

CIENTISTA DE DADOS E PESQUISA: UMA ANÁLISE EXPLORATÓRIA DO PERFIL PROFISSIONAL E DAS TENDÊNCIAS DO MERCADO GLOBAL

Leticia S. Almeida¹, Mohammad S. S. Silva², Pedro V. R. Almeida³, Rodrigo F. Pereira⁴, Bethina R. Camargo⁵, Felipe A. Camargo⁶

^{1,2,3,4,5,6} Centro Universitário SENAI São Paulo (UniSENAI) - Campus Santo Amaro, São Paulo, Brasil (leticia.s.almeida19@aluno.senai.br)

Resumo: Este estudo analisa a profissão de Cientista de Dados com base em cerca de 1.655 registros entre 2020 e 2023. O objetivo é compreender o perfil do cargo no mercado de tecnologia e identificar fatores que impulsionam sua valorização. Os resultados indicam que a área se destaca pela integração entre habilidades analíticas, técnicas e visão de negócios, sendo essencial em um cenário global orientado por dados.

Palavras-chave: Mercado de Trabalho; Tendências Salariais; Carreira em Dados; Remuneração; Ciência de Dados.

INTRODUÇÃO

Na atual era da tecnologia e da globalização, a produção de dados atingiu escala sem precedentes. Estima-se que, em apenas um minuto, a internet receba milhões de novos registros digitais, desde o compartilhamento de conteúdo em redes sociais até interações complexas de mercado (Insper, 2025). Diante desse volume crescente de informações, a capacidade de coletar, processar e interpretar esses dados tornou-se um diferencial competitivo para organizações de diferentes setores. Nesse contexto, a ciência de dados configura-se como um campo interdisciplinar voltado à extração de padrões, geração de conhecimento e apoio à tomada de decisão.

A área de Ciência de Dados tem apresentado forte expansão nos últimos anos, impulsionada pela crescente necessidade das organizações em realizar análises preditivas e orientadas a dados, capazes de apoiar decisões estratégicas e antecipar tendências de mercado indo além das análises descritivas tradicionalmente associadas às ferramentas de *Business Intelligence* (BI). O impacto desses profissionais nas empresas é expressivo, permitindo identificar padrões relevantes, o desenvolvimento de soluções baseadas em Inteligência Artificial e a geração de valor a partir dos dados (Voce, 2025). Esse cenário é corroborado por diferentes estudos e

relatórios de mercado. Segundo o Fórum Econômico Mundial, no relatório *Jobs of Tomorrow* (2020), os cientistas de dados figuram entre as funções com maior taxa de crescimento global. De acordo com VOOE (2024), aponta um aumento de 27% na empregabilidade relacionada à área de dados entre 2018 e 2023. No Brasil o destaque é para o aumento expressivo na busca por vagas da área, com estudos que indicam que cerca de 33% dos profissionais consideram essa carreira uma das mais relevantes para o mercado nacional até 2030 (Insper, 2025).

A profissão Cientista de Dados continua sendo amplamente reconhecida por sua relevância estratégica no mercado, sendo inclusive descrita como a "profissão mais *sexy* do século XXI" (Davenport, 2012). Sendo um perfil essencialmente analítico e multidisciplinar, que integra conhecimentos de estatística, matemática, programação e técnicas de *Machine Learning* para solucionar problemas complexos e apoiar decisões baseadas em dados. Porém, essa alta demanda também mostra diferenças na forma como o mercado valoriza e remunera a profissão.

Dados recentes da plataforma Glassdoor (Glassdoor, 2025), indicam que a remuneração base no Brasil varia expressivamente, podendo oscilar entre aproximadamente R\$ 6 mil a R\$ 12 mil, e com

valores que podem atingir cerca de R\$ 17 mil em níveis mais avançados de experiência. Essa variação levanta o problema central deste estudo: como a alta demanda global por esse profissional se traduz, na prática, em termos de compensação financeira e oportunidades de carreira ao redor do mundo?

Diante desse contexto, o objetivo deste artigo é analisar e mapear o panorama salarial do cargo de Cientista de Dados, com o intuito de identificar os fatores sociodemográficos e corporativos que mais influenciam sua consolidação no mercado de trabalho. Para conduzir essa análise exploratória, o estudo investigará variáveis fundamentais como a remuneração, a localização geográfica (país), o nível de experiência do profissional e o porte da empresa. A partir dessa abordagem, pretende-se compreender de forma empírica como o mercado global diferencia a remuneração dos profissionais que atuam em uma área considerada estratégica na economia orientada por dados.

MATERIAL E MÉTODOS

O projeto foi desenvolvido com base no conjunto de dados *Jobs and Salaries in Data Science*, disponibilizado publicamente na plataforma Kaggle (Qaasim, 2023). Trata-se de uma base estruturada contendo por volta de 9.355 informações e 12 colunas referentes ao período de 2020 a 2023, contemplando diferentes aspectos das profissões na área de ciência de dados, com predominância norte-americana, tendo em vista que o conjunto de dados é originado dos Estados Unidos. Entre as variáveis disponíveis, destacam-se remuneração, nível de experiência, tipo de contrato, localização geográfica, porte da empresa e modalidade de trabalho, conforme descrito na Tabela 1.

Tabela 1. Variáveis do conjunto de dados, com tradução para o português

Colunas originais	Colunas traduzidas
<i>work_year</i>	ano_do_trabalho
<i>job_title</i>	titulo_do_cargo
<i>job_category</i>	cargos
<i>salary_currency</i>	moeda_salarial
<i>salary</i>	salário
<i>salary_in_usd</i>	salário_em_dólar
<i>employee_residence</i>	residência_do_funcionário
<i>experience_level</i>	nível_de_experiência
<i>employment_type</i>	tipo_de_contrato

<i>work_setting</i>	modelo_de_trabalho
<i>company_location</i>	localização_da_empresa
<i>company_size</i>	tamanho_da_empresa

As análises foram conduzidas utilizando a linguagem de programação Python (Python Software Foundation, 2026), amplamente empregada em projetos de análise de dados (Instituto Infnet, 2026), em ambiente Google Colab (Google, 2018) com suporte ao Jupyter Notebook. Esse ambiente possibilita a integração entre código, documentação e interpretação analítica, favorecendo a clareza do processo de análise.

Para a manipulação e organização dos dados, foi utilizada a biblioteca Pandas (McKinney, 2010), responsável pela leitura dos dados em formato de *dataframes*. Na etapa de visualização, foram empregadas as bibliotecas Matplotlib (Hunter, 2007) e Seaborn (Waskom, 2021), que possibilitam a construção de gráficos estatísticos e facilitam a identificação de padrões, distribuições e relações entre variáveis.

O desenvolvimento do projeto foi estruturado em etapas, inicialmente, realizou-se uma análise exploratória preliminar com o objetivo de compreender a estrutura do conjunto de dados. Nessa etapa foram verificadas as variáveis disponíveis, quantidade de registros e seus valores, identificação da tipagem das variáveis, bem como a dimensão do banco de dados e o consumo de memória. A análise preliminar foi fundamental para observar possíveis inconsistências, orientar o tratamento de dados e garantir a confiabilidade do estudo.

Na sequência, foi realizada a etapa de tratamento de dados, essencial para organização e padronização das informações, assegurando maior clareza e qualidade analítica. Com base no conhecimento geral do banco de dados, procedeu-se a verificação de valores nulos e registros duplicados, uma vez que tais inconsistências são capazes de distorcer análises de dados (Chen, 2024). Dessa forma, foram adotadas técnicas de limpeza e remoção desses registros, utilizando da análise dessas inconsistências e comandos como o *drop_duplicate* e *isnull*, ambos foram aplicados na cópia do conjunto de dados, para não realizar a alteração no *dataframe* original, garantindo com essas alterações maior robustez e segurança às etapas subsequentes.



Na etapa seguinte, foi implementada a transformação dos dados, envolvendo a alteração da tipagem das variáveis, quanto a padronização linguística. Durante essa etapa, identificou-se a presença de variáveis com tipo inadequado, o que ocasionava maior consumo da memória e inconsistências de análise. Para mitigar esses problemas, foi utilizado o método *select.dtypes* e *astype*, realizada a identificação e conversão de variáveis do tipo *object* para *category*, quando apropriado, fomentando eficiência no processamento.

Adicionalmente, foi realizada a tradução das colunas e de registros relevantes, como nível de experiência, tamanho da empresa, modalidade e categoria de trabalho e tipo de contrato do inglês para o português, com o propósito de alcançar o público-alvo almejado da pesquisa, majoritariamente familiarizado com a língua portuguesa. Como etapa final da transformação, aplicou-se a ordenação das informações com base na hierarquia, contribuindo para uma melhor organização dos dados e facilitando a interpretação dos resultados de forma clara.

Por fim, após as etapas de verificação, transformação e tratamentos das variáveis, obteve-se *dataframe* estruturado e adequado para análises mais aprofundadas. A partir desse ponto, foram explorados os registros para a construção de visualizações gráficas, com o objetivo de facilitar a interpretação dos dados e evidenciar tendências. Nesse processo, foram utilizados diferentes tipos gráficos: como histogramas, *boxplots*, gráficos de barras, linhas e setores, para analisar a distribuição das variáveis, padrões em determinado período e detectar valores atípicos (*outliers*).

Com base nos dados disponíveis, foram realizadas análises voltadas à compreensão da evolução da profissão de cientista de dados, incluindo o estudo do crescimento da área ao longo dos anos, a influência de fatores associados a contratação e modalidade de trabalho sobre a remuneração, além disso, foram conduzidas o cruzamento de variáveis categóricas e numéricas, como salário e nível de experiência, com o objetivo de calcular medidas estatísticas, incluindo medidas de tendência central (média e mediana) e de dispersão. Também foram explorados aspectos geográficos, como, por exemplo, a pesquisa de distribuição de profissionais por país, bem como possíveis vieses associados a essa distribuição. Além de, análises referentes ao porte da empresa,

remuneração e quantidade de funcionários, buscando compreender o impacto.

A partir dessas análises, o conjunto obtido possibilitou a construção de discussões abrangentes sobre o cenário estudado, bem como a elaboração de visualizações gráficas pertinentes, apresentadas na seção a seguir.

Cabe destacar que o conjunto de dados analisado representa registros salariais de profissionais da área de ciência de dados, e não necessariamente contratações efetivas. Dessa forma, os resultados devem ser interpretados como uma representação do cenário salarial e profissional, e não como um indicador direto do volume de vagas ou admissões no mercado.

RESULTADOS E DISCUSSÃO

Após a aplicação das etapas metodológicas, especialmente da análise exploratória dos dados, foram elaboradas, visualizações gráficas, considerando cerca de 1622 registros, 3 variáveis numéricas e 9 variáveis categóricas, distribuídas em 12 colunas, depois dos tratamentos de valores duplicados e nulos, tipagem das variáveis e seleção dos dados referentes ao cargo de Cientista de Dados e Pesquisa, com o objetivo de sintetizar os padrões observados e subsidiar a interpretação dos resultados. Essas representações permitiram identificar tendências e relações entre variáveis no conjunto de dados analisado.

A Figura 1 apresenta a evolução temporal do número de registros associados à área de cientista de dados e pesquisa, no período de 2020 a 2023, realizada por meio de um gráfico de linhas. Verifica-se uma tendência de crescimento contínuo, com intensificação ao longo dos anos, no número de profissionais de ciência de dados, com aumento mais acentuado entre 2022 e 2023. Considerando os dados analisados, o número de registros passou de aproximadamente 400 em 2022 para cerca de 1.200 em 2023, representando um crescimento aproximado de 72,2%, calculado a partir da variação relativa entre o número de registros observados em 2022 e 2023. Esse comportamento, sugere uma expansão recente da demanda por profissionais da área, hipótese compatível com o avanço da adoção de práticas orientadas por dados nas organizações.

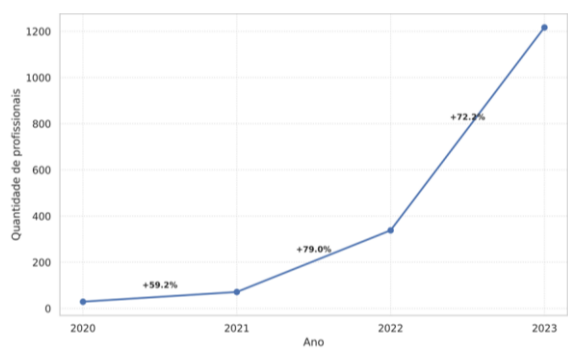


Figura 1. Crescimento anual da profissão Cientista de Dados e Pesquisa. Fonte: Elaborada pelos autores.

A Figura 2 apresenta a distribuição geográfica dos registros, considerando a localização das empresas por país no período de 2020 a 2023. Observa-se forte concentração nos Estados Unidos, que respondem por 86% das ocorrências analisadas, enquanto os demais países apresentam participações percentuais significativamente menores.

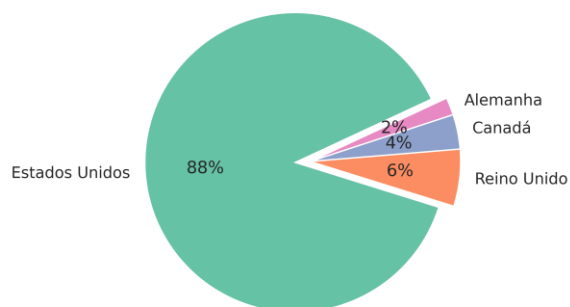


Figura 2. Países com maiores índices de contratações na área de Ciência de Dados e Pesquisa. Fonte: Elaborada pelos autores.

Esse resultado deve ser interpretado com cautela, pois pode refletir, ao menos em parte, um viés de composição do conjunto de dados, caso sua origem seja predominantemente norte-americana. Ainda assim, a concentração observada é compatível com o papel central dos Estados Unidos nos ecossistemas de tecnologia e inteligência artificial. Segundo o relatório AI Index 2025, elaborado pelo Stanford Human-Centered Artificial Intelligence (HAI), os Estados Unidos ampliaram sua liderança global em investimento privado em inteligência artificial, o que contribui para a centralidade do país no mercado tecnológico contemporâneo (Stanford Hai, 2025).

A Figura 3 apresenta a distribuição salarial para o cargo de Cientista de Dados e Pesquisa. Observa-se a

maior concentração de valores em faixas salariais intermediárias, com maiores valores em torno de 120 mil a 160 mil dólares anuais. A distribuição apresenta diferenças à direita, mostrando a presença de salários mais elevados em menor frequência. Esse comportamento sugere que existem casos de maior valorização dos salários. Como por exemplo, estimativas atuais do Glassdoor indicam remuneração média por ano de aproximadamente 154 mil dólares para cientistas de dados nos Estados Unidos, com mudanças conforme experiência e contexto de organizações (Glassdoor, 2025).

Dessa forma, os padrões no conjunto de dados mostram-se compatíveis com tendências no mercado, ainda que a comparação deva ser realizada com cuidado devido a diferenças de métodos entre os dados analisados.

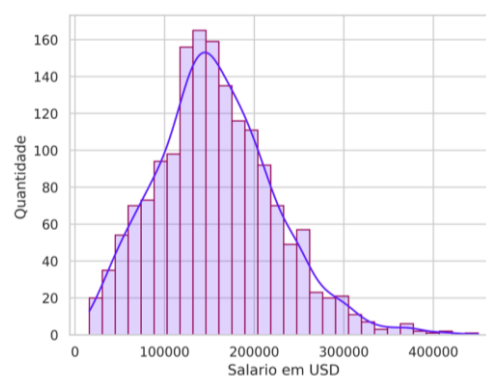


Figura 3. Distribuição salarial do cargo de Cientista de Dados e Pesquisa. Fonte: Elaborada pelos autores.

A Figura 4 apresenta a distribuição dos salários anuais em dólares (USD) segundo o nível de experiência para o cargo de Cientista de Dados e Pesquisa, permitindo uma análise mais detalhada das diferenças econômicas entre os grupos júnior, pleno, sênior e executivo. O *boxplot* mostra a mediana, a dispersão dos valores e a presença de diferenças, o auxilia a comparação entre os diferentes níveis de senioridade.

Os resultados indicam uma relação entre experiência e remuneração, com elevação constante das medianas à medida que aumenta a senioridade. Os grupos sênior e executivo concentram os maiores níveis salariais, enquanto os níveis pleno e sênior apresentam maior dispersão e presença de *outliers*. Esse comportamento propõe diferenças nessas categorias, relacionando-se a fatores como

especialização técnica, responsabilidades do cargo e características organizacionais.

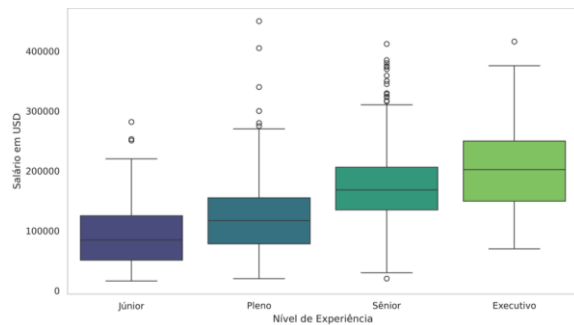


Figura 4. Distribuição salarial por nível de experiência. Fonte: Elaborada pelos autores.

A Figura 5 apresenta a distribuição dos registros por nível de experiência e modalidade de trabalho, considerando os regimes presencial, remoto e híbrido, para o cargo de Cientista de Dados e Pesquisa.

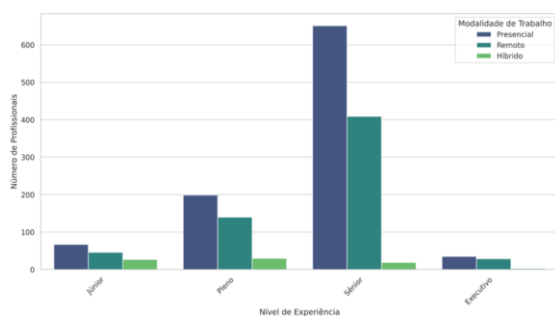


Figura 5. Distribuição dos registros por nível de experiência e modalidade de trabalho. Fonte: Elaborada pelos autores.

Observa-se predominância do regime presencial em todos os níveis analisados, especialmente entre profissionais seniores, embora o trabalho remoto também apresente participação relevante no conjunto de dados. Esse resultado deve ser interpretado com cautela, pois pode refletir características específicas da base analisada, e não necessariamente uma tendência universal do mercado. Ainda assim, a presença expressiva do trabalho remoto é compatível com transformações recentes nas formas de organização do trabalho, intensificadas após a pandemia de COVID-19. Evidências internacionais mostram que vagas com possibilidade de trabalho remoto cresceram de forma substancial entre 2020 e 2023, sobretudo em ocupações com forte

componente digital, como as relacionadas à ciência de dados.

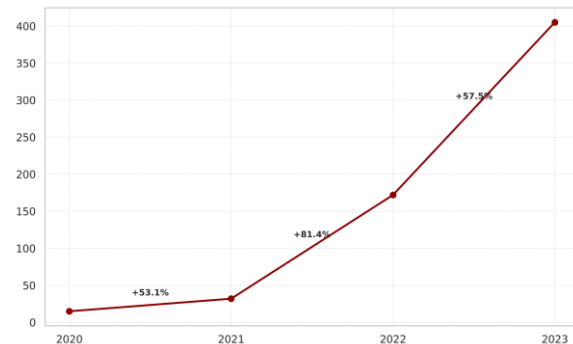


Figura 6. Crescente do trabalho remoto o período de 2020 a 2023 referente a profissão Cientista de Dados. Fonte: Elaborada pelos autores.

A Figura 7 sintetiza a relação entre a remuneração anual em dólares (USD) e diferentes variáveis explicativas, incluindo nível de experiência, modalidade de trabalho, tipo de contrato e porte da empresa, permitindo uma análise integrada dos fatores associados ao salário no cargo de Cientista de Dados e Pesquisa.

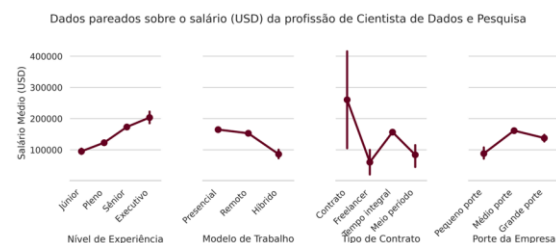


Figura 7. Dados pareados do salário médio anual em dólares (USD) e diferentes variáveis da profissão de Cientista de Dados e Pesquisa. Fonte: Elaborada pelos autores.

Em relação ao nível de experiência, observa-se progressão salarial consistente, com aumento das médias à medida que cresce a senioridade, indicando que a experiência constitui um dos fatores mais fortemente associados à remuneração. Quanto à modalidade de trabalho, as diferenças entre os regimes são menos acentuadas, embora o modelo presencial apresente valores ligeiramente superiores; esse resultado, entretanto, deve ser analisado com cautela, já que a Figura 5 mostra maior concentração de registros nessa categoria.

Comparado ao tipo de contrato, observa um padrão contínuo no conjunto de dados analisados, no qual



profissionais relacionados ao tempo integral mostram maiores médias salariais em comparação a outras formas de contratação, como freelancer e meio período. Além de que, verifica-se alta variação nos valores, evidenciado pelo aumento representado no gráfico com salários que se estendem aproximadamente de 100 mil até 400 mil (USD). Essa dispersão sugere a influência de fatores adicionais, como nível de experiência, especialização técnica e contexto organizacional, refletindo as mudanças do mercado de trabalho na área.

No que se refere ao porte das empresas os resultados indicam um comportamento padrão no conjunto de dados analisado, no qual empresas de médio porte apresentam médias salariais maiores às de grande porte no período de 2020 a 2023. Esse resultado deve ser analisado com precaução, pois pode estar associado a características específicas do conjunto de dados. Em outros observa-se uma tendência, na qual empresas de grande porte, como organizações globais de tecnologia e instituições financeiras, apresentam níveis salariais maiores para profissionais da área de ciência de dados (Glassdoor, 2025).

CONCLUSÃO

Concluindo, os resultados desse estudo mostram que o cargo de Cientista de Dados e Pesquisa possui crescimento constante no cenário profissional. Observa que a remuneração relacionada ao cargo é elevada, com valores que variam entre 100 e 300 mil dólares anuais, no contexto nacional, entre aproximadamente 7 mil e 20 mil reais mensais, isso demonstra ser influenciada por diferentes fatores, como, o nível de experiência e o tipo de contratação. As análises realizadas evidenciam a complexidade do mercado de trabalho, permitindo identificar padrões e tendências à distribuição salarial e às características profissionais. Esses resultados contribuem para uma compreensão mais estruturada da dinâmica da área, bem como dos fatores associados a melhores oportunidades e diferenciação no mercado.

A elaboração deste estudo também possibilitou reforçar a percepção de crescimento contínuo do cargo de Cientista de Dados e Pesquisa em escala global, acompanhada de valorização profissional. Embora os dados analisados estejam concentrados majoritariamente nos Estados Unidos e limitados até o ano de 2023, os resultados indicam tendências que extrapolam esse contexto específico.

Diante dessas limitações, propõe-se como continuidade da pesquisa a realização de análises com bases de dados mais recentes e geograficamente diversificadas, de modo a ampliar a compreensão do cenário global e validar os padrões identificados.

Por fim, os achados indicam que o cargo de Cientista de Dados e Pesquisa não se configura como uma tendência passageira, mas como uma área consolidada e em expansão, com relevância crescente no mercado de trabalho e potencial de desenvolvimento contínuo, especialmente diante do avanço das tecnologias orientadas a dados. Nesse sentido, a área se posiciona como um eixo central na transformação digital sendo uma das principais forças de inovação e competitividade em todo o mundo.

AGRADECIMENTOS

Os autores agradecem ao Centro Universitário SENAI São Paulo (UniSENAI), campus de Santo Amaro pelo apoio acadêmico, aos professores orientadores Dr. Felipe A. Camargo e Dra. Bethina R. Camargo pelo suporte no desenvolvimento do trabalho e à instituição Centro Educacional SESI pelo fomento da bolsa Passaporte Universitário que contribuíram para a realização dessa pesquisa.

REFERÊNCIAS

- MICHAEL, C. Problemas e soluções relacionados à duplicação de dados. **Oracle**, Austin, 2024. Disponível em: <https://www.oracle.com/br/data-duplication/>. Acesso em: 29 abr. 2026.
- DAVENPORT, T. H.; PATIL, D. Data Scientist: The Sexiest Job of the 21st Century. **Harvard Business Review**, Boston, 2012. Disponível em: <https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century>. Acesso em: 29 abr. 2026.
- GLASSDOOR. Salários de Data Scientist. **Glassdoor**, San Francisco, 2025. Disponível em: https://www.glassdoor.com.br/Sal%C3%A1rios/%C3%A9tats-unis-data-scientist-sal%C3%A1rio-SRCH_IL.0,10_IN1_KO11,25.htm. Acesso em: 29 abr. 2026.
- GOOGLE. Google Colaboratory: Ambiente de computação em nuvem. **Google**, Mountain View, 2018. Disponível em: <https://colab.research.google.com/>. Acesso em: 29 abr. 2026.
- HUNTER, J. D. Matplotlib: A 2D graphics environment. **Computing in Science &**



Engineering, Piscataway, v. 9, n. 3, p. 90-95, 2007. Acesso em: 29 abr. 2026.

INSPER. Ciência de dados é uma das áreas com maior demanda de profissionais. **Insper**, São Paulo, 2025. Disponível em: <https://www.insper.edu.br/pt/conteudos/dados/ciencia-de-dados-maior-demanda-profissionais>. Acesso em: 29 abr. 2026.

INSTITUTO INFNET. As linguagens de programação para ciência de dados. **Instituto Infnet**, Rio de Janeiro, 2026. Disponível em: <https://blog.infnet.com.br/data-science/as-linguagens-de-programacao-para-ciencia-de-dados-mais-usadas/>. Acesso em: 29 abr. 2026.

MCKINNEY, W. Data structures for statistical computing in Python. **Proceedings of the 9th Python in Science Conference**, Austin, p. 51-56, 2010. Acesso em: 29 abr. 2026.

PYTHON SOFTWARE FOUNDATION. Python. **Python Software Foundation**, Wilmington, 2026. Disponível em: <https://www.python.org>. Acesso em: 29 abr. 2026.

QAASIM, H. Jobs and salaries in data science. **Kaggle**, San Francisco, 2023. Disponível em: <https://www.kaggle.com/datasets/hummaamqaasim/jobs-in-data>. Acesso em: 29 abr. 2026.

SINGH, D. Data Analyst Job Outlook: Salaries, trends and opportunities. **Skillify Solutions**, Toronto, 2026. Disponível em: <https://skillifysolutions.com/blogs/data-science/data-analyst-job-outlook/>. Acesso em: 29 abr. 2026.

STANFORD HAI. The 2025 AI Index Report. **Stanford Institute for Human-Centered Artificial Intelligence**, Stanford, 2025. Disponível em: <https://hai.stanford.edu/ai-index/2025-ai-index-report>. Acesso em: 29 abr. 2026.

VOEE. An analysis of Data Science Occupations, Skills Demand, and Graduate Outcomes in Virginia. **Virginia Office of Education Economics**, Richmond, 2025. Disponível em: <https://voee.org/wp-content/uploads/2025/03/Data-science-FINAL.pdf>. Acesso em: 29 abr. 2026.

WASKOM, M. Seaborn: statistical data visualization. **Journal of Open Source Software**, Austin, v. 6, n. 60, p. 3021, 2021. Acesso em: 29 abr. 2026.

WORLD ECONOMIC FORUM. Jobs of Tomorrow: Mapping Opportunity in the New Economy. **World**

Economic Forum, Genève, 2020. Disponível em: https://www3.weforum.org/docs/WEF_Jobs_of_Tomorrow_2020.pdf. Acesso em: 29 abr. 2026.