



ANÁLISE EXPLORATÓRIA DO CARGO DE ENGENHEIRO DE DADOS

Juan Matheus O. Lopes¹, Julia M. S. Santana², João Pedro C. Santos³, Kevin Joel G. Alaca⁴, Bethina R. Camargo⁵, Felipe A. Camargo⁶

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

Resumo: Este trabalho analisa o cargo de Engenharia de Dados com base no conjunto de dados *Jobs and Salaries in Data Science*, do Kaggle, usando Python. A distribuição salarial apresenta assimetria à direita, tornando a mediana mais representativa que a média. Há concentração de registros nos Estados Unidos e crescimento de registros em níveis mais elevados de experiência. Por ser exploratório, o estudo não estabelece relações causais, mas oferece uma visão inicial sobre o perfil do cargo.

Palavras-chave: Remuneração; Diferença salarial; Nível de experiência; Mercado de trabalho; Ciência de dados.

INTRODUÇÃO

A área de ciência de dados tem assumido papel cada vez mais relevante nas organizações contemporâneas, especialmente em contextos nos quais a tomada de decisão depende da coleta, do tratamento e da interpretação de grandes volumes de dados. Nesse cenário, a capacidade de transformar dados em informação útil e valor estratégico tornou-se um diferencial competitivo (Provost e Fawcett, 2013), ao passo que transformações recentes no mercado de trabalho indicam crescimento da importância de funções ligadas à tecnologia e ao uso intensivo de dados (World Economic Forum, 2025).

Nesse contexto, diferentes profissionais desempenham funções complementares. Enquanto analistas de dados interpretam informações para apoiar decisões e cientistas de dados desenvolvem modelos preditivos, cabe ao Engenheiro de Dados construir e manter as infraestruturas capazes de coletar, armazenar, processar e disponibilizar dados de forma confiável e escalável, papel estruturante sem o qual as demais atividades analíticas não seriam possíveis. A organização desses fluxos é, portanto, essencial para viabilizar análises consistentes e processos decisórios baseados em dados (Kleppmann, 2017).

Diante da importância desse cargo e da crescente demanda por profissionais qualificados na área, este trabalho realiza uma análise exploratória do cargo de Engenheiro de Dados com base em um conjunto de dados público, buscando descrever a distribuição salarial, identificar a concentração geográfica dos registros e analisar a relação entre nível de experiência e remuneração ao longo do tempo.

MATERIAL E MÉTODOS

A pesquisa foi desenvolvida com base no conjunto de dados público *Jobs and Salaries in Data Science* (Qaasim, 2023), disponível na plataforma Kaggle, composto por 9.355 registros e 12 variáveis, abrangendo profissionais da área de dados de diferentes países entre 2020 e 2023. As principais variáveis utilizadas neste estudo foram: ano de trabalho (*work_year*), cargo (*job_title*), nível de experiência (*experience_level*), remuneração padronizada em dólar (*salary_in_usd*) e localização da empresa (*company_location*).

As análises foram conduzidas em ambiente Python 3.12.13, utilizando as bibliotecas Pandas (McKinney, 2010) para manipulação e tratamento dos dados, e Matplotlib (Hunter, 2007) e Seaborn (Mwaskom, 2021) para visualização gráfica.

Inicialmente, realizou-se uma inspeção descritiva da base, seguida por etapas de tratamento, incluindo a seleção dos registros referentes ao cargo de Engenheiro de Dados, resultando em 2.260 observações, e a remoção de duplicatas, com o método *drop_duplicates*, que reduziu o conjunto final para 1.160 registros, sem valores ausentes. Para garantir comparabilidade, utilizou-se a variável de remuneração já padronizada em dólar.

Posteriormente, foram realizadas análises exploratórias com o objetivo de descrever a distribuição dos salários, a concentração geográfica dos registros e a relação entre experiência profissional e remuneração ao longo do período analisado.

RESULTADOS E DISCUSSÃO

A distribuição salarial observada apresenta assimetria à direita, como apresentado na Figura 1, com presença de valores elevados que impactam a média. Nesse contexto, embora a média anual seja de aproximadamente US\$ 143.000, a mediana (US\$ 135.000) mostra-se mais adequada para representar o comportamento típico dos salários, por ser menos sensível a valores extremos.

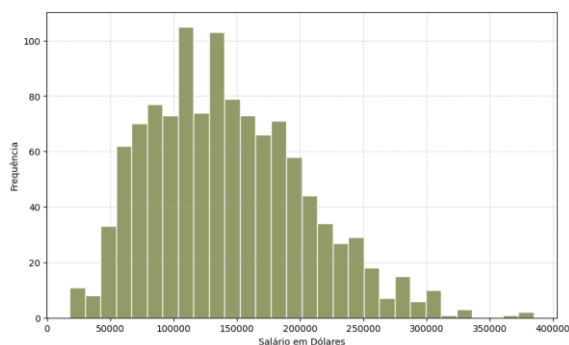


Figura 1. Histograma de distribuição salarial do cargo de engenheiro de dados. Fonte: Elaborada pelos autores.

Com base na Figura 2, a análise da distribuição geográfica indica forte concentração de registros nos Estados Unidos, que representam aproximadamente 85,9% das observações do conjunto de dados. No entanto, esse resultado deve ser interpretado com cautela, pois pode refletir características de coleta e cobertura do conjunto de dados, e não necessariamente a distribuição global real das oportunidades.

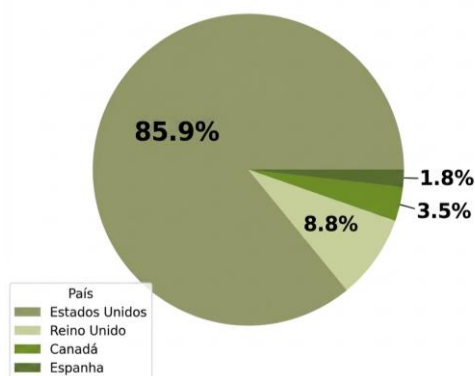


Figura 2. Proporção das oportunidades de trabalho para engenheiros de dados por país, com destaque para a predominância dos Estados Unidos. Fonte: Elaborada pelos autores.

Para facilitar a visualização dos demais países, a Figura 3 apresenta uma representação gráfica excluindo os Estados Unidos. Esse procedimento, contudo, deve ser entendido apenas como recurso visual complementar, não como representação mais fiel da realidade. Entre os países restantes, o Reino Unido se destaca com o maior número de registros,

seguido por Canadá e Espanha, sugerindo que as oportunidades na área fora dos Estados Unidos estão concentradas predominantemente em países de língua inglesa e na Europa.

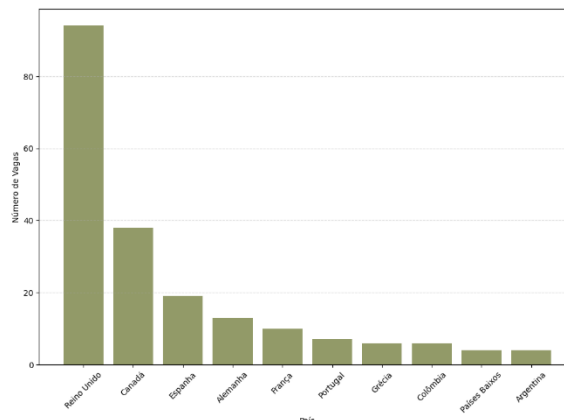


Figura 3. Distribuição das oportunidades de trabalho para engenheiros de dados por país. Fonte: Elaborada pelos autores.

Quanto à experiência profissional, a Figura 4A evidencia crescimento no número de registros associados a níveis mais elevados, especialmente no nível sênior, que apresentou um aumento de 150% em 2023 em relação ao ano anterior. Adicionalmente, a Figura 4B ilustra que níveis mais altos de experiência estão associados a maiores médias salariais, embora não seja possível estabelecer relação causal a partir da análise realizada. Destacam-se o aumento de aproximadamente 33% na média salarial no nível sênior entre 2020 e 2023, e o crescimento linear de aproximadamente 175% no nível executivo entre 2020 e 2022.

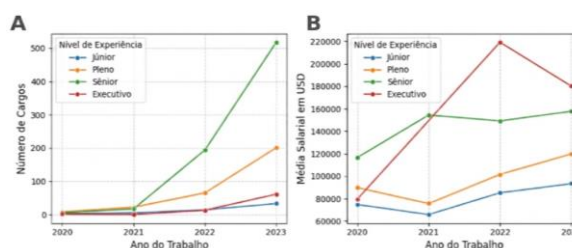


Figura 4. (A) Número de cargos anual por nível de experiência (2020 - 2023); (B) Média salarial anual por nível de experiência (2020 - 2023). Fonte: Elaborada pelos autores.

CONCLUSÃO

A análise exploratória permitiu identificar padrões relevantes no conjunto de dados, como a assimetria na distribuição salarial, a concentração geográfica dos registros e a associação entre níveis mais elevados de experiência e maiores remunerações.



No entanto, por se tratar de um estudo exploratório, não é possível atribuir causalidade a fatores externos específicos. Os resultados devem ser interpretados como uma descrição dos dados analisados, e não como representação definitiva do mercado.

Como limitações, destacam-se possíveis vieses de cobertura do conjunto de dados, especialmente em relação à concentração geográfica, além da ausência de variáveis complementares que poderiam aprofundar a análise.

Para estudos futuros, recomenda-se a inclusão de variáveis como habilidades técnicas, formação acadêmica e modalidade de trabalho, visando ampliar a compreensão sobre o perfil do cargo.

AGRADECIMENTOS

Os autores agradecem à UniSENAI Suíço-Brasileira pelo suporte institucional ao desenvolvimento deste trabalho, ao Professor Dr. Felipe de Almeida Camargo pela orientação e contribuições ao longo da análise, e ao SESI pelo apoio financeiro por meio do Passaporte Universitário, que viabilizou a participação dos estudantes nesta iniciativa científica.

REFERÊNCIAS

- HUNTER, John D. Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, v. 9, p. 90–95, 2007.
- KLEPPMANN, Martin. *Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems*. O'Reilly Media, Sebastopol, 2017.
- MCKINNEY, Wes. Data structures for statistical computing in Python. *Proceedings of the 9th Python in Science Conference*, p. 51–56, 2010.
- MWASKOM, Michael. Seaborn: statistical data visualization. *Journal of Open Source Software*, v. 6, n. 60, p. 3021, 2021.
- PROVOST, Foster; FAWCETT, Tom. *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking*. O'Reilly Media, Sebastopol, 2013.
- QAASIM, Hummaam. *Jobs and Salaries in Data Science*. Kaggle, 2023. Disponível em: <https://www.kaggle.com/datasets/hummaamqaasim/jobs-in-data>. Acesso em: 12 abr. 2026.
- WORLD ECONOMIC FORUM. *The Future of Jobs Report 2025*. World Economic Forum, Geneva, 2025.