a Oklahoma State University analisou lesões relacionadas a futebol americano usando relatórios e análise de dados preditiva. O projeto seguiu a

(Continua)

Caso aplicado 1.4 (Continuação)

metodologia CRISP-DM (a ser descrita no Capítulo 4) para identificar recomendações sobre gestão de lesões, entender os vários dados coletados a respeito, limpar os dados, desenvolver visualizações para fazer diversas inferências, construir MPs para analisar períodos de recuperação de lesões e obter regras sequenciais para prever relações entre as lesões e as várias partes do corpo afetadas.

O conjunto de dados envolveu mais de 560 registros de lesão no futebol americano, que foram divididos em categorias de lesão em si – parte do corpo/local/lateralidade, ação realizada, gravidade, tipo de lesão, data de início e de recuperação – e em categorias de atleta e de evento – ID do jogador, posição de atuação, atividade, início e local do jogo. O tempo de recuperação foi calculado para cada registro e classificado em diferentes categorias temporais: de 0 a 1 mês, de 1 a 2 meses, de 2 a 4 meses, de 4 a 6 meses e de 6 a 24 meses.

Diversas visualizações foram desenvolvidas para fazer inferências a partir de conjuntos de dados envolvendo o tempo de recuperação associado às posições dos jogadores; a gravidade das lesões e o tempo de recuperação; o tratamento oferecido e o tempo de recuperação associado; principais lesões em cada parte do corpo e assim por diante.

Modelos de redes neurais foram construídos para prever as categorias de recuperação usando o IBM SPSS Modeler. O sucesso da classificação de categorias de recuperação foi bastante bom: a precisão foi de 79,6%. Com

base nessa análise, muitas recomendações foram sugeridas, incluindo obter mais informação dos especialistas logo após a ocorrência da lesão, em vez de deixar a equipe de preparação física examinar jogadores lesionados, treinar jogadores de posições defensivas a evitar lesões e aplicar mecanismos rigorosos de prevenção de lesões em treinamento.

Questões para discussão

1. Quais tipos de análises estatísticas são aplicados no exame de lesões?
2. Como as visualizações ajudam a entender os dados e a filtrar recomendações a partir deles?
3. O que é um problema de classificação?
4. O que pode ser obtido a partir de análises sequenciais?

O que podemos aprender com este caso aplicado

Para qualquer projeto de análise de dados, é sempre importante entender o domínio a ser estudado e o estado atual de seus problemas por meio de uma análise extensiva do único recurso: dados históricos. Muitas vezes, visualizações representam uma ótima ferramenta para obter *insights* iniciais a partir dos dados, o que pode ser refinado ainda mais pelas opiniões de especialistas, identificando a importância relativa dos dados relacionados ao problema. As visualizações também ajudam a gerar ideias a respeito de problemas obscuros, o que pode ser obtido desenvolvendo-se MPs capazes de auxiliar no processo decisório de organizações.

Fonte: Sharda, R., Asamoah, D., & Ponna, N. (2013). "Research and Pedagogy in Business Analytics: Opportunities and Illustrative Examples." *Journal of Computing and Information Technology*, 21(3), 171–182.

Análise de dados prescritiva

Uma terceira categoria de análise de dados é denominada **análise de dados prescritiva**. A meta da análise de dados prescritiva é reconhecer o que está acontecendo, bem como

o que deve vir a acontecer, e tomar decisões para garantir o melhor desempenho possível. Historicamente, esse grupo de técnicas costumava ser estudado sob o guarda-chuva da pesquisa operacional ou das ciências administrativas, e seu objetivo geral é otimizar o desempenho de um sistema. A meta aqui é chegar a uma decisão ou a uma recomendação para uma ação específica. Tais recomendações podem assumir a forma de uma decisão pontual do tipo sim/não a um problema, uma quantidade específica (o preço de um determinado item, digamos) ou um conjunto completo de planos de produção. A decisão pode ser apresentada a um tomador de decisões ou pode ser usada diretamente em um sistema automatizado de regras decisórias (como em sistemas de precificação de companhias aéreas). Por isso, esses tipos de análise de dados também podem ser denominados **análise de dados normativa ou decisória**. O Caso Aplicado 1.5 oferece um exemplo de tais aplicações analíticas prescritivas. Aprenderemos outros aspectos da análise de dados prescritiva no Capítulo 6.

Caso aplicado 1.5

Uma empresa de barras de aço especiais utiliza análise de dados para determinar prazos de promessa de entrega

Este caso aplicado se baseia em um projeto que envolve um dos autores. Uma empresa que prefere não revelar seu nome (nem seu setor de atuação) estava tendo dificuldades ao se decidir sobre o estoque de matérias-primas a ser mantido para satisfazer seus clientes. Essa empresa fornece barras de aço customizadas para seus clientes. Essas barras podem ser cortadas em formatos e tamanhos específicos e podem obedecer a exigências singulares de material e acabamento. A empresa adquire matérias-primas por todo o mundo e as armazena em seu depósito. Quando um cliente em potencial entra em contato com a empresa para orçar barras especiais com exigências materiais específicas (composição, origem do metal, qualidade, formatos, tamanhos, etc.), o representante de vendas tem pouquíssimo tempo para produzir tal orçamento, incluindo a data em que o produto poderá ser entregue e, é claro, preços e outros detalhes. Ele precisa tomar decisões sobre disponibilidade de entrega, determinando em tempo real os prazos para quando pode prometer os produtos que os clientes consultaram durante a etapa de orçamento. Anteriormente, um representante de vendas tinha de tomar tais decisões analisando relatórios sobre o estoque disponível de matérias-primas. E parte da matéria-prima

disponível podia já ter sido comprometida para a encomenda de outro cliente. Assim, talvez nem todo estoque pudesse ser considerado livremente disponível. Por outro lado, talvez haja uma carga de matéria-prima a ser entregue no futuro próximo e que também poderia ser usada para satisfazer a encomenda desse cliente em potencial. Por fim, pode até haver a oportunidade de cobrar mais caro para dar prioridade a uma nova encomenda, reaproveitando estoque já comprometido e atrasando uma encomenda anterior. Obviamente, tais decisões devem se basear em análises de custo/benefício quanto a atrasar ou não uma encomenda anterior. O sistema deve, portanto, ser capaz de extrair em tempo real dados sobre estoque, encomendas comprometidas, entrada de matéria-prima, limitações de produção e assim por diante.

Para embasar tais decisões sobre prazos de entrega, um DSS em tempo real foi desenvolvido para encontrar um aproveitamento ideal do estoque disponível e nortear análises adicionais do tipo “e se”. O DSS utiliza um pacote de modelos de programação mesclada com números inteiros que são solucionados usando-se software comercial. A empresa incorporou o DSS ao seu sistema de planejamento de recursos empresariais para racionalizar seu uso de análise de negócios.

(Continua)

Caso aplicado 1.5 (Continuação)

Questões para discussão

1. Por que a realocação de estoque de um cliente para outro pode ser uma importante questão para discussão?
2. De que forma um DSS pode ajudar na tomada de tais decisões?

Fonte: Pajouh Foad, M., Xing, D., Hariharan, S., Zhou, Y., Balasundaram, B., Liu, T., & Sharda, R. (2013). "Available-to-Promise in Practice: An Application of Analytics in the Specialty Steel Bar Products Industry." *Interfaces*, 43(6), 503–517. <http://dx.doi.org/10.1287/inte.2013.0693> (acessado em julho de 2016).

Análise de dados aplicada em diferentes domínios

As aplicações da análise de dados em vários ramos de atuação deram origem a muitas áreas relacionadas ou pelo menos a novos vocabulários. Está na moda vincular a expressão análise de dados a qualquer ramo ou tipo de dados. Além da categoria geral de análise textual – voltada a agregar valor a partir de texto (a ser estudada no Capítulo 5) – ou da análise da Web – que analisa fluxos de dados na Internet (também no Capítulo 5) – muitos fluxos/profissões de análise de dados específicos de um ramo de atuação ou de um problema foram desenvolvidos. Dentre os novos terrenos de aplicação da análise de dados estão marketing, varejo, prevenção de fraudes, transportes, saúde, esportes, recrutamento de talentos, ciência comportamental, entre outros. A Seção 1.1 introduziu o exemplo da análise de dados esportivos. O Caso Aplicado 1.1 também poderia ser considerado um caso de estudo de análise de dados em companhias aéreas. A seção a seguir introduzirá em termos gerais sua aplicação na área de saúde e de mercado. Literalmente toda análise sistemática de dados em um setor específico está sendo rotulada como análise de dados de (“preencha a lacuna”). Embora isso possa levar longe demais o conceito de análise de dados, o benefício é que mais pessoas de ramos específicos estão cientes do poder e do potencial dessa ferramenta. Isso também proporciona um foco para profissionais que desenvolvem e aplicam conceitos da análise de dados em um setor vertical. Ainda que muitas das técnicas para desenvolver aplicações de análise de dados sejam comuns, há aspectos singulares dentro de cada segmento vertical que influenciam o modo como os dados podem ser coletados, processados, analisados e as aplicações implementadas. Por isso, a diferenciação da análise de dados com base em um foco vertical é vantajosa para o crescimento em geral da disciplina.

Análise de dados ou ciência de dados?

Embora o conceito de análise de dados esteja recebendo mais atenção por parte da indústria e dos círculos acadêmicos, outro termo já foi introduzido e está ganhando popularidade. O novo termo é *ciência de dados*. Assim, praticantes da ciência de dados são cientistas de dados. D. J. Patil, do LinkedIn, é às vezes creditado como o criador do termo *ciência de dados*. Já foram feitas algumas tentativas de descrever as diferenças entre analista de dados e cientistas de dados (consulte, por exemplo, emc.com/collateral/about/news/emc-data-science-study-wp.pdf). Uma opinião é que *analista de dados* é apenas mais um termo para designar os profissionais que estão praticando BI na forma de compilação e limpeza de dados, extração de relatórios e talvez alguma visualização.