

# A ciência de dados e a cientista de dados

## Data science and data scientist

Renata Cristina Gutierrez Castanha<sup>1</sup>

<sup>1</sup> Anheuser Busch Inc. ORCID: <https://orcid.org/0000-0002-0167-4213>

Autor para correspondência/Mail to: Renata Cristina Gutierrez Castanha, regutierrez@gmail.com



**Renata Cristina Gutierrez Castanha** é cientista de dados na empresa AB-InBev (Brasil). Licenciada em Matemática pela UNESP - Univ Estadual Paulista, mestra e doutora em Ciência da Informação pela UNESP - Univ Estadual Paulista, com período sanduíche na UWM - University of Wisconsin-Milwaukee sob orientação do Dr. Dietmar Wolfram e financiamento CAPES. Membro da Associação Nacional de Pesquisa e Pós-Graduação em Ciência da Informação (ANCIB), International Society for Knowledge Organization (ISKO) e do Grupo de Pesquisa em Estudos Métricos da Informação (EMI), atuando nas temáticas relacionadas à organização do conhecimento, estudos métricos da informação, análise de domínio, análise de redes sociais, avaliação da pós-graduação, genealogia acadêmica, indicadores topológicos, estatística multivariada, data science fundamentais e machine learning.



Copyright © 2021 Castanha. Todo o conteúdo da Revista (incluindo-se instruções, política editorial e modelos) está sob uma licença Creative Commons Atribuição 4.0 Internacional. Ao serem publicados por esta Revista, os artigos são de livre uso em ambientes educacionais, de pesquisa e não comerciais, com atribuição de autoria obrigatória. Mais informações em <http://revistas.ufpr.br/atoz/about/submissions#copyrightNotice>.

## Resumo

A Dr.<sup>a</sup> Renata Cristina Gutierrez Castanha responde questões sobre o conceito, aplicação e projetos de ciência de dados, bem como sobre a atuação do e da cientista de dados.

**Palavras-chave:** Dados; Ciência de Dados; Cientista de Dados.

## Abstract

Dr. Renata Cristina Gutierrez Castanha answers questions about the concept, application and data science projects, as well as about the performance of the data scientist.

**Keywords:** Data; Data science; Data scientist.

## 1. Profissionalmente, quem é a Renata?

Sou uma pessoa que procura sempre estar aprendendo alguma coisa e a *keyword* da moda que designa esse comportamento é *lifelong learner*. Nesse sentido, eu passei a minha vida inteira procurando e encontrando um conhecimento que me era atrativo e me especializando em tópicos que eu particularmente gostava muito, que utilizaria no futuro, mas não sabia exatamente onde aplicar. Eu sabia que queria aprender sempre mais. E assim me tornei uma profissional híbrida, em vários sentidos. Comecei minha carreira com a licenciatura e acabei em um ambiente corporativo, dois domínios completamente opostos, mas que, de alguma maneira, se encaixaram perfeitamente no que eu queria: aplicar todo conhecimento teórico que eu obtive ao longo da minha vida a problemas reais, que trouxessem valores para o negócio da empresa em que eu trabalho, e a evangelização e educação sobre questões e metodologias quantitativas e a cultura *data-driven*.

Hoje, me sinto uma profissional completa, pois continuo na parte educacional, lecionando disciplinas relativas a dados e metodologias quantitativas na UFSCar, e ainda pesquisando e aplicando todo meu conhecimento na parte corporativa, atuando como cientista de dados e, mais recentemente, como gerente de produtos de dados, na Ambev, mais especificamente produtos que envolvam soluções com inteligência artificial.

## 2. O que é a ciência de dados?

Para falar de ciência de dados, precisamos dar um passo atrás e falar de como dados são importantes. Não é novidade que estamos vivendo em um mundo que está se afogando em dados – tantos dados, que daí vem o termo big data. Desde os websites que estão rastreando cada clique nosso, até o celular que registra nossa localização, smartwatches que utilizam pedômetros e estão sempre registrando nossos batimentos cardíacos, hábitos de movimento e padrões de sono. Basicamente, tudo que fazemos no nosso dia a dia gera dados de alguma maneira – e muitos deles, atingindo 59 zettabytes em 2020.

Profissionais da [International Data Corporation \(2021\)](#) apontam que a produção de dados dobra a cada dois anos. A quantidade de dados criados nos próximos três anos será maior do que os dados criados nos últimos 30 anos. Isto posto, precisamos ter em mente que os dados, sozinhos, não nos dizem nada. É aí que entra a ciência de dados. Em linhas gerais, a ciência de dados é a arte de obter conhecimento utilizando dados brutos. Não existe uma definição formal em consenso amplamente utilizada no campo, mas, mais especificamente, entende-se a ciência de dados como um domínio de conhecimento multidisciplinar, que utiliza o método científico e técnicas de análise de dados, machine learning e inteligência artificial para resolver problemas reais, trazendo à luz conhecimento estratégico e insights acionáveis para o negócio. Esses insights são utilizados para que as empresas aprimorem seu processo de tomada de decisão, otimizando as estratégias de negócios. A Netflix, por exemplo, utiliza dados do usuário para entender seus padrões de visualização, o que aumenta seu interesse por um gênero específico e os ajuda a entender quais filmes produzir e também o que recomendar. A ciência de dados também

envolve o desenvolvimento de um produto de dados, que é um ativo técnico que usa dados para gerar resultados baseados em algoritmos, como um software, um app, por exemplo – e é essa área que eu estou atuando agora.

### **3. Em qual contexto os dados ganharam visibilidade na economia mundial?**

Com o período pós-guerra, a chegada da internet e o desenvolvimento tecnológico, fomos observando o aumento do volume de dados, e com isso enormes possibilidades de armazenamento e processamento. Essa mudança trouxe uma oportunidade de transformar a economia tradicional. Observamos corporações tomarem como ativos propriedades, fábricas, equipamentos, estoque, dinheiro e propriedade intelectual. De uns anos para cá, um novo tipo de ativo surge, os dados, e com isso, empresas estão coletando e analisando grandes volumes de dados, gerando uma nova “economia de dados” moldada para gerar valor por meios internos e externos.

Muitos já ouviram a máxima que “dados são o novo petróleo”, cunhada pelo matemático Clive Humby. E realmente, estamos vivendo em uma época em que os dados são mais valiosos do que nunca: adotam um papel crítico com o crescimento da economia digital. Nos últimos anos, empresas com dados no centro dos seus modelos de negócios passaram a dominar os rankings das mais valiosas corporações do mundo, vide Amazon, Facebook, Apple, Microsoft, entre tantas. Porém, contrariando a máxima do petróleo, o valor dos dados está nos insights gerados por meio de análises e combinações de diferentes conjuntos de dados que fornecem o valor real – o resultado real da aplicação da ciência de dados.

### **4. Como a ciência de dados pode auxiliar na tomada de decisão nas organizações?**

Como mencionado anteriormente, estamos vivendo em um mundo inundado de dados. Nesse sentido, a ciência de dados desempenha papel basilar em quase todos os aspectos das operações e estratégias de negócios das corporações. Tomando um ponto de vista mais operacional, algumas iniciativas que compreendem o uso de ciência de dados para tomada de decisões poderia ser, por exemplo, a otimização do gerenciamento estoques de produtos, redes de distribuição, e até atendimento customizado ao cliente. Em um nível mais fundamental, os dados apontam o caminho para aumentar a eficiência e reduzir custos, permitindo que as empresas criem planos e estratégias de negócios baseados em dados, como análises derivadas do comportamento do cliente, tendências de mercado e concorrência, por exemplo. Sem os dados, podemos perder oportunidades e tomar decisões equivocadas.

Um exemplo muito claro de decisões *data-driven* ocorreu em 2012, em que a campanha do então presidente americano Barack Obama empregou dezenas de cientistas de dados que trabalharam com *data mining* e identificaram eleitores que precisavam de atenção extra, escolhendo programas e recursos de arrecadação de fundos específicos de doadores e concentrando esforços para conseguir votar onde eles provavelmente seriam úteis. Esses esforços desempenharam um papel importante na reeleição do presidente, o que significa que é uma aposta segura que as campanhas políticas do futuro se tornarão cada vez mais baseadas em dados.

Falando mais especificamente de aplicações na área de marketing, a utilização de dados para tomada de decisão poderia criar campanhas de marketing mais fortes e publicidade direcionada a nichos específicos de clientes para aumentar as vendas de produtos, por exemplo. Dentro da área financeira, outro exemplo seria o gerenciamento de riscos, detecção de transações fraudulentas e controle de inadimplência. Em ambientes industriais, podemos trabalhar com manutenção preditiva, evitando quebras de equipamentos, por exemplo. Na área da saúde, por exemplo, os dados podem ser utilizados para diagnóstico de condições médicas, análise de imagem, planejamento de tratamento e pesquisa médica. Outra área interessante é a desportiva, em que as equipes analisam o desempenho do jogador e planejam estratégias de jogo por meio da ciência de dados. Ou como você acha que a Alemanha goleou o Brasil por 7x1 na Copa do Mundo de 2014?

### **5. Quem é o cientista de dados? Quais competências são fundamentais para um cientista de dados?**

Não é toa chamado de o “trabalho mais sexy do século 21” por Davenport e Patil (2012), o cientista de dados é o profissional responsável por extrair insights de dados brutos por meio de técnicas analíticas avançadas, ao passo em que atende às necessidades e aos objetivos de negócios específicos de uma empresa. A função do cientista de dados está se tornando cada vez mais importante à medida que as empresas dependem cada vez mais dos dados para conduzir sua tomada de decisão.

Pessoalmente, eu acredito que, muito além de ferramentas específicas e conhecimento acadêmico, os cientistas de dados são principalmente pessoas solucionadoras de problemas, uma vez que trabalhar com dados também significa entender o objetivo, onde podemos chegar, que tipo de informação podemos extrair e como isso é importante para a empresa. Os cientistas de dados determinam as perguntas que precisam de respostas e apresentam diferentes abordagens até resolver o problema de negócio proposto.

As funções do cientista de dados geralmente incluem: análises estatísticas, identificação de padrões ou trends, criação de ferramentas, processos ou produtos baseados em inteligência artificial, desenvolvimento e teste de novos algoritmos, desenvolvimento de modelos preditivos com machine learning, construção de painéis para visualização de dados, e algumas competências soft, como apresentação de resultados e reunião com as partes interessadas (stakeholders) do seu projeto ou produto.

O conhecimento core exigido, em sua maioria, do profissional de dados é composto por disciplinas duras como

a Matemática e a Estatística. Para ser capaz de extrair insight dos dados, é imprescindível tomá-los de uma perspectiva quantitativa. Nesse sentido, a experiência em conceitos matemáticos como álgebra linear, estatística clássica e bayesiana é necessária para desempenhar tarefas como correlações de dados e descoberta de padrões. Ainda, esse profissional deve ser bem versado em linguagens de programação como SQL e Python ou R. Um cientista de dados precisa ser capaz de entender códigos complexos e traduzi-los em formatos mais legíveis e comprehensíveis, bem como desenvolver novos algoritmos conforme o solicitado. Por fim, um cientista de dados deve ter um bom conhecimento sobre as decisões táticas a serem tomadas no negócio e como o mesmo funciona. O cientista de dados é capaz de aconselhar a equipe de negócios depois de mergulhar fundo nos dados e extrair informações açãoáveis dos mesmos.

## **6. Conte-nos sobre a sua trajetória profissional até chegar a atuar como cientista de dados?**

Como eu mencionei antes, eu sempre gostei de estudar e estava atenta a tópicos quantitativos que me interessavam. Eu comecei lecionando matemática, ainda que voluntariamente, quando eu estava na graduação, e fui evoluindo essa vontade de ser professora quando entrei no mestrado e posteriormente, no doutorado. Em ambos, eu lecionei como professora bolsista e substituta em disciplinas de bibliometria/cientometria e estatística aplicada a diversos domínios, como pedagogia, fisioterapia, filosofia, biblioteconomia e arquivologia. Além disso, eu comecei a ter mais interesse por metodologia de pesquisa quantitativa e comecei a ver muito essa “nova” profissão de cientista de dados estava se estabelecendo no país. Foi aí que comecei a ler sobre essa carreira emergente, e estudar cada vez mais sobre o tema.

Algum tempo depois, eu comecei a procurar emprego ativamente por conta da situação em que a educação brasileira se encontrava – e ainda se encontra: falta de incentivo, concursos, bolsas. Eu fui bolsista por quase toda a minha vida “profissional” – aqui entenda-se minha trajetória acadêmica na pós-graduação, e eu precisava de experiência profissional e estabilidade financeira. Foi aí que coloquei meu currículo à disposição no LinkedIn e um recrutador se interessou pelo meu currículo e assim eu fui convidada para uma entrevista com a área de *Advanced Analytics* da Ambev, que estava se estruturando. Depois de aprovada no processo seletivo, eu comecei sendo a cientista de dados responsável pelo departamento de *Supply* da cervejaria, depois fui para a área de *Sales Finance* e estou há três anos atuando na área.

## **7. Quais recomendações você daria para quem está iniciando na ciência de dados?**

Seja curioso. Não existe caminho fácil. Estude bastante, mas não foque na teoria: faça cursos e/ou veja vídeos com uma abordagem mais “mão na massa”, e que utilizem dados mais próximos da realidade. O arcabouço teórico quantitativo é super importante, mas o que faz diferença no dia a dia é como você tem habilidades de resolver os problemas que são propostos.

## **8. O que mais te anima (ou te motiva) em relação ao momento atual da ciência de dados?**

Pra mim, o mais animador é que estamos apenas no começo do desenvolvimento desse universo de ciência de dados, e temos muito pela frente! Uma das partes mais empolgantes é o processo de aprendizado pelo qual passamos e o crescimento que eu venho ganhando como indivíduo. Cientistas de dados estão constantemente aprendendo e trazendo soluções inovadoras para o mundo, em diversas áreas.

## **9. Em relação à ciência de dados, você tem algum mentor ou pessoas que te influenciam ou inspiram?**

Sem dúvida, minhas mentoras são as pessoas que me direcionaram para esse caminho e sempre me apoiaram incondicionalmente: minha orientadora de mestrado e doutorado, a Profa. Dra. Maria Cláudia, e também a Profa. Dra. Ely Tannuri. Além disso, eu trabalho com pessoas incríveis que me inspiram todo dia e temos uma troca de conhecimento inestimável.

## **10. Você tem um projeto favorito em ciência de dados? O que este projeto tem de diferente?**

Tenho! Esse projeto, chamado AB Credit, que começou em meados de 2018, acabou virando um produto de inteligência artificial que utilizamos diariamente na Ambev. Como cientista de dados responsável pelo departamento de *Sales Finance*, eu fui a responsável pelo conceito, planejamento, design do produto, parte do desenvolvimento, validação, lançamento e suporte total até hoje.

Esse produto foi desenvolvido em parceria com a incubadora *Beer Garage*, da própria Anheuser-Busch InBev (a Ambev “americana”), localizada no Vale do Silício, e é composto por uma solução de inteligência artificial compreendendo uma cadeia de algoritmos de *machine learning* que analisa os dados históricos dos nossos clientes e faz previsão do risco de crédito, garantindo uma otimização do nosso capital empregado e controle de inadimplência. Foi uma experiência intensa, e o resultado foi muito melhor que o esperado, inclusive citado em matérias nos cadernos de inteligência artificial do *Wall Street Journal*, por [Sreeharsha \(2019\)](#) e *Forbes*, por [Marr \(2019\)](#).

## REFERÊNCIAS

- Davenport, T., & Patil, D. (2012, 1 de Outubro). *Data Scientist: The Sexiest Job of the 21st Century*. Harvard Business Review. Recuperado de <https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century>
- International Data Corporation. (2021). *The premier global market intelligence firm*. Recuperado de <https://www.idc.com/>
- Marr, B. (2019, 9 de Setembro). *The Amazing Ways The Brewers of Budweiser Are Using Artificial Intelligence To Transform The Beer Industry*. Forbes. Recuperado de <https://www.forbes.com/sites/bernardmarr/2019/09/09/the-amazing-ways-the-brewers-of-budweiser-are-using-artificial-intelligence-to-transform-the-beer-industry/?sh=5811b4b7422c>
- Sreeharsha, V. (2019, 25 de Novembro). *AB InBev Uses AI to Assess Beer Quality, Creditworthiness of Distributors*. Wall Street Journal. Recuperado de <https://www.wsj.com/articles/ab-inbev-uses-ai-to-assess-beer-quality-creditworthiness-of-distributors-11574677800?mod=searchresults&page=1&pos=3>

---

Como citar esta entrevista (APA):

Castanha, R. C. G. (2021). A ciência de dados e a cientista de dados. *AtoZ: novas práticas em informação e conhecimento*, 10(2), 1 – 4. Recuperado de: <http://dx.doi.org/10.5380/atoz.v10i2.79822>