



ANÁLISE EXPLORATÓRIA DO DESENVOLVIMENTO DO CARGO DE LIDERANÇA NA ÁREA DE CIÊNCIA DE DADOS

Kauã F. L. Pereira¹, Rafael C. S. Silva², Júlio C. D. Peres³, Marcos V. G. Teixeira⁴, Bethina R. Camargo⁵, Felipe A. Camargo⁶

^{1,2,3,4,5,6} Centro Univesitário SENAI São Paulo (UniSENAI) - Campus Santo Amaro, São Paulo, Brasil (kaua.pereira34@senaisp.edu.br)

Resumo: Este estudo analisa a evolução das carreiras em Ciência de dados para cargos de liderança com base em dados de 2020 e 2023. O objetivo é compreender a progressão profissional e identificar fatores associados à ascensão na área. A análise exploratória evidenciou predominância de profissionais em nível sênior e crescimento salarial conforme a experiência, além da concentração de cargos estratégicos em funções de liderança.

Palavras-chave: Ciência de Dados, Progressão de Carreira, Liderança em Dados, Gestão de Pessoas, análise de dados.

INTRODUÇÃO

Atualmente, há um crescimento exponencial na geração de dados provenientes de diferentes fontes como países, empresas e pessoas e toda esta informação, está armazenada em grandes bancos de dados (*data centers*) e na internet. Grandes volumes de dados são gerados a cada instante, tornando o processamento dessas informações algo de extrema importância para o mundo atual, sendo algo de suma importância para a longevidade de algumas organizações. Esse aumento de dados motivou o surgimento e a rápida ascensão de cargos especializados (Alura, 2023). Segundo o World Economic Forum (2020), lista a área de dados como uma das maiores crescentes no mercado de trabalho no ano de 2020, e a *Harvard Business Review* classifica a área de dados como a mais atrativa do século XXI (Davenport e Harris 2017). Atualmente, podemos observar que a Ciência de Dados tem se consolidado como um componente essencial em sistemas de software orientados a dados, abrangendo análises descritivas, preditivas e prescritivas.

No Brasil, a área é apontada como uma das que apresentam maior demanda por profissionais, com crescimento expressivo na busca por especialistas nos últimos anos (Insper, 2025). Esse cenário reflete a valorização de competências capazes de organizar dados complexos e gerar análises aplicáveis ao contexto corporativo. Este cenário reforça a relevância de cargos como; Analistas de Dados, *Business Intelligence* (BI), Cientista de Dados, Engenheiro de dados, e Liderança de Dados, uma vez

que o ambiente organizacional atual, é fortemente orientado por dados.

A importância desses profissionais surge na capacidade de transformar dados brutos em conhecimento viável e decisões estratégicas de forma assertivas. Segundo Naomi (2023), a adoção de uma cultura que baseia questões estratégicas, processos e ações através da análise de dados concretos, em vez de decisões baseadas exclusivamente em intuição. Assim, permitindo que as empresas consigam se sobressair em relação as quais não operam desta maneira, atingindo resultados mais certos (Bean, 2021; Davenport e Harris, 2017). Segundo Müller (2018), empresas que investem em Ciência de Dados tiveram aumento de produtividade entre 3% e 7%. Assim, ao integrar grandes volumes de dados (*Big Data*) e Inteligência Artificial (IA), as organizações não apenas inovam em seus produtos, mas alcançam um desempenho substancialmente superior em suas operações gerais.

Apesar do alto crescimento do mercado de dados, surge uma categoria tão importante quanto as funções focadas estritamente em programação e engenharia. Sendo esta categoria a liderança estratégica em dados. No cenário brasileiro, cargos voltados para Liderança e Gestão estão em alta, sendo responsáveis pela gestão de times, definição do plano de metas (*roadmap*) e na tomada de decisão estratégica. Observa-se que no âmbito profissional, a promoção não ocorre de forma súbita, mas gradual. O processo de construção de um líder em dados, inicia-se no amplo conhecimento técnico, além disso, habilidades



como comunicação clara, organização e tomada de decisão, se tornam de suma importância para ser tornar um bom líder. A liderança dentro da área de dados, ocorre até mesmo antes da formalização do cargo.

Neste contexto, o papel do líder pode mudar de empresa para empresa, mas de modo geral, ele atua como uma “ponte” essencial, garantindo que a área técnica não trabalhe de forma desconecta e esteja inteirada sobre a realidade da empresa. Atualmente, a liderança se diversifica em níveis como *Tech Lead* (foco em arquitetura), Coordenador (gestão e viabilidade técnica), Gerente (estratégia e pessoas) e Diretor (visão macro do negócio).

Contudo, o problema central reside na ausência de uma trilha de carreira que oriente os profissionais que almejam chegar nestes cargos. Como esta área é mais “recente” quando comparada com as profissões mais tradicionais. As especializações necessárias, *softskills* ainda não estão consolidadas na literatura, tornando as exigências muito divergentes de uma empresa para outra. Além disso, quando se fala em liderança, ainda persiste os estereótipos do executivo sênior, já em final de carreira. O objetivo central deste artigo, é exemplificar os caminhos de progressão dentro da área de dados, mostrando que existem oportunidades de liderança que vão além do executivo tradicional. Além disso, pretendemos mostrar quais competências são solicitadas para quem almeja obter tal cargo, como gestão de pessoas, conhecimento de diversas ferramentas, como por exemplo BI.

MATERIAL E MÉTODOS

O presente trabalho foi desenvolvido com o objetivo de analisar de forma exploratória o mercado de trabalho na área de Ciência de Dados. Para garantir a precisão dos resultados e a reprodutibilidade do estudo, adotou-se um fluxo de trabalho dividido em etapas: coleta, tratamento, tradução e visualização de dados.

O desenvolvimento do código foi centralizado na linguagem Python (Rossum, 2022), escolhida por sua robustez e ampla biblioteca de suporte estatístico. Como plataformas de execução, foram utilizados o Google Colab (Google, 2024) e o Jupyter Notebook (Project jupyter, 2024). Ambas as plataformas permitiram a criação de um ambiente interativo no qual blocos de código e descrições textuais coexistem, facilitando a documentação das etapas do processo.

Para o desenvolvimento do projeto, três bibliotecas fundamentais foram empregadas:

- **Pandas:** Foi utilizada como base para a manipulação de dados, permitindo a estruturação de DataFrames e a execução de cálculos de forma otimizada (McKinney, 2022).
- **Matplotlib:** Foi utilizada para a construção dos gráficos e controle dos elementos visuais (Hunter, 2003).
- **Seaborn:** Foi aplicada para aprimorar para elevar a estética das visualizações, oferecendo gráficos estatísticos mais elaborados e padrões de cores que facilitam a interpretação dos dados (Waskom, 2010).

O fluxo de trabalho seguiu uma sequência lógica para garantir que a base de dados fosse não apenas limpa, mas também compreensível para o público geral.

O banco de dados utilizado neste estudo foi o *Jobs and Salaries in Data Science*, obtido publicamente na plataforma Kaggle (Qasim, 2023). O processo iniciou-se com o carregamento dos dados. Em seguida, realizou-se uma análise exploratória inicial para identificar a estrutura do banco de dados, incluindo a quantidade de linhas, e colunas disponíveis e os tipos de variáveis (numéricas ou categóricas). Essa fase foi essencial para obtenção de *insights* preliminares, como as variações salariais médias e a distribuição de profissionais por cargo. Sem tratamento, os dados podem ter inconsistências, como valores nulos ou ausentes. Por isso, foi realizada uma etapa de limpeza, na qual utilizou-se o método *drop_duplicates*, como resultado, ele removeu todas as linhas que apresentavam duplicatas, garantindo que cada registro de salário fosse único. Além deste comando, utilizou-se também o método *isnull* para verificar a presença de valores nulos. Esse processo garantiu uma base de dados consistente, evitando distorções que comprometessem a precisão das análises.

Na etapa seguinte, realizou-se a transformação e padronização dos dados através da técnica de mapeamento de dicionários, usando a biblioteca Pandas como principal ferramenta. As variáveis categóricas originalmente em inglês foram traduzidas para o português, incluindo termos técnicos, cargos e categorias. Neste caso, foi utilizado o método *.map()*, com o objetivo de padronizar a base e facilitar a análise dos resultados. Além disso, foi realizada a adequação dos tipos de dados (tipagem) via *astype*, assegurando compatibilidade com as operações estatísticas e diminuindo o armazenamento necessário para gerir esse banco de dados. Para variáveis que exigiam uma hierarquia lógica, como o nível de experiência, utilizou-se uma técnica de categorização ordenada, garantindo uma correta geração das visualizações gráficas, contribuindo para a consistência e confiabilidade da análise.

Ao final desse processo, os dados foram organizados e estruturados, possibilitando sua utilização em análises estatísticas. A combinação das etapas de limpeza e padronização permitiu que a análise seguinte se concentrasse no comportamento do mercado de Ciência de Dados, especialmente em relação às tendências salariais.

RESULTADOS E DISCUSSÃO

O banco de dados original continha 9.355 linhas e 12 colunas. Após o processo de limpeza, remoção de duplicatas idênticas e valores nulos, a base final utilizada para a análise resultou em 5.339 linhas, garantindo que não houvesse sobreposição de dados salariais. As variáveis presentes na base incluem informações como 'Ano de Trabalho', 'Nível de Experiência', 'Tipo de Emprego', 'Título do Cargo', 'Salário', 'Moeda', 'Residência do Funcionário', 'Local da Empresa' e 'Porte da Empresa'. Entre elas, as variáveis contínuas são representadas pelo 'Salário' e 'Salário em USD', enquanto as demais são classificadas como categóricas. É importante ressaltar que o banco de dados é de abrangência global, contendo dados de diversos países e regiões do mundo, o que permitiu uma análise ampla do comportamento do mercado de Ciência de Dados internacional.

A Figura 1 representa a proporção dos profissionais da área de dados por nível de experiência. Existe uma predominância do nível sênior correspondente a 63,5% de amostra, os níveis pleno e executivo, apresentaram, respectivamente, 20,5% e 13,1%, enquanto o nível júnior representou 2,8%, configurando a menor parcela.

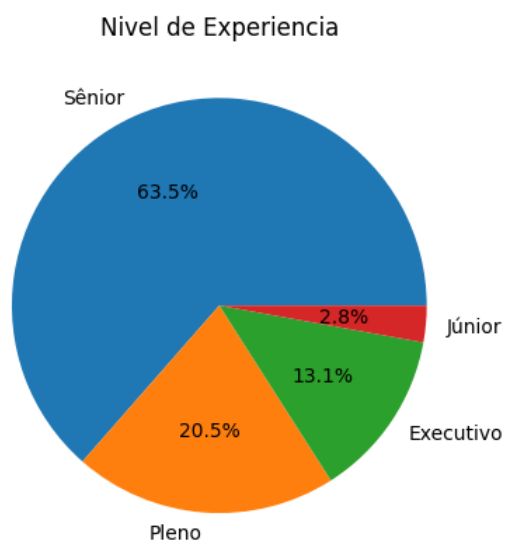


Figura 1. Proporção dos profissionais por nível de experiência. Fonte: Elaborada pelos autores.

Com base nestes resultados, observou-se que a estrutura do mercado de dados é composta em sua maioria por profissionais com experiência consolidada, fato demonstrado pela predominância do nível Sênior. A baixa representatividade do nível Executivo mostra que há uma maior dificuldade de acesso a estes cargos. Além disso, o baixo número de profissionais de nível Júnior, aponta que as posições de entrada são mais restritas, mas que aqueles que entram, conseguem se tornar sênior. Verificou-se também que a presença de profissionais do nível executivo indica uma exceção visto que tais encargos correspondem a uma posição mais estratégica como diretores, e alta gestão, sendo pouco frequentes em níveis inferiores, como o júnior. Desta forma, concluiu-se que, embora os dados indiquem predominância de profissionais mais experientes, existem exceções ao longo dos níveis hierárquicos.

Na Figura 2, apresenta a quantidade de profissionais por cargo de liderança. Observa-se maior concentração em funções como Engenheiro de Dados, Gerente de Dados e Líder de Dados, indicando que a liderança na área está associada a cargos estratégicos e técnicos de maior complexidade.

A análise destaca os cargos de liderança mais demandados no conjunto de dados, com o cargo de engenheiro de dados liderando o gráfico com cerca de 212 ocorrências. Na sequência, aparecem os cargos de gerente de dados 85 e líder de dados 42,

Enquanto funções de suporte e especialidades menores apresentam frequência reduzida na amostra.

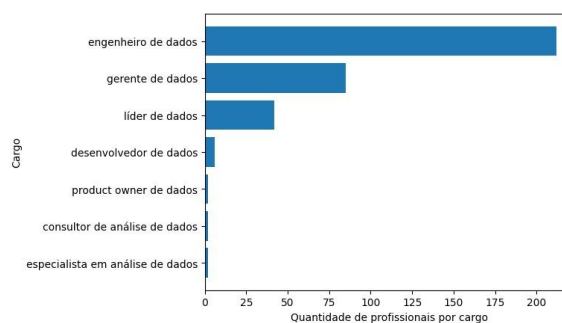


Figura 2. Gráfico dos cargos com maior demanda. Fonte: Elaborada pelos autores.

Observou-se ainda que cargos como Desenvolvedor de Dados e Especialista em Análise de Dados, apresentaram menor número de ocorrências, por se caracterizarem como posições mais técnicas e operacionais. Indica que o conjunto de dados se concentra em cargos de maior nível hierárquico e com foco em liderança, os quais, por sua vez, estão associados a uma progressão de carreira que se inicia em funções técnicas, como analistas e

desenvolvedores, refletindo uma estrutura consolidada na área de dados.

Na Figura 3, apresentou-se a distribuição salarial por nível de experiência em cargos de liderança, considerando valores em dólares (USD), observou-se a presença de cargos de liderança em todos os níveis de experiência, embora níveis inferiores representem exceções. As médias salariais indicam um crescimento consistente dentro destas carreiras, partindo de USD 81.012 para júnior, USD 122.611 para pleno, USD 153.097 sêniores e USD 185.376 para executivos, o que identifica uma progressão expressiva dentro dos diferentes níveis de experiência mesmo em posições de liderança.

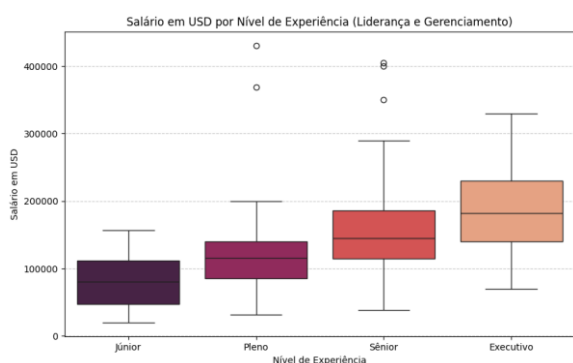


Figura 3. Gráfico do salário por nível de experiência. Fonte: Elaborada pelos autores.

A análise das distribuições salariais permitiu identificar a presença de valores discrepantes (*outliers*), possivelmente associados a fatores como porte da empresa, nível de responsabilidade do cargo ou localização geográfica. Além disso, observou-se um aumento expressivo nos salários pelo nível de experiência, consolidação dos cargos de liderança, e evidenciando que, mesmo nesses cargos, o nível de experiência influencia diretamente a variação salarial.

CONCLUSÃO

A partir das análises exploratórias realizadas neste estudo, identificou-se a estruturação do desenvolvimento da liderança na área de Ciência de Dados, bem como fatores que influenciam a progressão profissional para cargos de gestão e liderança. Os resultados mostraram que cargos de liderança estão relacionados a níveis de experiência, com maior concentração de profissionais no nível sênior, embora tenham sido observadas exceções em níveis inferiores. Além disso, observou-se a concentração de cargos de liderança em funções específicas, como Engenheiro de *Analytics*, Gerente de Dados e *Head* de Dados, indicando que, na área de dados a liderança não é atribuída de forma

homogênea, mas sim organizada em posições estratégicas dentro das organizações. Dessa forma, considera-se que o objetivo do estudo foi alcançado, uma vez que foram identificadas as formas de progressão para cargos de liderança na área de dados, além das competências necessárias para alcançar essas posições, dentre elas o conhecimento técnico, a visão de negócio, a comunicação e a gestão de pessoas.

Entre os principais achados, destaca-se a análise salarial que demonstrou um aumento significativo de salários conforme o aumento no nível de experiência, evidenciando que, mesmo em cargos de liderança, a experiência profissional permanece sendo um fator determinante para a remuneração. Ademais, foram identificadas discrepâncias salariais (*outliers*), possivelmente relacionadas ao porte da empresa, nível de responsabilidade ou localização geográfica.

Como limitações do trabalho, destaca-se que a análise foi realizada com base em um único banco de dados, o que pode limitar a representatividade dos resultados em relação ao mercado de trabalho global, além de não considerar fatores como cultura organizacional, porte da empresa e localização geográfica, que acabam por influenciar tanto a progressão de carreira quanto a remuneração. Para trabalhos futuros, sugere-se a ampliação deste estudo por meio da inclusão de novas variáveis, como porte da empresa, país, nível de escolaridade e área de especialização. Destaca-se a importância da utilização de dados mais recentes, a fim de obter uma análise mais atual do mercado de trabalho, bem como a realização de estudos mais aprofundados sobre as habilidades técnicas e comportamentais mais exigidas para cargos de liderança na área de dados.

AGRADECIMENTOS

Gostaríamos de expressar nosso agradecimento à UniSENAI Suíço-Brasileira pela oportunidade concedida de participar do evento COBICET. O apoio institucional foi fundamental para a realização deste estudo sobre o desenvolvimento de cargos de liderança na área de Ciência de Dados. Além disso, agradecemos de forma especial ao professor Felipe Camargo pelo auxílio, orientação e suporte técnico ao longo de todo o desenvolvimento deste projeto.

REFERÊNCIAS

- ALURA. Carreira em dados: quais as habilidades necessárias e como está o mercado de trabalho. Disponível em: <https://www.alura.com.br/artigos/carreira-em-dados>. Acesso em: 31 mar. 2026.
- BEAN, R. Fail Fast, Learn Faster: Lessons in Data-Driven Leadership in an Age of Disruption, Big Data, and AI. 2021.



DAVENPORT, T. H.; HARRIS, J. G. Competing on Analytics: The New Science of Winning. 2017.

HUNTER, J. D. Matplotlib: a 2D graphics environment. 2003. Disponível em: <https://matplotlib.org/>. Acesso em: 31 mar. 2026.

INSPER. Ciência de dados é uma das áreas com maior demanda de profissionais. Disponível em: <https://www.insper.edu.br/pt/conteudos/dados/ciencia-de-dados-maior-demanda-profissionais>. Acesso em: 02 abr. 2026.

MÜLLER, O.; FAY, M.; VOM BROCKE, J. The effect of big data and analytics on firm performance: an econometric analysis considering industry characteristics. 2018. Disponível em: <https://www.researchgate.net/publication/323773046>. Acesso em: 31 mar. 2026.

NAOMI, V.; JOHNSON, N. Percepção dos profissionais da área de dados sobre a carreira no Brasil. 2023. Disponível em: <https://cadernopaic.fae.edu/cadernopaic/article/view/548>. Acesso em: 31 mar. 2026.

QAASIM, H. Jobs and Salaries in Data Science. 2023. Disponível em: <https://www.kaggle.com/datasets/hummaamqaasim/jobs-in-data>. Acesso em: 31 mar. 2026.

ROSSUM, G. Python. Disponível em: <https://www.python.org/>. Acesso em: 31 mar. 2026.

WASKOM, M. Seaborn: statistical data visualization. Disponível em: <https://seaborn.pydata.org/>. Acesso em: 31 mar. 2026.

MCKINNEY, Wes. Python for Data Analysis. 3. ed. Sebastopol: O'Reilly Media, 2022. Disponível em: <https://wesmckinney.com/book>. Acesso em: 12 abr. 2026.