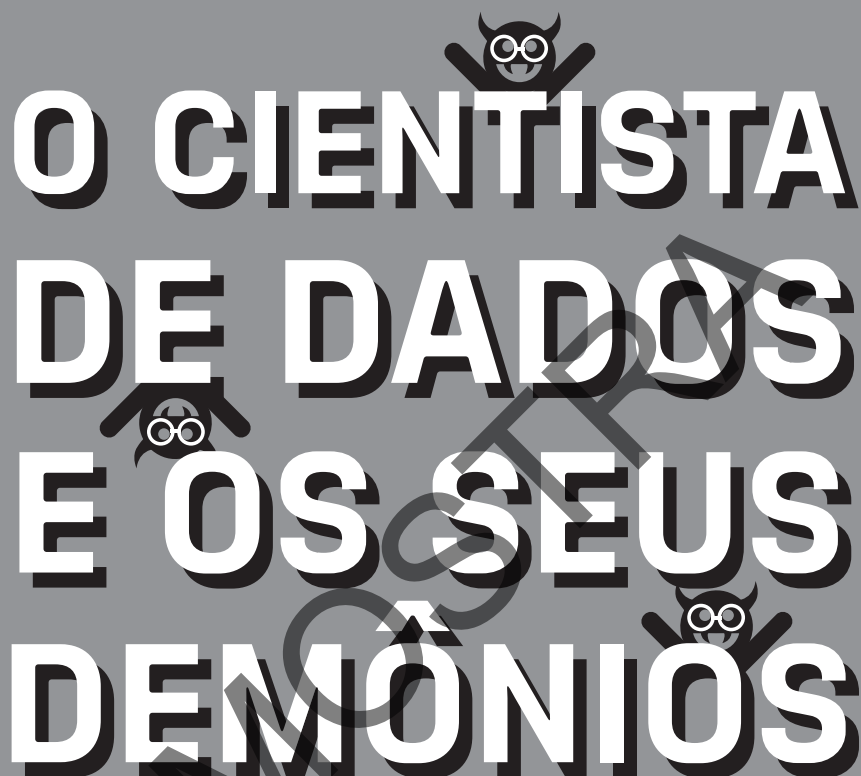


Heverton Anuniação

O CIENTISTA
DE DADOS
E OS SEUS
DEMÔNIOS

The title is written in large, bold, white, sans-serif capital letters. The words are stacked vertically. Small owl icons with large eyes are placed above the 'O' in 'CIENTISTA', above the 'E' in 'E OS SEUS', and above the 'O' in 'DEMÔNIOS'. A large, semi-transparent watermark 'AM' is overlaid diagonally across the title.

Conselhos de grandes
profissionais da área de dados
do Brasil e do exterior

Prefácio por:

Bill Inmon, o Pai do Data Warehouse
e Marilu Lopez, Presidente da DAMA International



ALTA BOOKS
GRUPO EDITORIAL
Rio de Janeiro, 2023

SUMÁRIO

SOBRE O CURADOR E AUTOR: HEVERTON ANUNCIÇÃO	V
AGRADECIMENTOS	VII
Profissionais convidados do Brasil e do exterior	VIII
INTRODUÇÃO POR HEVERTON ANUNCIÇÃO:	
Sabedoria e intuição digitais	1
PREFÁCIO POR BILL INMON — ESTADOS UNIDOS	
Matando o Data Warehouse	7
Comentários do autor Heverton Anunção	13
PREFÁCIO POR MARILU LOPEZ — MÉXICO	
Gestão de dados como base para a transformação digital	15
Comentários do autor Heverton Anunção	22
AVIV GRUBER — ISRAEL	
O que é preciso para se tornar um cientista de dados?	23
Comentários do autor Heverton Anunção	26
WALDINEI GUIMARÃES — BRASIL	
O cientista de dados e seus demônios	28
Comentários do autor Heverton Anunção	30
THIAGO MARQUES — BRASIL	
Como começar a aprender estatística e ciência de dados	32
Comentários do autor Heverton Anunção	41

ERIC SIEGEL — ESTADOS UNIDOS

Falácia da exatidão: A cobertura da IA pela mídia é falsa 42
Comentários do autor Heverton Anunciação 51

JIM STERNE — ESTADOS UNIDOS

Descobrimo a descoberta 52
Comentários do autor Heverton Anunciação 67

DR. KIRK BORNE — ESTADOS UNIDOS

Condução segura na empresa de autocondução 69
Comentários do autor Heverton Anunciação 76

JESSE ANDERSON — ESTADOS UNIDOS

Os dois tipos de engenharia de dados 77
Comentários do autor Heverton Anunciação 80

MARIE WALLACE — IRLANDA

Cientistas de dados e profissionais de análise:
Somos mais fortes juntos 81
Comentários do autor Heverton Anunciação 90

KRISTEN KEHRER — ESTADOS UNIDOS

Ingredientes-chave para ser orientado por dados 92
Comentários do autor Heverton Anunciação 103

AMIT AGARWAL — ÍNDIA

Estrutura analítica de autoatendimento 104
Comentários do autor Heverton Anunciação 116

TIMOTHY KOOI — SINGAPURA

A ciência de dados requer um conjunto diferente
de mentalidades empresariais 117
Comentários do autor Heverton Anunciação 122

CHARLES GIVRE — ESTADOS UNIDOS

Iniciando uma carreira na Ciência de dados	123
Comentários do autor Heverton Anuniação	128

BRAM NAUTS — PAÍSES BAIXOS

A habilidade do futuro — Ideias-chave para melhorar a sua	130
Comentários do autor Heverton Anuniação	142

DR. RAJKUMAR BONDUGULA — ESTADOS UNIDOS

Sua equipe de ciência de dados está preparada para o sucesso?	143
Comentários do autor Heverton Anuniação	156

ROBERTO OLIVEIRA — BRASIL

Considerações para um projeto de Business Intelligence	157
Comentários do autor Heverton Anuniação	168

ANGÉLICA NARDO CASERI, PHD — BRASIL

Aprendendo ciência de dados, na prática.	169
Comentários do autor Heverton Anuniação	174

RONALD WASSERSTEIN — ESTADOS UNIDOS

Um ótimo momento para ser um estatístico.	176
Comentários do autor Heverton Anuniação	181

FERNANDA MACIEL, PHD — ESTADOS UNIDOS

Preciso mesmo fazer uma estatística descritiva?	183
Comentários do autor Heverton Anuniação	192

CLAUDIA MONTENEGRO — BRASIL

Os dados obtidos mediante inferências na área da saúde e os impactos na LGPD	193
Comentários do autor Heverton Anuniação	198

CONCLUSÃO

e a evolução continua... 200

FERRAMENTAS PARA O CIENTISTA DE DADOS 202

Lei Geral de Proteção de Dados Pessoais (LGPD) 203

Autoridade Nacional de Proteção de Dados 213

GUIA PARA O CORPO DE GERENCIAMENTO DE DADOS DO CONHECIMENTO (DAMA-DMBOK)

Estrutura funcional. 218

ASSOCIAÇÕES PROFISSIONAIS OU ACADÊMICAS

PARA O PROFISSIONAL DE CIÊNCIA DE DADOS 221

Ferramentas e Recursos para o Cientista de Dados 222

QUADRO DE ALGUNS SOFTWARES DE ANÁLISE DE

DADOS DE CÓDIGO ABERTO EM 2021 233

Glossário de Data Warehouse & Business Intelligence 234

REFERÊNCIA BIBLIOGRÁFICA 250

Índice. 251

AMOSTRA

*Trate seu próximo como você gostaria de ser tratado
— mesmo se for seu cliente!*

Jesus Cristo

AMOSTRA



INTRODUÇÃO POR HEVERTON ANUNCIÇÃO:

SABEDORIA E INTUIÇÃO DIGITAIS

VOCÊ É O QUE COMPARTILHA, creio eu. Por essa razão, decidi me tornar o curador e criei este livro.

Eu acabo de me deparar com uma matéria da revista *Superinteressante*, da Editora Abril, com o seguinte título: “**Nos EUA, algoritmo promete prever crimes uma semana antes que eles ocorram.**” (*Superinteressante*, 2022).¹

O que me faz lembrar de um outro momento. Você se lembra daquele filme chamado *Minority Report*? No qual o ator **Tom Cruise** consegue também atuar na prevenção de crimes antes deles acontecerem?

Enquanto isso, também nos Estados Unidos, o **Google** acaba de demitir um engenheiro que abriu ao público seu projeto de Inteligência Artificial (IA), conforme matéria no *Yahoo Finanças*: “Engenheiro do Google é afastado por afirmar que robô da empresa ‘tem vida.’” (*Yahoo Finanças*, 2022).²

De quais algoritmos estamos falando aqui?

-
1. Disponível em: <<https://super.abril.com.br/sociedade/nos-eua-algoritmo-promete-prever-crimes-uma-semana-antes-que-eles-ocorram/>>. Acesso em: jul. 2022.
 2. Disponível em: <<https://br.financas.yahoo.com/news/engenheiro-do-google-e-afastado-por-afirmar-que- robo-da-empresa-tem-vida-141414872.html> — julho 2022>. Acesso em: jul. 2022.

Quais engenheiros de softwares são esses?

De onde vêm esses cientistas de dados?

Como eles vivem?

O que eles comem?

Como eles se reproduzem?

Um profissional, não importa a que área ele pertença, acredito, nunca deve pensar que sua verdade é definitiva ou que sua maneira de fazer ou resolver algo é a melhor.

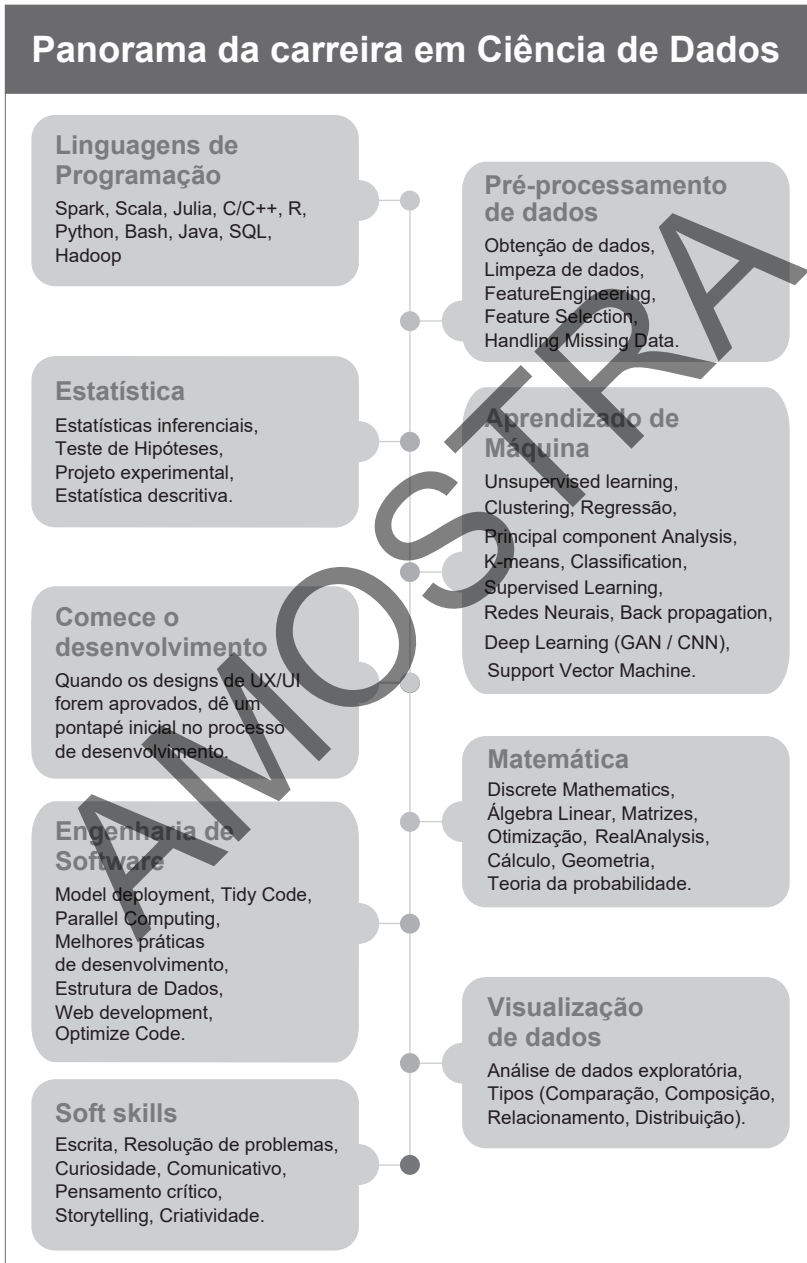
E, logicamente, tive que fazer certo e errado para chegar a essa simples conclusão.

Agora, o que isso tem a ver com o propósito deste livro? No qual reuni dicas e conselhos importantes de uma elite de profissionais da ciência de dados de vários setores e experiência respeitável.

E, neste mundo de algoritmos, é obrigação de qualquer profissional que precisa de dados estar conectado às melhores práticas, ou, quem sabe, até inventar suas próprias.

Eu resolvi fazer um infográfico com um resumo das principais características comportamentais ou técnicas que esse profissional da área de dados deve enriquecer-se diariamente, bem como sua equipe:

Figura 1 – Principais características comportamentais



Fonte: Conteúdo autoral.

Agora, isso me faz lembrar outro filme chamado *Jerry Maguire*, que teve também como o ator o mesmo **Tom Cruise**. Há uma parte do filme no qual os atores principais questionam: “*show me the money*” (mostre-me a grana), afinal, de que vale tanto esforço se não há a recompensa?

Adaptando para o nosso livro, de que vale o profissional de dados ter bilhões em registros se no final não obtiver a sabedoria digital ou “riqueza” para melhores decisões?

Em um de meus livros, *A Matemática das Emoções*, eu crio a **Fórmula da Expectativa**, em que $S = R - E$ (Satisfação é igual ao Resultado menos Expectativa). Qual a relação dessa com a carreira de um profissional de dados? É que nenhuma empresa irá recompensar justamente esse profissional se ele não entregar o resultado acima da média.

Figura 2 - Precisamos de dados



Fonte: Conteúdo autoral.³

Depois de ter trabalhado em centenas de projetos de consultoria e implementação de melhores práticas em Marketing de Relacionamento (CRM), Business Intelligence (BI) e Experiência do Cliente (CX), bem como em inúmeros projetos de Tecnologia da Informação, uma verdade é absoluta: precisamos de dados!

3. Extraído do livro do próprio autor, Heverton Anunciação, livro publicado na Amazon KDP Kindle, e publicação própria.

Encontrei empresas que queriam proporcionar uma excelente experiência ao cliente, mas não tinham os dados certos!

Encontrei empresas que queriam calcular quais canais eram os mais rentáveis, mas não tinham os dados corretos!

Encontrei empresas que queriam avaliar quais são os melhores produtos e clientes que dão os melhores retornos, mas que só tinham dados fora de sincronia!

Estes são apenas alguns exemplos de problemas que encontramos hoje, nas empresas. A maioria das empresas diz que faz tudo perfeitamente, mas não é mostrado na mídia ou na imprensa a dor de cabeça que as áreas de Tecnologia da Informação sofrem para unir os dados corretos. E quando elas conseguem se unir e torná-los disponíveis, o tempo para comercializar e as possíveis oportunidades já foram perdidos.

Portanto, se uma empresa quer ser considerada excelência em governança corporativa e satisfazer as áreas jurídica, marketing, vendas, atendimento ao cliente, tecnologia, logística, produtos, entre outras, deve começar o mais rápido possível para se tornar uma empresa voltada para os dados e em tempo real.

Dizem que os dados e as informações são o novo petróleo. Então, quem tem as informações corretas e sabe como utilizá-las em sua empresa?

**70% do tempo do cientista de dados é
limpeza e preparação de dados**

**90% do tempo do minerador é gasto na
perfuração de ouro**

Muitas empresas continuam a praticar seu planejamento estratégico anual. Encontramos algumas que passaram do tradicional modelo de projeto “cachoeira” para o modelo “ágil” para acompanhar a velocidade do mercado. Com isso, aplicando o conceito: cometer um erro rápido, aprender rápido e corrigir rapidamente.

Errar é importante, mas errar com o menor custo possível. Para isso, recomendo às empresas que procurem suas intuições digitais, inspirações digitais.

Explico melhor: a intuição é a capacidade de entender as coisas no momento, sem a necessidade de realizar raciocínios complexos. É também a capacidade ou o ato de perceber, discernir ou sentir as coisas, independentemente do raciocínio ou da análise. A intuição é uma percepção criada dentro do indivíduo ou da empresa e, com ela, decisões ou ações são tomadas.

Agora, imagine combinar intuição com inspiração? A inspiração é, em um sentido figurativo, o ato de inspirar uma boa ou má ação. Isto é, algo ou insight que vem de “fora” do indivíduo ou “fora” da empresa, que causa decisões ou ações a serem tomadas.

Que pergunta eu gostaria de fazer? Pergunte-se: por que as mulheres têm um sexto sentido muito mais aguçado do que os homens? Isso é intuição feminina ou o sexto sentido. Sentimento, sensação, sexto sentido, o nome não importa, porque todos os sinônimos se referem à mesma coisa: intuição, ou seja, a percepção que toda mulher tem e ninguém sabe de onde vem. É um som, uma voz, um eco interior. E isso não tem absolutamente nada de sobrenatural!

Assim, com este livro, proponho a todos os funcionários e empresas que chegará um dia em que todos saberão utilizar, a partir de seus dados, seu sexto sentido. O sexto sentido é uma percepção extrassensorial, que vai além de nossos cinco sentidos básicos: visão, audição, paladar, olfato e tato. É uma sensação de intuição, que de certa forma nos permite ter sensações de “clarividência” e até mesmo visões de eventos futuros.

Uma empresa só alcançará essa capacidade se começar imediatamente a aplicar a verdadeira governança de dados. E os ilustres cientistas de dados, que fazem parte deste livro, lhe mostrarão o caminho para dar o primeiro passo.

Este livro faz parte da Trilogia CRM e Customer Experience chamada CX Trilogy cujo objetivo é unir a comunidade mundial de profissionais da CX, serviço ao cliente, ciência de dados e CRM. Acredito que esta união facilitaria a contratação de nosso setor e profissão, assim como a identificação dos melhores profissionais do mercado.

Aproveite a viagem!

HEVERTON ANUNCIÇÃO

www.heverton.com.br



PREFÁCIO POR BILL INMON — ESTADOS UNIDOS

MATANDO O DATA WAREHOUSE

NÃO HÁ dúvidas sobre isso. É um mundo confuso lá fora. Além disso, o mundo não está parado. O que você aprendeu ontem funciona hoje sob um novo conjunto de regras. O mundo muda a cada dia. A cada hora. A cada minuto.

Você faz escolhas o tempo todo que o afetam e aqueles ao seu redor, continuamente...

Então, com toda a complexidade, toda a confusão, toda a mudança constante que é simplesmente um fato da vida no mundo, como você mantém sua sanidade? Como você mantém sua bússola moral? Como você se mantém centrado quando tudo o mais ao seu redor está enlouquecendo? Como você encontra a paz em um mundo de caos?

Aqui estão alguns conselhos que me serviram bem.

1. Confie em sua intuição. Isso vale para o intelecto. Emoção. E escolher e manter amigos. Parece que todos nós temos um sexto sentido sobre em quem devemos confiar e o que devemos aprender e fazer. Ouça esse sexto sentido interior.
2. Tratar cada dia como um presente, porque é exatamente isso que é. Cada dia é precioso. Tire o máximo proveito de cada dia.

3. Se você não consegue explicar algo a uma criança de 6 anos, você mesmo não entende. Esta citação foi de Albert Einstein. Mas é tão verdadeira e tão relevante quanto no dia em que disse. No final do dia, ele era um sujeito bastante brilhante.

4. Não importa o que você esteja fazendo, saiba em cada momento de sua vida profissional como o que você está fazendo se relaciona com o valor comercial. Se você não pode relacionar o que está fazendo em sua vida profissional com o valor comercial, então você não deveria estar fazendo isso em primeiro lugar.

5. Quando você trabalha para fazer três coisas:

- ofereça valor a seu cliente;
- proporcione valor a seu acionista; e
- proporcione dor à sua concorrência.

6. Reconheça que você aprenderá por toda sua vida. Sua educação formal termina um dia, mas sua experiência de aprendizagem continua diariamente, pelo resto de sua vida. Abraça a aprendizagem e as mudanças que ocorrerão todos os dias em sua vida.

Se seguir esses conselhos, você e sua família nunca passarão fome. E acordarão todas as manhãs com um senso de propósito e vitalidade.

Matando o Data Warehouse

É um esporte industrial. No início, os teóricos do banco de dados queriam matar o armazém de dados, porque ele não se enquadrava em sua teoria do que deveria ser um banco de dados. Então, os fornecedores que suportavam esses teóricos do banco de dados queriam matar o Data Warehouse (DW). O armazém de dados “moveu seu queijo”. E os fornecedores não gostavam de ninguém que movesse esse queijo. Outros vieram e tentaram chamar um Data Mart de armazém de dados. Mais confusão. Então apareceram os vendedores de Big Data e anunciaram que poderiam substituir um armazém de dados por Big Data.

O Data Warehouse parecia um mercado maduro e tentador. Então, os vendedores de Big Data tentaram um pouco de marketing para convencer as pessoas de que o Big Data era o substituto para um armazém de dados.

Todos queriam matar o armazém de dados.

O armazém de dados era como o saco de pancada no ginásio. Todo pugilista ou aspirante a pugilista aparece e bate no saco de boxe (armazém de dados). Como Rodney Daingerfield, “o armazém de dados não recebe nenhum respeito”.

Mas, por meio dele, todo o armazém de dados ficou maior e mais forte com todos esses detratores e inimigos querendo substituí-lo/matá-lo/demovê-lo.

O que é o armazém de dados que gerou este tremendo poder de permanência? A característica essencial, número um de um Data Warehouse, é que ele é uma arquitetura, não uma tecnologia. O Data Warehouse resolve um problema muito essencial para o mundo — permite que os dados baseados em aplicações coexistam harmoniosamente com os dados corporativos. Enquanto houver aplicações e necessidade de dados corporativos em toda a empresa, haverá um armazém de dados.

Mas isso era o passado. Um a um, as pessoas que tentaram matar o armazém de dados caíram (ou estão caindo).

Quando você olha para o futuro, há algumas luzes muito brilhantes no futuro do Data Warehouse. Uma extensão natural do armazenamento de dados é o cofre de dados. O cofre de dados tem muitos benefícios. Quando há necessidade de um alto grau de confiança e confiabilidade dos dados, o cofre de dados é a resposta. Dan Linstedt fez o trabalho pioneiro em um cofre de dados que lidera o caminho. De maneira diferente, quando a confiabilidade e credibilidade dos dados é um problema, o cofre de dados é a resposta.

Mas há outra luz muito brilhante para o armazém de dados, além do cofre de dados. Essa luz brilhante é a adição de dados baseados em texto ao armazém de dados. Tradicionalmente, o armazenamento de dados era baseado em dados estruturados, baseados em transações. Aplicações processavam transações e produziam bancos de dados que precisavam ser integrados a um armazém de dados. Mas agora — com o ETL (Extract, Transform, Load) textual — é possível começar a incluir dados baseados em texto no armazém de dados. Duas coisas acontecem simultaneamente