

As habilidades em análise de dados mais requisitadas no mercado de trabalho

Déborah Lima Scalioni^{1*}; Renato Máximo Sátiro

SOBRE OS AUTORES

¹Especialista em Data Science e Analytics. Rua Antônio Jordão, 443, Residencial Jardim Califórnia, CEP: 13857-000, Estiva Gerbi, SP, Brasil

² Universidade Federal de Goiás - UFG. Doutor no Programa de Pós Graduação em Administração. Avenida Esperança s/n, Câmpus Samambaia - Prédio da Reitoria. 74690-900 Goiânia, GO, Brasil.

*Autor correspondente: deborahscalioni@gmail.com

COMO CITAR

Scalioni D.L.; Sátiro R.M.; As habilidades em análise de dados mais requisitadas no mercado de trabalho. Revista E&S. 2025; 6: e20240033.



Devido à transformação digital, impulsionada com o advento da pandemia, as empresas geram enormes quantidades de dados todos os dias. Nesse contexto, tem aumentado a competitividade entre as empresas na busca por profissionais de análise de dados que possuam as habilidades necessárias na área. Os empregos de engenheiro de dados e cientista de dados estão entre as ocupações mais procuradas em 2020, com taxas de crescimento anual de 78% e 75%, respectivamente^[1]. Esses profissionais são responsáveis por construir toda a infraestrutura de dados e implementar métodos para capturar, processar e realizar análises preventivas e preditivas dentro da organização. A manipulação correta de grandes volumes de dados pode gerar poderosos insights obter para a empresa, guiando-a para as melhores tomadas de decisão^[1].

No entanto, esse crescimento substancial da área, associado à sua multidisciplinariedade, gerou uma confusão em relação a quais são as habilidades esperadas desses profissionais, e em relação a definições básicas de ferramentas, métodos e funções. Tais equívocos acabam por gerar altos custos para empregadores e profissionais, que precisam lidar com a falta de clareza nas descrições dos cargos nos processos seletivos^[2].

Neste sentido, observa-se que empregadores buscam profissionais que combinem habilidades práticas e cognitivas em análise de dados, computação e negócios, com competências sociais e características de personalidade específicas. A demanda por essas habilidades varia de acordo com o nível do cargo e as responsabilidades envolvidas. No entanto, muitos executivos reconhecem que os candidatos raramente possuem todas as competências desejadas, o que os obriga a investir em treinamento adicional para preparar esses profissionais adequadamente para as funções^[3]. Ou seja, observa-se que a busca das organizações por profissionais com habilidades técnicas e comportamentais se dá por elas serem fundamentais no processo de transformação digital, uma vez que fornecem informações que dão embasamento à tomada de decisão rápida e assertiva que garantem que as empresas se mantenham competitivas no mercado.

Ao abordar essa temática Dubey e Gunasekaran^[4] listaram alguns exemplos de quais são essas competências. As chamadas de *hard skills* são as habilidades técnicas que podem ser aprendidas, como estatística, previsão, otimização, finanças, contabilidade financeira, marketing; enquanto as *soft skills* são habilidades que estão mais ligadas à personalidade, como liderança, trabalho em equipe, escuta, aprendizagem, atitude positiva, comunicação, relacionamento interpessoal, paciência e paixão.

Nas áreas de análise e ciência de dados há uma pluralidade de possibilidades de atuação e de cargos para esses profissionais, tais como: Analista de Dados, Engenheiro de Dados, Cientista de Dados, Arquiteto de Dados, Engenheiro de Big Data, Especialista em Machine Learning etc. Há diversos caminhos possíveis e diversas habilidades, que podem coincidir ou não. Para o contexto deste trabalho foi necessário estreitar esse amplo universo e, para isso, foram utilizados como referência *frameworks* que listaram e categorizaram os conhecimentos em dados utilizados para elaborar a lista a partir da qual análises foram aqui desenvolvidas^{[1],[5]}.

Outro ponto também levado em consideração foi o de que a era digital e o distanciamento social, provocados pela COVID-19, expandiram o uso de ferramentas on-line nos processos de recrutamento e seleção, uma vez que tal modalidade se tornou o principal meio de atração de candidatos para a maioria das empresas. Dentre os principais impulsionadores dessa adoção estão: a facilidade para os candidatos, velocidade do processo, escalabilidade e alcance. Além disso, poder reagir mais rapidamente aos concorrentes, atraindo candidatos à frente da concorrência, pode ser um outro diferencial competitivo para as organizações^[6].

Devido à relevância das ferramentas digitais no mercado de trabalho, este trabalho teve como objetivo analisar o processo descrito e fornecer informações concretas aos profissionais e empresas sobre as habilidades mais relevantes para as vagas do mercado de trabalho de dados, com maior enfoque nos

cargos de analista de dados (*data analyst*), engenheiro de dados (*data engineer*) e cientista de dados (*data scientist*), que atualmente são as posições mais procuradas pelas empresas brasileiras. Por conta de sua ampla utilização e posição de referência enquanto rede profissional, optou-se por utilizar o LinkedIn como fonte de dados para o desenvolvimento dessa análise.

A metodologia utilizada foi dividida em quatro etapas. No primeiro momento fez-se a extração de dados por meio do processo de *web scraping*. A seguir foi feita a análise descritiva dos dados coletados; nesta etapa foram selecionadas e categorizadas as principais habilidades buscadas na área, com intuito de verificar as características das vagas publicadas. No terceiro momento, foi medida a associação entre as habilidades em questão. Por fim, foram utilizadas técnicas de *text mining* para entender o contexto e extrair informações significativas dos campos textuais, determinando as palavras e frases mais utilizadas.

Tabela 1. Síntese dos métodos utilizados no trabalho

Método	Descrição
Web scraping	Extração automatizada de dados de websites para coletar informações sobre as vagas publicadas, incluindo campos textuais e outras informações.
Análise descritiva	Seleção e categorização das principais habilidades buscadas nas vagas, verificando características relevantes e padrões nos dados coletados.
Medida de associação	Avaliação das relações entre as habilidades identificadas, a fim de entender como elas estão conectadas ou frequentemente associadas.
Text mining	Análise dos campos textuais para identificar palavras e frases mais utilizadas, entendendo o contexto e extraindo informações significativas.

Fonte: Elaborado pelos autores (2024)

Web scraping

Web scraping é uma técnica utilizada para extrair dados da internet e armazená-los para posterior utilização ou análise^[7]. Ela tem sido amplamente utilizada para coleta de grandes volumes de dados, especialmente devido à grande quantidade de dados gerados on-line diariamente. Neste trabalho optou-se por utilizá-la em razão de ser uma técnica de fácil customização e automatização e, para implementá-la, utilizou-se a biblioteca Selenium e a linguagem Python.

A pesquisa realizada na seção de vagas da plataforma LinkedIn foi feita em seis buscas distintas, utilizando-se as palavras-chave "analista de dados", "*data analyst*", "engenheiro de dados", "*data engineer*", "cientista de dados" e "*data scientist*", e definindo-se o filtro de localização como Brasil. O intuito do uso desses critérios foi abranger o máximo de vagas possíveis da área de dados, mesmo aquelas publicadas em outro idioma, bastante comuns em algumas empresas multinacionais.

A página de anúncios de vagas da plataforma exibe mais resultados à medida que o usuário rola para baixo no navegador, no entanto, depois de seis repetições desse processo, o carregamento deixa de ser automático e, para acessar mais resultados, o usuário precisa clicar no botão 'ver mais vagas'. Com base nessa dinâmica e a fim de otimizar o processo de coleta de dados, foi criado um *loop* que simulava a interação do usuário com a plataforma, considerando o total de vagas e a quantidade de novas vagas mostradas por clique. Após a coleta de todas as vagas disponíveis, foi gerado um segundo *loop* utilizando a *tag* da lista lateral ('li') para capturar as informações mais relevantes de cada vaga — como título, link, empresa, localização e data. Para as demais informações, como descrição da vaga,

nível de experiência, tipo de emprego, função e setor, foi necessário clicar no título de cada vaga para obter o detalhamento em uma aba à direita da tela (Figura 1).

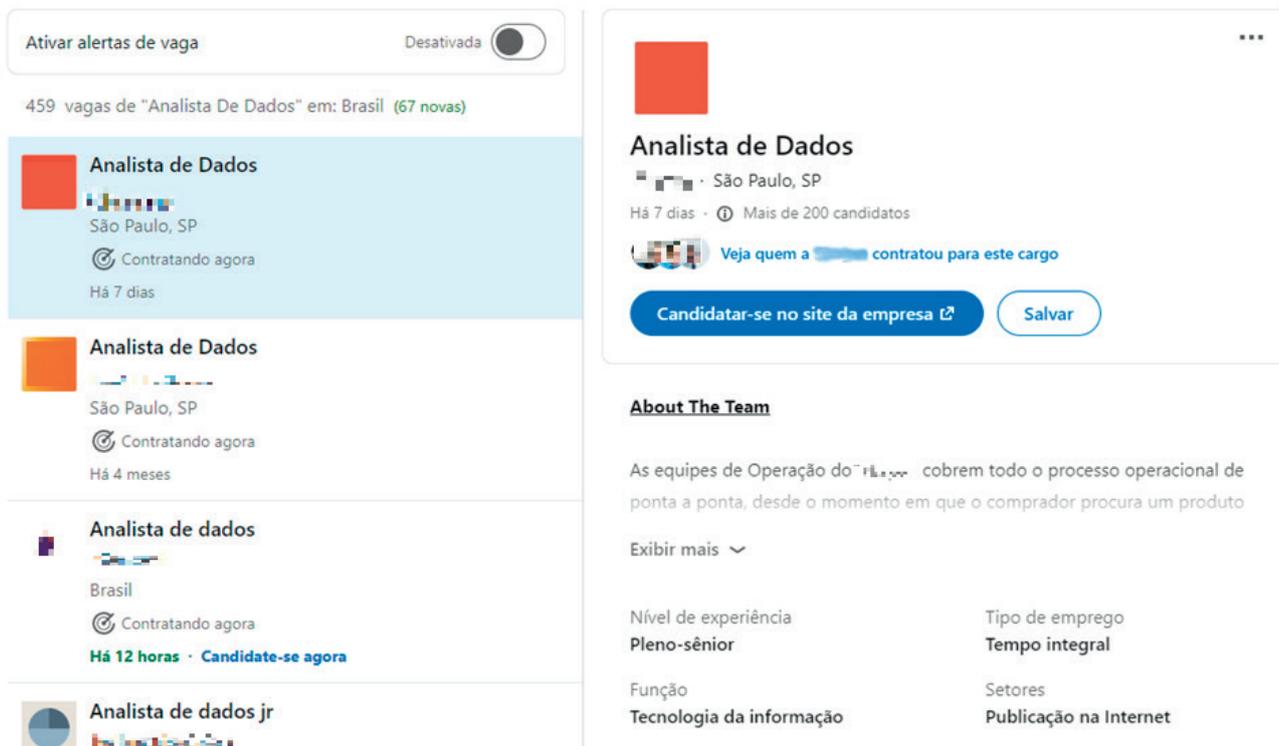


Figura 1. Página de vagas com busca por analista de dados do LinkedIn, 2022

Fonte: LinkedIn^[1]

Após a coleta, os dados foram estruturados em *dataframes*, combinados em um conjunto único e submetidos ao processo de limpeza e remoção de duplicatas. Ao final desse processo, foi obtido um conjunto de dados composto por 2498 publicações únicas.

Análise de dados

A análise descritiva é uma técnica utilizada para sintetizar e agrupar dados ou realizar comparações entre grupos de dados^[8]. Dada a natureza qualitativa dos dados coletados no LinkedIn, utilizou-se essa técnica por ser a mais adequada para compreensão das características nas distribuições, permitindo a identificação de agrupamentos numéricos como porcentagens, índices e médias.

A análise dos dados abrangeu a exploração de diversas variáveis, tais como localização, tipo, nível de senioridade, setor, empresa e o tipo de contrato. Além disso, como o objetivo também era identificar as particularidades de cada função dentro da área de dados, foi criada uma classificação nomeada "posição". A fim de identificar as particularidades de cada uma, assim como da área como um todo, foram realizadas diversas análises comparativas entre as posições, tanto separadamente quanto em conjunto.

Como ferramentas descritivas para representação dos dados foram utilizados gráficos e tabelas que apresentam as informações de forma clara e facilmente interpretável.

Associação entre as habilidades

Inicialmente, foi necessário delimitar o escopo da pesquisa. Para isso, e com base nas referências consultadas, foram selecionadas as principais competências técnicas e comportamentais requeridas

para profissionais da área de dados. Com a finalidade de facilitar a análise, essas habilidades foram agrupadas e categorizadas da seguinte forma:

- **Banco de Dados:** ETL, ELT, SQL, SQL Server, Oracle, Postgres, MySQL, NoSQL, Cassandra, MongoDB, Data Warehouse.
- **Dashboard e Relatórios:** Power BI, Tableau, Qlik, Business Objects, Excel, D3, Matplotlib, plotly.
- **Big Data:** Hadoop, Spark, Hive, Pig, HBase.
- **Programação:** Python, Scala, C/C++, C#, Java, .NET, HTML, PHP, Javascript.
- **Estatística:** R, SAS, SPSS, Stata, MATLAB, Regressão, Probabilidade.
- **Cloud:** Cloud, AWS, Azure, GCP.
- **Comportamentais:** Liderança, comunicação, relacionamento, apresentação, organização.

Para medir a associação entre as habilidades selecionadas, foi utilizado um algoritmo para identificar a presença de cada termo no campo de descrição das vagas publicadas. Em alguns casos, foram consideradas diferentes versões de cada termo (por exemplo: Power BI, PowerBI e PBI; Data Warehouse e DW), e também as traduções para vagas publicadas em inglês (por exemplo: comunicação e *communication*). O *dataframe* retornado pelo algoritmo foi composto por colunas binárias e analisado utilizando-se o "Phi Coefficient" — também conhecido como "Matthews Correlation Coefficient" (MCC) —, uma técnica amplamente utilizada em diferentes áreas para a medição da associação entre variáveis binárias^[9].

Text mining

A mineração de texto, ou text mining, é o processo de extrair informações relevantes de fontes de dados não estruturadas, como textos, para identificar palavras-chave e padrões. Essa técnica transforma textos em variáveis numéricas, permitindo sua análise com métodos estatísticos. No presente caso, utilizou-se a análise de n-grams, que detecta sequências de palavras recorrentes em um texto, para explorar de forma mais detalhada os campos textuais "título" e "descrição"^[9].

A partir da categorização de competências e aplicação das técnicas supracitadas, foram apresentadas as principais características das vagas no mercado de trabalho de dados, com foco principal em verificar quais eram as habilidades mais requisitadas e se havia diferenças entre as competências buscadas nos diferentes cargos de analistas, engenheiros e cientistas de dados. Ademais, por meio de técnicas de análise de texto, a análise permitiu identificar as relações entre as diferentes habilidades e as combinações mais frequentes de palavras. Os resultados obtidos poderão auxiliar estudantes e profissionais tanto para orientar suas decisões educacionais quanto fornecendo embasamento para uma busca mais assertiva de vagas de trabalho.

Análise descritiva

Por meio da análise descritiva dos dados buscou-se compreender e sintetizar suas principais características. A análise dos títulos das oportunidades apresentou uma grande pluralidade de termos. Para a criação do campo posição, foram necessários alguns tratamentos desses dados, tais como: remoção de caracteres especiais, conversão para maiúsculas e para o português, além de padronização dos termos de nível de senioridade. Dessa forma, verificou-se que 34,2% das oportunidades eram direcionadas a engenheiros de dados, 22,3% a analistas de dados e 12,9% a cientistas de dados. As

demais oportunidades foram classificadas na categoria "outro", que abrangeu engenheiros de software, engenheiros de *machine learning* e analistas de negócios (Figura 2).

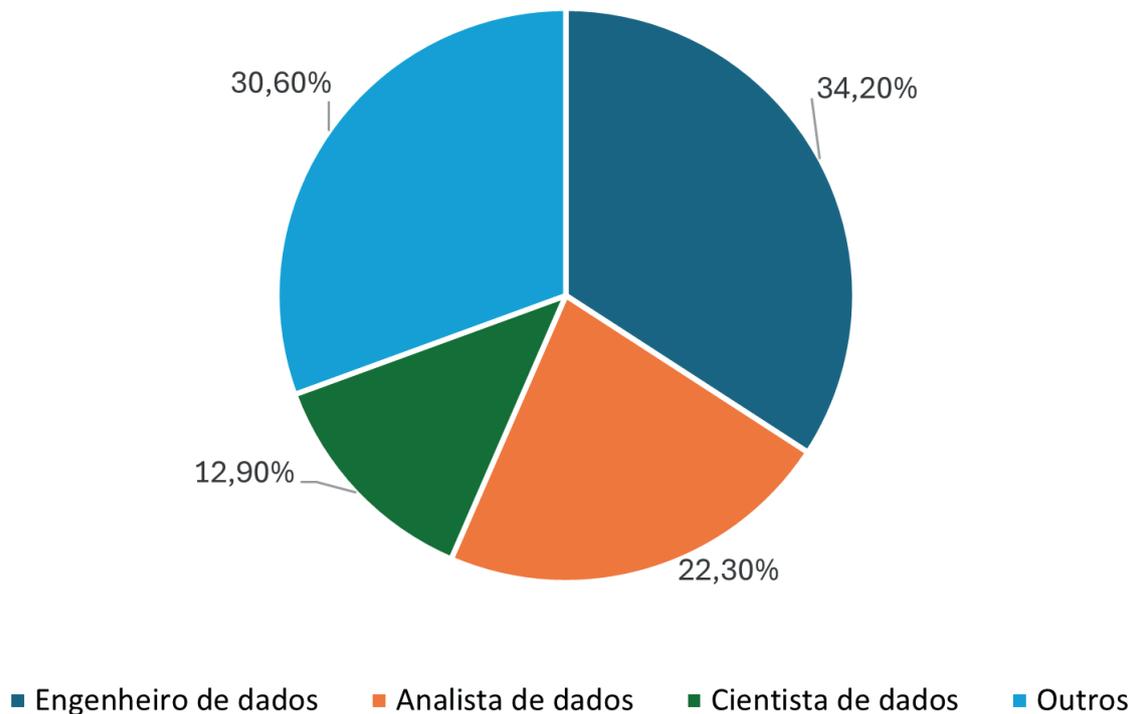


Figura 2. Distribuição das vagas por posição

Fonte: Resultados originais da pesquisa

Por ser uma informação que poderia ser feita em diferentes níveis (cidade, estado ou até mesmo país), também fez-se necessário um tratamento da coluna localização, separando-a nos níveis de cidade e estado. Observou-se que a maior disponibilidade das vagas de dados estava no estado de São Paulo, com 50,7% (Figura 3). No entanto, 209 vagas estavam classificadas no nível país e por isso não apareceram no gráfico de distribuição estadual, fator que se dá por muitas vagas da área serem remotas.

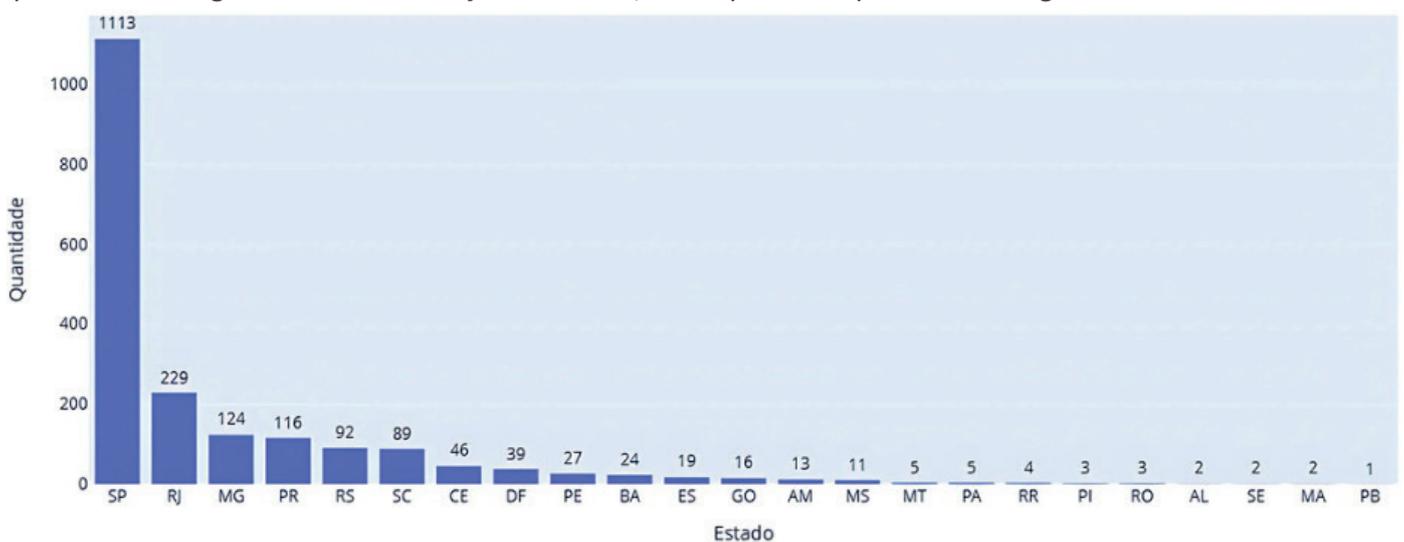


Figura 3. Quantidade de vagas por estado

Fonte: Resultados originais da pesquisa

No gráfico de distribuição de vagas por municípios, foram desconsideradas 447 vagas que **não** foram classificadas no nível de cidade. Observou-se que **São Paulo liderava entre** as 20 cidades com mais oportunidades na área de dados (Figura 4).

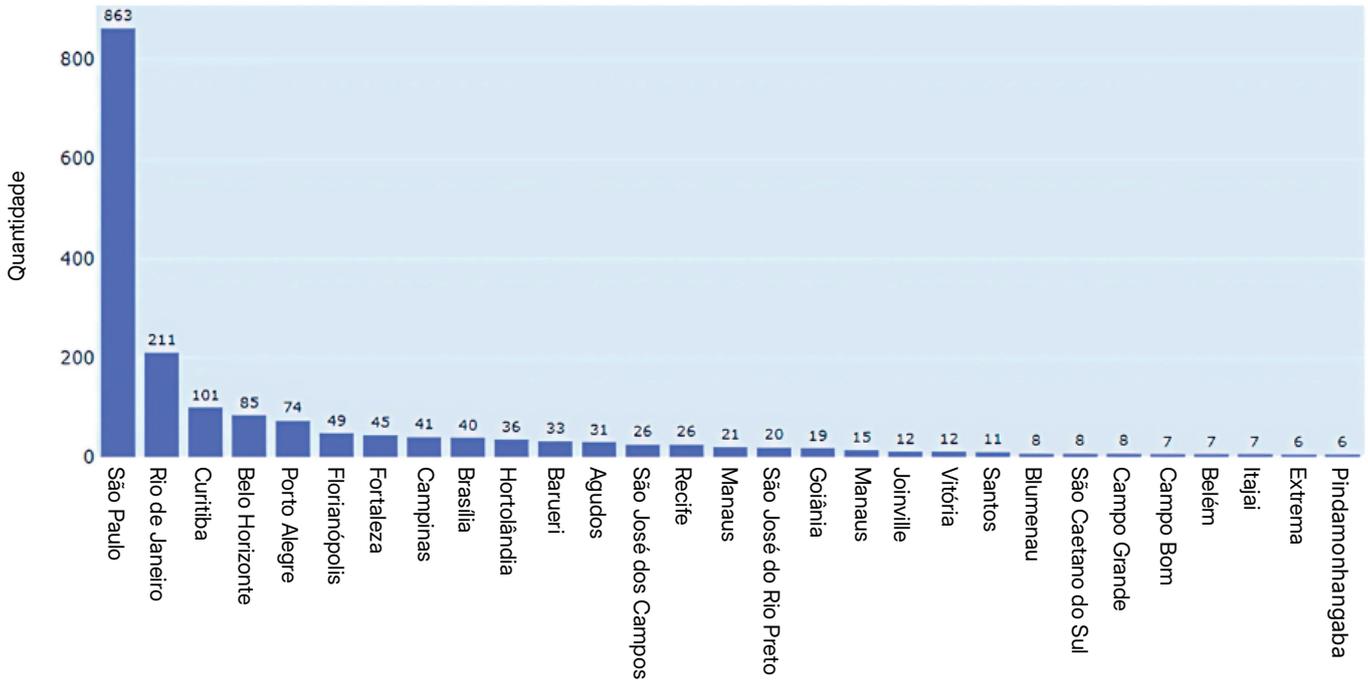


Figura 4. 20 cidades com maior disponibilidade de vagas na área de dados

Fonte: Resultados originais da pesquisa

Ao analisar o nível de senioridade das vagas, percebeu-se que a informação referente ao cargo muitas vezes não é preenchida ou é preenchida incorretamente nas vagas da plataforma. Isso pode ser observado, por exemplo, em vagas classificadas como de nível Assistente, mas com título ou descrição especificando um cargo de nível pleno ou sênior. A fim de minimizar esse problema, foi gerado o campo "position level" para verificar se o nível de senioridade estava especificado no título ou descrição da vaga e, quando não, se era feito o uso do campo disponibilizado pela própria ferramenta. Mesmo com o tratamento empregado, uma quantidade considerável de vagas ainda permaneceu com o nível de senioridade nulo ou sinalizado incorretamente.

Desconsiderando os dados nulos, analisou-se a distribuição dos níveis de senioridade para as vagas disponíveis (Figura 5). Observou-se que o nível júnior é mais significativamente representativo para a função de análise de dados (29,14%) que para as demais. Em contrapartida os níveis pleno-sênior e sênior concentravam-se nas vagas destinadas a engenheiros de dados (38,45%), enquanto as vagas para analistas e cientistas de dados ficaram numa posição intermediária (entre 15% e 17%) para os níveis sênior. Uma das hipóteses levantadas é que a análise de dados é uma porta de entrada para a área, enquanto a engenharia de dados é uma função que exige mais conhecimento técnico, levando as empresas a buscarem por profissionais mais seniores. É importante salientar que a categoria "outro" apresentou número significativo de vagas em níveis elevados por alocar cargos com uma senioridade alta, como *tech lead* ou engenheiros de *machine learning*.

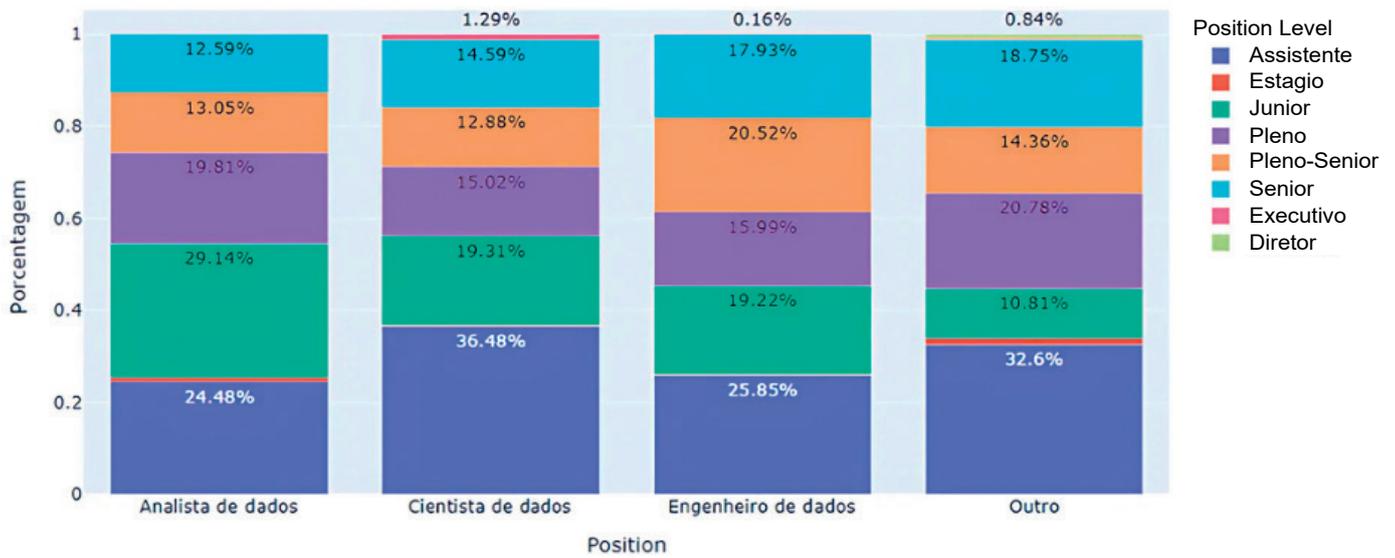


Figura 5. Distribuição dos níveis de senioridade por posição

Fonte: Resultados originais da pesquisa

Na análise de distribuição de vagas por empresas do mercado de dados, a IBM se destacou com o maior número de ofertas de empregos, com mais que o dobro de vagas publicadas pela segunda colocada (Figura 6).

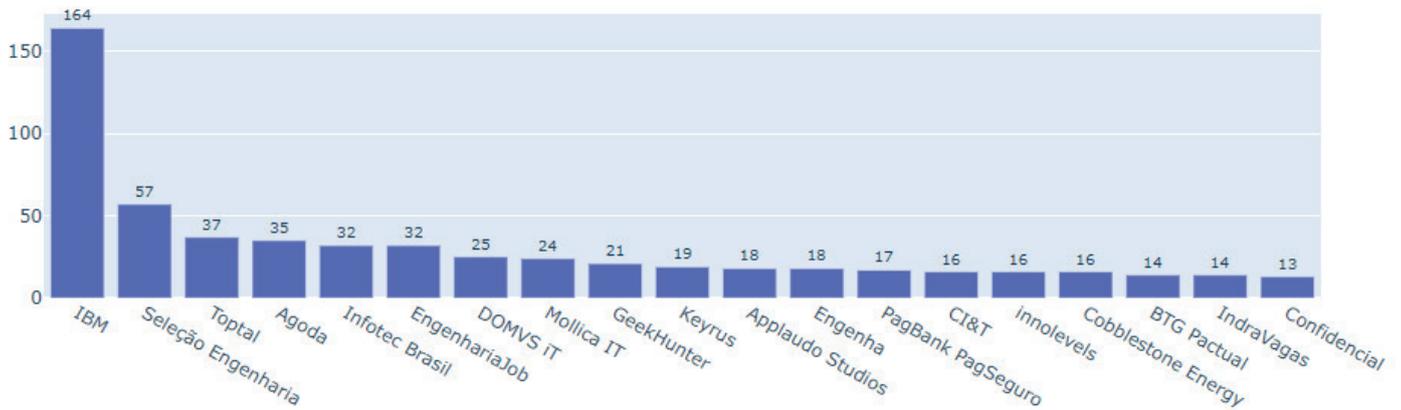


Figura 6. As 20 empresas com maior número de vagas

Fonte: Resultados originais da pesquisa

Em relação aos setores, os que apresentaram maior número de vagas abertas foram o setor de serviços e consultoria de TI e as empresas de recursos humanos (Figura 7). Esses dados evidenciam o crescimento que as consultorias de TI vêm apresentando nos últimos anos. Além disso, também destaca o papel das empresas de recursos humanos, que são contratadas por companhias de diversos segmentos para buscarem profissionais qualificados no mercado.

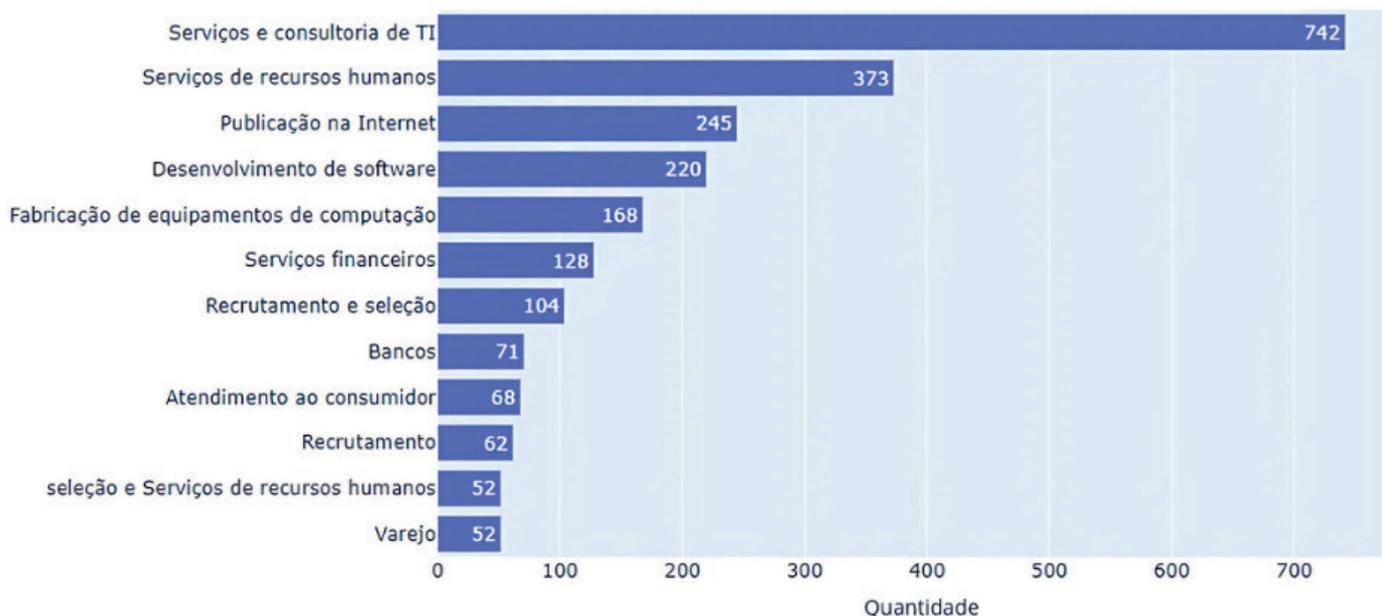


Figura 7. Os 12 setores com maior número de vagas

Fonte: Resultados originais da pesquisa

Quanto ao tipo de trabalho e função, estas não geraram informações relevantes para a análise. No entanto, ao analisar as categorias criadas, os resultados observados corresponderam ao esperado e as categorias de "Programação e Banco de Dados" foram as mais procuradas, com pelo menos uma das habilidades nelas listadas presentes em mais de 65% das vagas. As categorias "Cloud", "Dashboards e Relatórios", "Habilidades comportamentais" e "Big Data" também tiveram uma presença significativa, com valores próximos a 50% (Tabela 2).

Tabela 2. Número de vagas e percentual sobre o total por categoria

Categoria	Qtd.	%
Programação	1785	71.46
Banco de Dados	1669	66.81
Cloud	1259	50.40
Comportamental	1235	49.44
Dashboards e Relatórios	1198	47,96
Big Data	694	27.78
Estatística	579	23.18

Fonte: Resultados originais da pesquisa

Observando a distribuição de habilidades individualmente, vê-se que as cinco mais procuradas foram: SQL, Python, Cloud, Excel e AWS (Tabela 3). Foi possível notar que SQL e Python teve uma presença significativa para todas as funções, mas suas posições variaram de acordo com cada função. Percebeu-se que analista de dados foi a única na qual Python não apareceu em primeiro lugar, ficando atrás de SQL e Power BI; enquanto para engenheiro e cientista de dados, Python liderou e SQL apareceu na sequência. Outro ponto de destaque é que as habilidades Cloud, AWS, ETL, Spark, Scala, Java e Azure foram mais presentes para a função engenheiro de dados, Power BI para o analista de dados e o R para cientista e analista de dados. Curiosamente, comunicação e organização, duas

competências comportamentais, estiveram entre as 15 competências mais procuradas, sugerindo que as empresas buscam profissionais que, além do conhecimento técnico, consigam se comunicar bem e sejam organizados.

Tabela 3. Número de vagas por habilidades e posição

Categoria	Habilidade	Analista de Dados	Engenheiro de Dados	Cientista de Dados	Outro	Total
Banco de Dados	Sql	384	654	175	275	1488
Programação	Python	252	667	262	230	1411
Cloud	Cloud	75	561	108	192	936
Dashboard e Relatórios	Excel	251	214	88	249	802
Cloud	Aws	84	402	79	156	721
Comportamental	Comunicação	166	206	103	222	697
Banco de Dados	Etl	110	436	47	45	638
Big Data	Spark	38	460	73	55	626
Dashboard e Relatórios	Power Bi	293	121	54	114	582
Programação	Scala	50	300	64	135	549
Programação	.Net	53	208	84	202	547
Comportamental	Organização	106	212	60	158	536
Programação	Java	50	237	36	163	486
Estatística	R	133	62	141	83	419
Cloud	Azure	36	242	47	77	402

Fonte: Resultados originais da pesquisa

Em seguida, buscando-se compreender a ocorrência de mais competências específicas de cada cargo, foi calculada a representatividade de cada habilidade dentro de cada posição, e para isso, foram necessários pesos diferentes, pois a quantidade de vagas retornadas nas pesquisas foi diferente para cada posição.

Na posição de analista de dados, Power BI se destacou em relação às demais, com 61,95% contra 16,64%, 19,65% e 1,75% nas posições de engenheiro, cientista e outros, respectivamente (Tabela 4). Outras duas ferramentas de visualização de dados que também apareceram, foram Tableau e Qlik.

Tabela 4. Habilidades com maior delta para Analista de Dados em relação as demais

Habilidade	Analista_%	Engenheiro_%	Cientista_%	Outro_%
power bi	61.95	16.64	19.65	1.75
tableau	59.56	17.89	20.76	1.78
qlik	51.98	27.77	19.17	1.08
html	51.94	12.99	27.51	7.56
sql server	49.25	35.19	13.14	2.42

Fonte: Resultados originais da pesquisa

No requisito HBase 92,30% das vagas eram direcionadas a engenheiros de dados. Cassandra, ELT, Hive e Data Warehouse também fizeram parte das habilidades mais requisitadas para engenheiros de dados, pois, em geral, habilidades relacionadas a banco de dados e big data têm maior peso em funções de engenheiro (Tabela 5).

Tabela 5. Habilidades com maior delta para engenheiro de dados em relação às demais

Habilidade	Analista_%	Engenheiro_%	Cientista_%	Outro_%
hbase	2.76	92.30	4.74	1.197419
cassandra	11.73	77.80	8.08	1.064810
elt	19.56	73.63	5.32	1.149562
hive	7.74	71.46	20.43	1.335699
data warehouse	15.37	69.16	14.62	1.267120

Fonte: Resultados originais da pesquisa

No caso dos cientistas de dados, observou-se que as habilidades relacionadas a estatística e pacotes de visualização se destacaram, com a maioria delas tendo um delta de aproximadamente 80% em relação às demais. Destaca-se, ainda, que Stata apresentou 95,94% das oportunidades voltadas para essa posição (Tabela 6).

Tabela 6. Habilidades com maior delta para cientista de dados em relação às demais

Habilidade	Analista_%	Engenheiro_%	Cientista_%	Outro_%
stata	0.00	0.00	95.94	4.06
plotly	6.77	0.00	93.23	0.00
regressão	7.15	1.16	90.72	0.98
probabilidade	6.48	2.53	89.29	1.70
matplotlib	5.76	9.36	84.25	0.63

Fonte: Resultados originais da pesquisa

Por fim, foram exploradas as habilidades mais buscadas em cada categoria e observou-se que, para banco de dados, o SQL se destacou como a mais popular (Figura 8). Do total de 2.498 vagas de emprego analisadas, 1.488 delas apresentavam preferência por um candidato com conhecimento ou experiência com SQL (Figura 8).

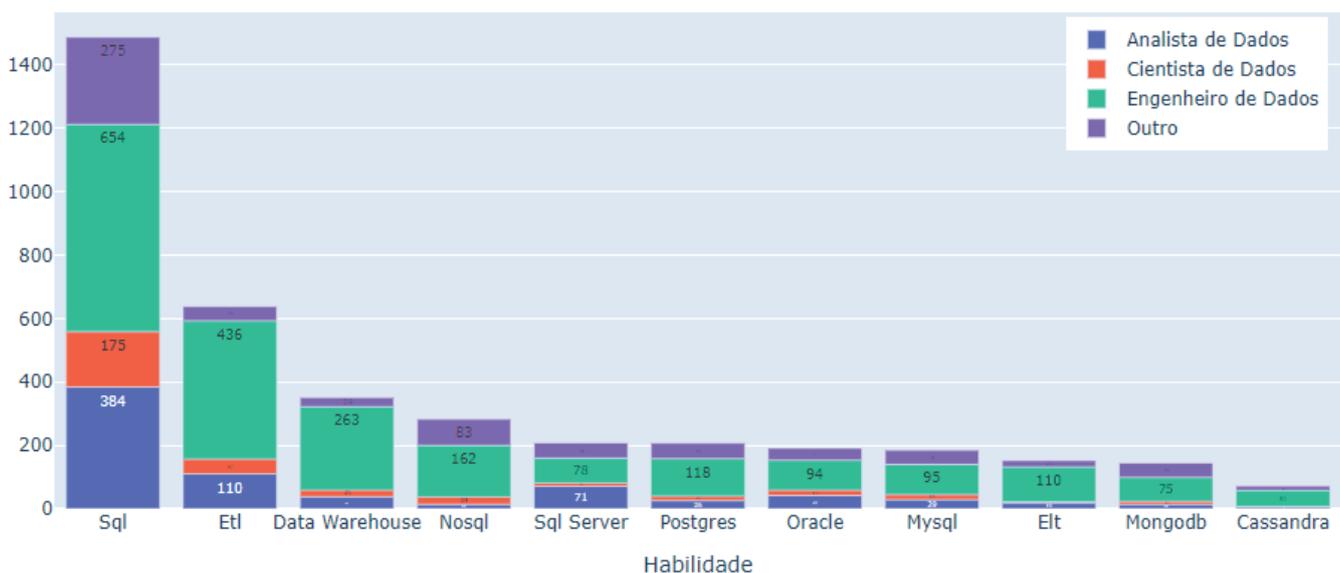


Figura 8. Habilidades mais procuradas da categoria banco de dados

Fonte: Resultados originais da pesquisa

Na categoria de programação, também foi possível observar a predominância de uma linguagem em relação às demais. Em um total de 1.411 vagas de programação, 56,49% citavam o Python, um número mais de duas vezes superior ao da Scala (549 vagas), segunda colocada (Figura 9). Caso o R, alocado na categoria estatística, também fosse considerado para programação, ele ocuparia a quinta posição (419 vagas), atrás de Java (486 vagas).

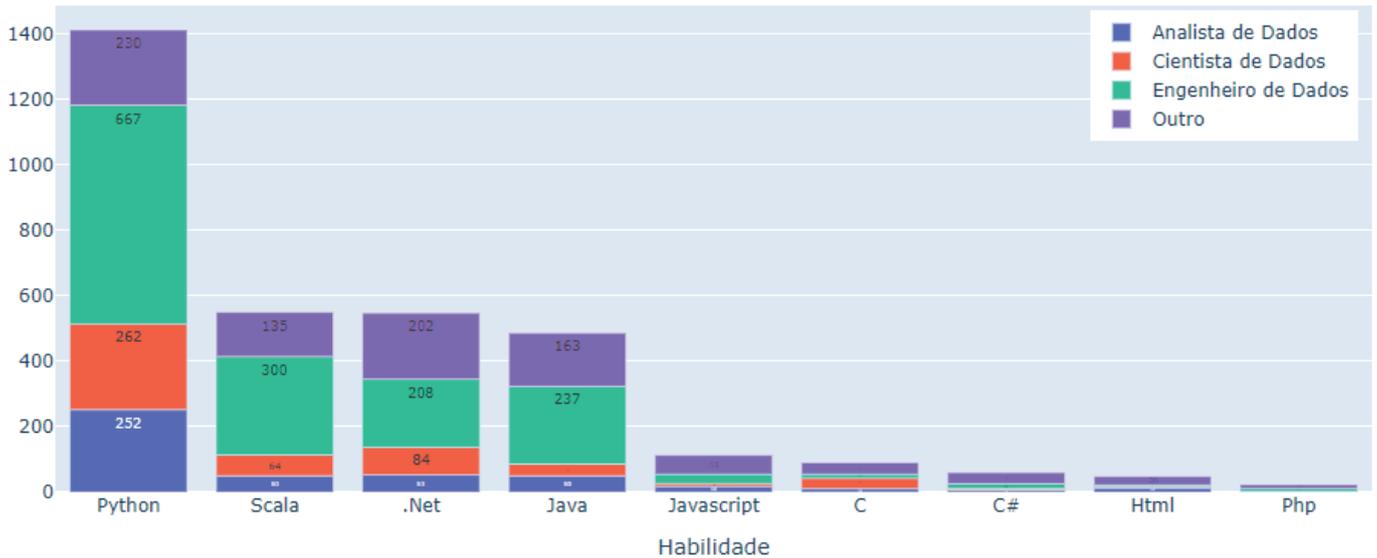


Figura 9. Habilidades mais procuradas da categoria programação

Fonte: Resultados originais da pesquisa

Para a categoria cloud, optou-se por analisar também a própria palavra, para verificar se a habilidade também seria citada sem especificação de plataforma. O termo cloud apareceu 936 vezes, seguida pelas plataformas AWS (721 vezes), Azure (402 vezes) e GCP (320), refletindo o *market share* do mercado brasileiro de nuvem (Figura 10).

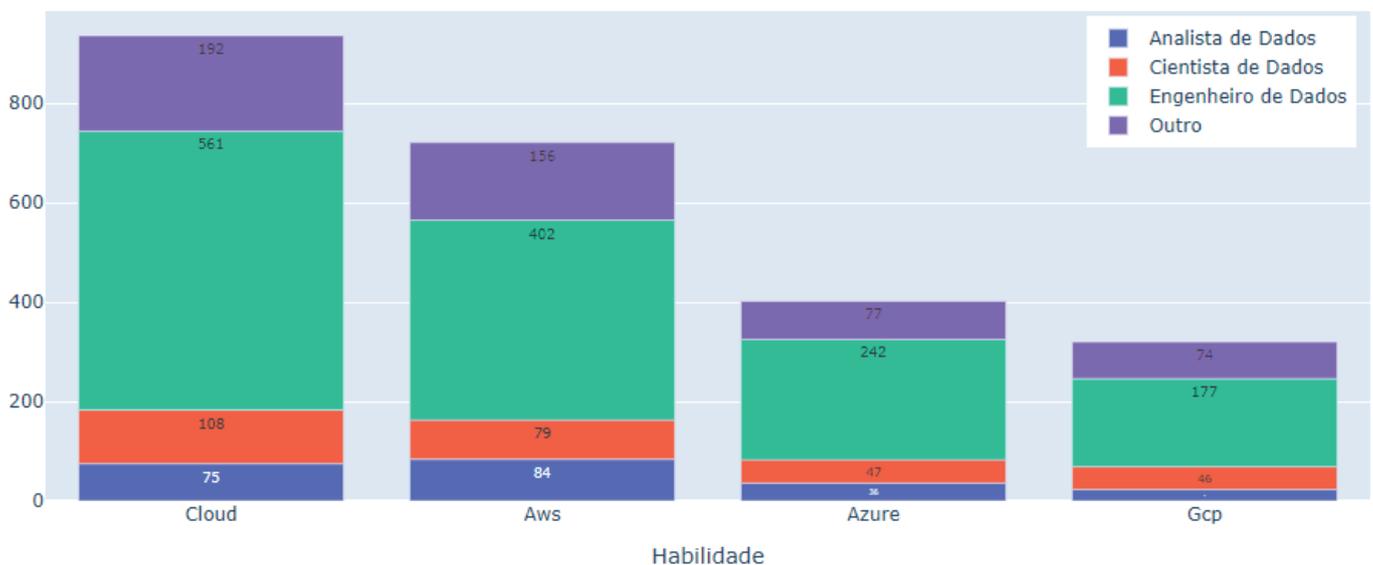


Figura 10. Habilidades mais procuradas da categoria cloud

Fonte: Resultados originais da pesquisa

Na categoria de dashboards e relatórios os mais populares foram Excel, Power BI e Tableau, e, como já esperado, foram mais numerosas para as posições de analista de dados. Elas também tiveram alto volume de aparição para as posições "outro", provavelmente relacionado às posições de analista de negócios, função para a qual exigem conhecimento em ferramentas de visualização e análise de dados, principalmente o Excel, visto como a porta de entrada para essa área (Figura 11).

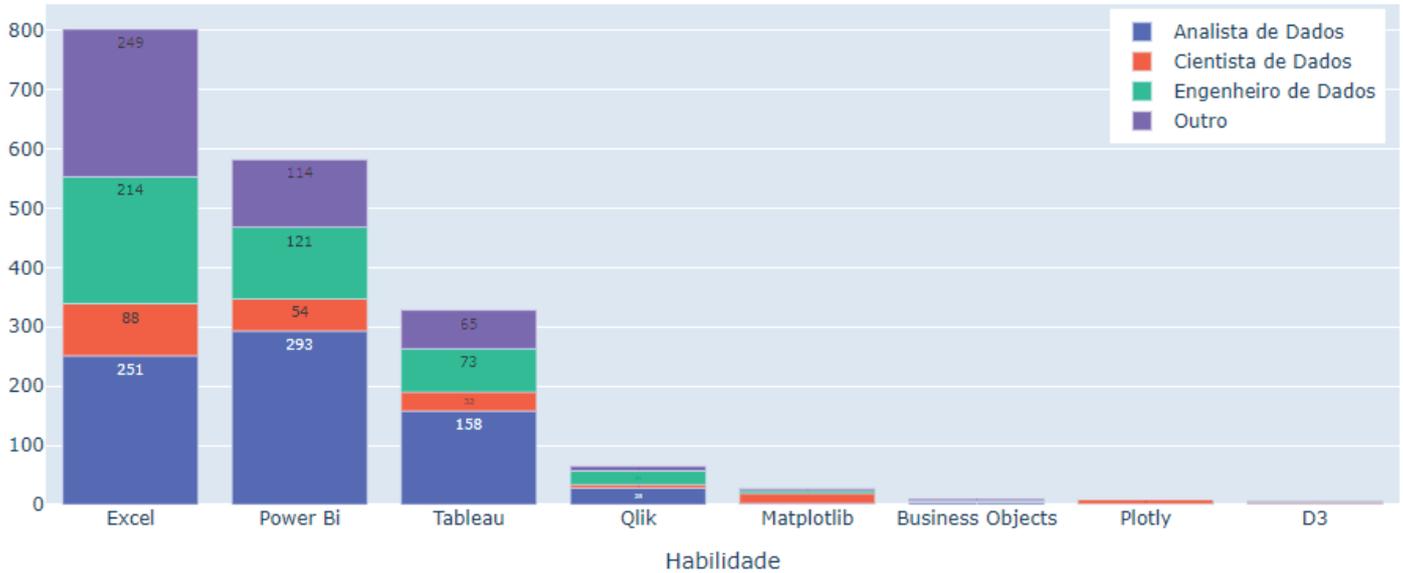


Figura 11. Habilidades mais procuradas da categoria dashboards e relatórios

Fonte: Resultados originais da pesquisa

Na categoria de *big data*, observou-se um destaque de Spark, Hadoop e Hive como as habilidades mais procuradas. É importante observar que estas são, de longe, as mais requisitadas para engenheiros de dados (Figura 12).

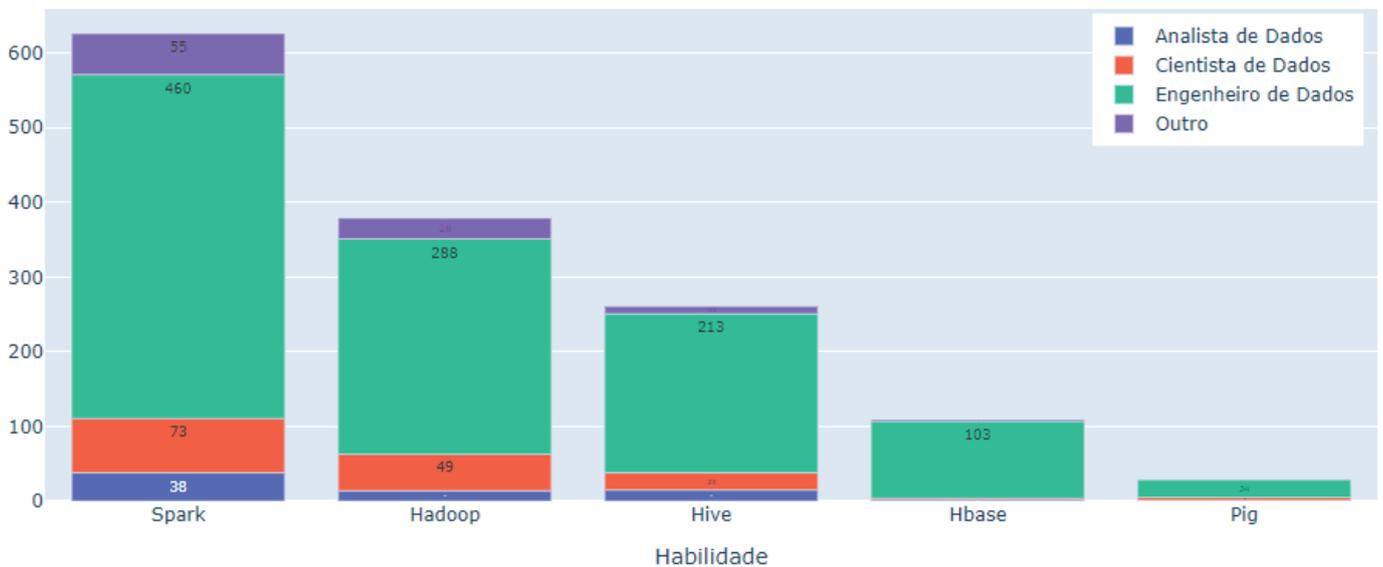


Figura 12. Habilidades mais procuradas da categoria big data

Fonte: Resultados originais da pesquisa

Na categoria estatística, também analisou-se a estatística enquanto habilidade. Ela esteve entre as quatro habilidades mais populares para a categoria, com 5,5 %, junto com R, que liderou com 16,77%.

Pandas (6%) e SAS (5%) ficaram em segundo e quarto lugar, respectivamente (Figura 13). Nesta categoria foi possível perceber que os níveis de valores para cientistas de dados foram maiores ou iguais em relação às demais — o que não aconteceu em outras categorias, corroborando a importância da estatística para esses cargos.

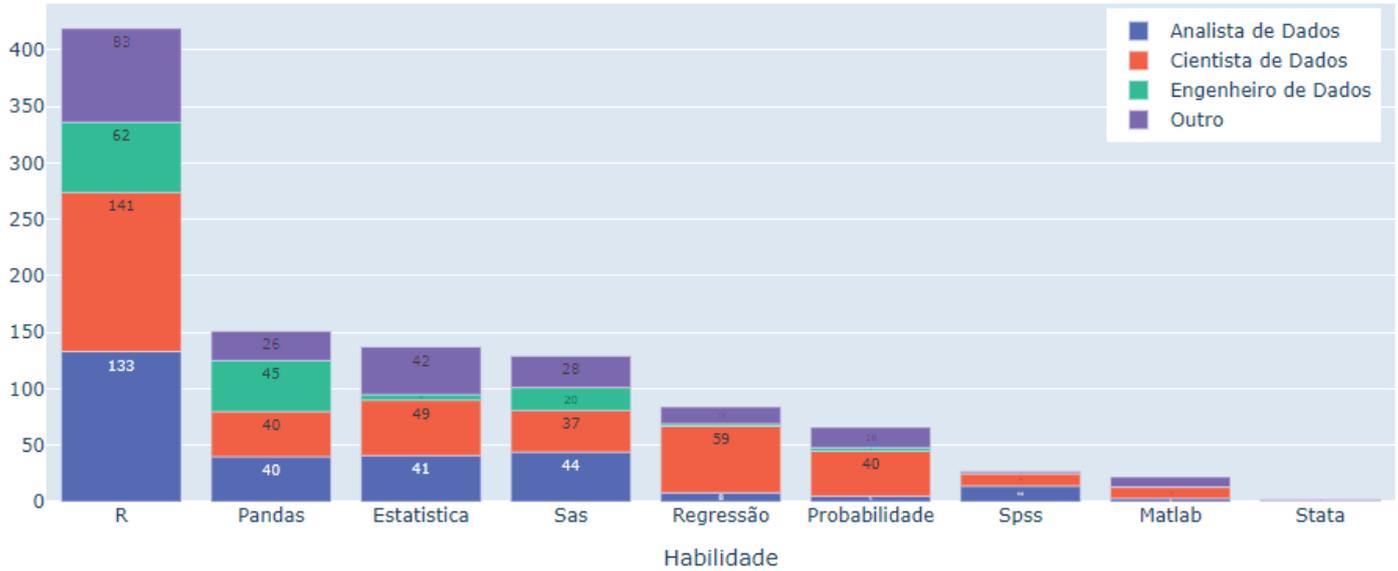


Figura 13. Habilidades mais procuradas da categoria estatística
 Fonte: Resultados originais da pesquisa

Por fim, entre as habilidades comportamentais, comunicação e organização se destacaram como as mais predominantes (Figura 14).

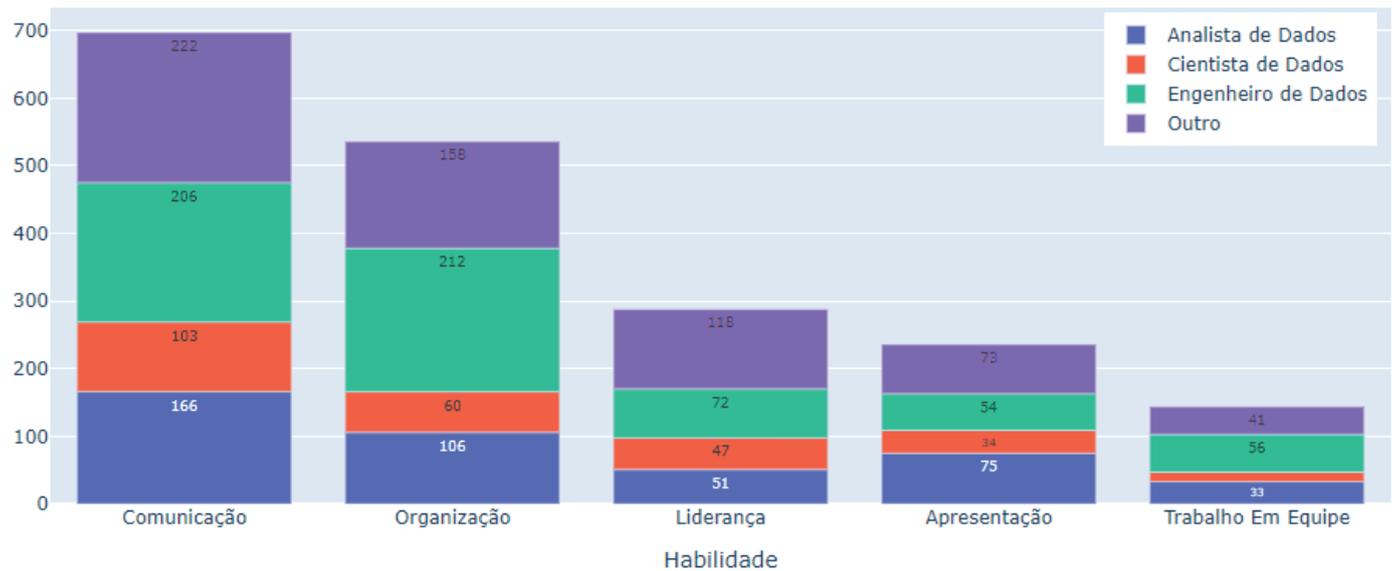


Figura 14. Habilidades mais procuradas da categoria habilidades comportamentais
 Fonte: Resultados originais da pesquisa

Associação entre as habilidades

Para analisar a associação entre as habilidades, considerou-se as vinte principais, e foi possível identificar correlações positivas fortes entre algumas delas. A maior dessas associações foi observada

entre Hadoop e Spark, tecnologias diferentes, porém consideradas estruturais para a categoria *big data*. Na mesma linha, Spark e Python também tiveram fortes relações, demonstrando que o Python é amplamente requerido em conjunto com o Spark. Power BI e Tableau tiveram associação de destaque e, por serem ferramentas análogas, gerou-se a hipótese de que as empresas são flexíveis em relação a qual ferramenta de BI o candidato precisa ter experiência, levando à conclusão que elas são solicitadas separadamente não em conjunto. A associação entre Python e SQL reafirmou a importância de dominar ambas as habilidades no atual mercado de dados. As altas correlações entre cloud e plataformas cloud, como o GCP, não diferiram do esperado e, portanto, não foram consideradas como fontes insight. As demais associações se mostraram moderadas ou fracas (Figura 15).

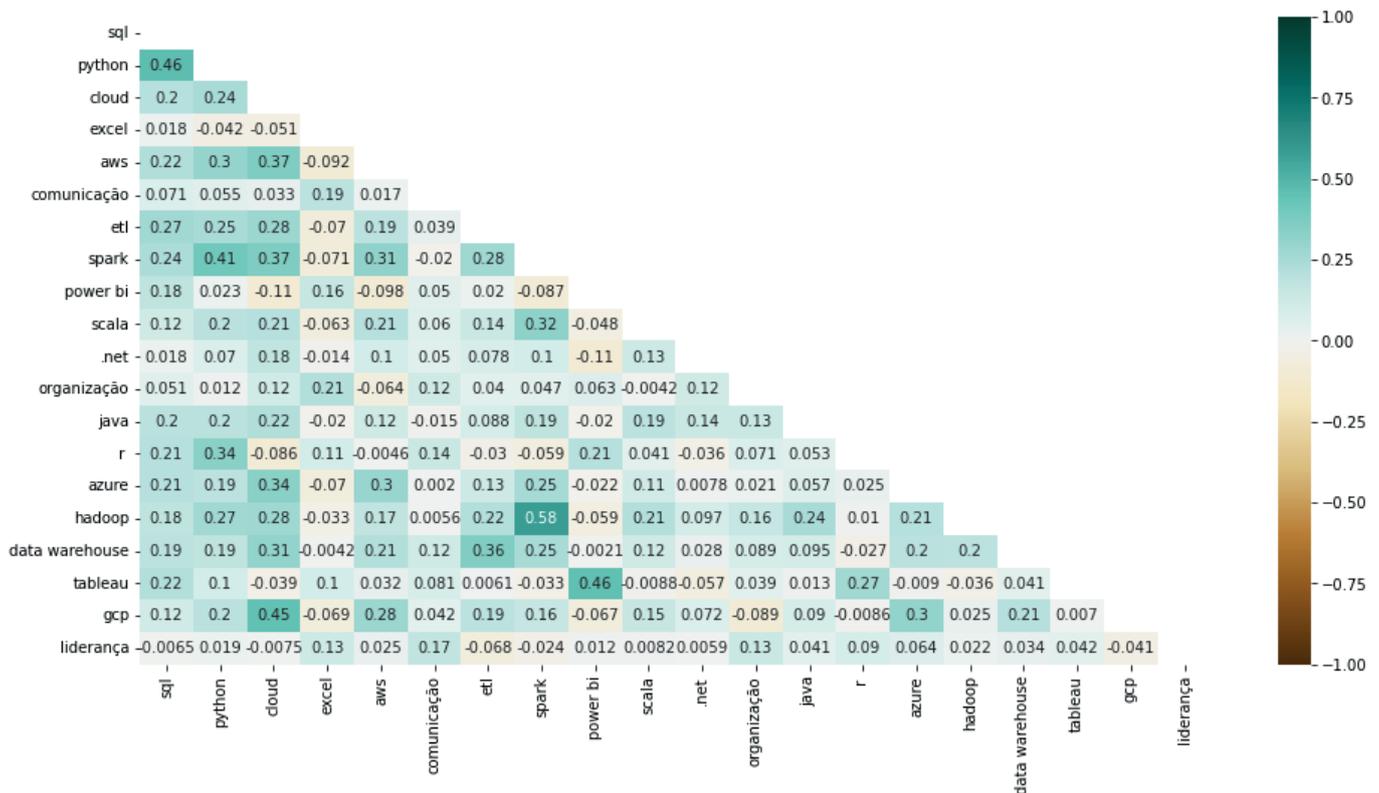


Figura 15. Heatmap da associação entre as habilidades mais procuradas

Fonte: Resultados originais da pesquisa

Text mining

Para não se limitar a analisar somente as habilidades pré-selecionadas, também foi realizado o processo de mineração de texto para buscar informações desconhecidas que pudessem ser úteis dentro das publicações. Para isso foi utilizada a biblioteca Tidytext para buscar os n-gramas, ou seja, a ocorrência de sequências de palavras nos campos textuais título e descrição. Os tratamentos realizados foram: colocar todas as palavras em letra minúscula; remover pontuações, caracteres especiais e "stopwords", em inglês e português — palavras comuns aos idiomas, mas que não agregavam à análise, como, por exemplo artigos e preposições (o, a, em, no).

Os unigramas dos títulos das vagas, que são as principais palavras individuais encontradas, foram representados visualmente em uma nuvem de palavras (Figura 16) de modo a dar maior destaque às palavras que aparecem com mais frequência. No geral, as palavras "dados", "engenheiro", "analista",

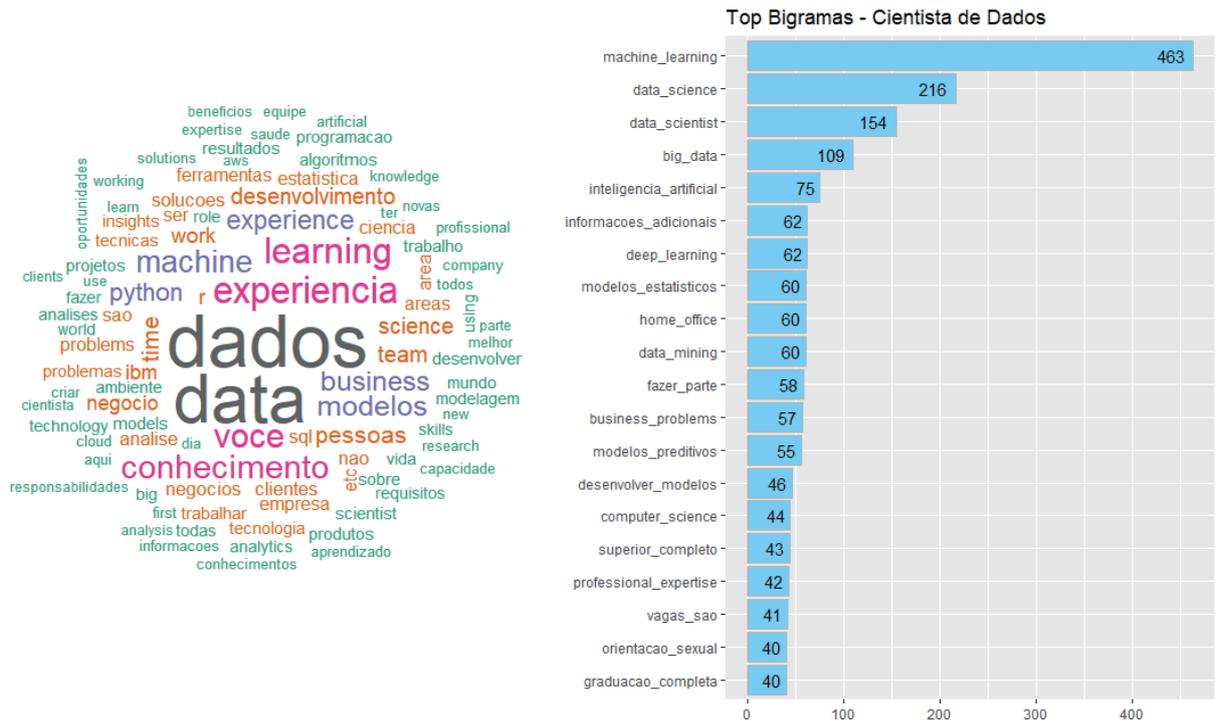


Figura 21. Principais unigramas e bigramas nas vagas de Cientista de Dados
 Fonte: Resultados originais da pesquisa

Nas vagas destinadas a engenheiros de dados, os termos em destaque foram: "big data", "data lake", "data warehouse", "data pipeline" e termos relacionadas a "cloud" (Figura 22).

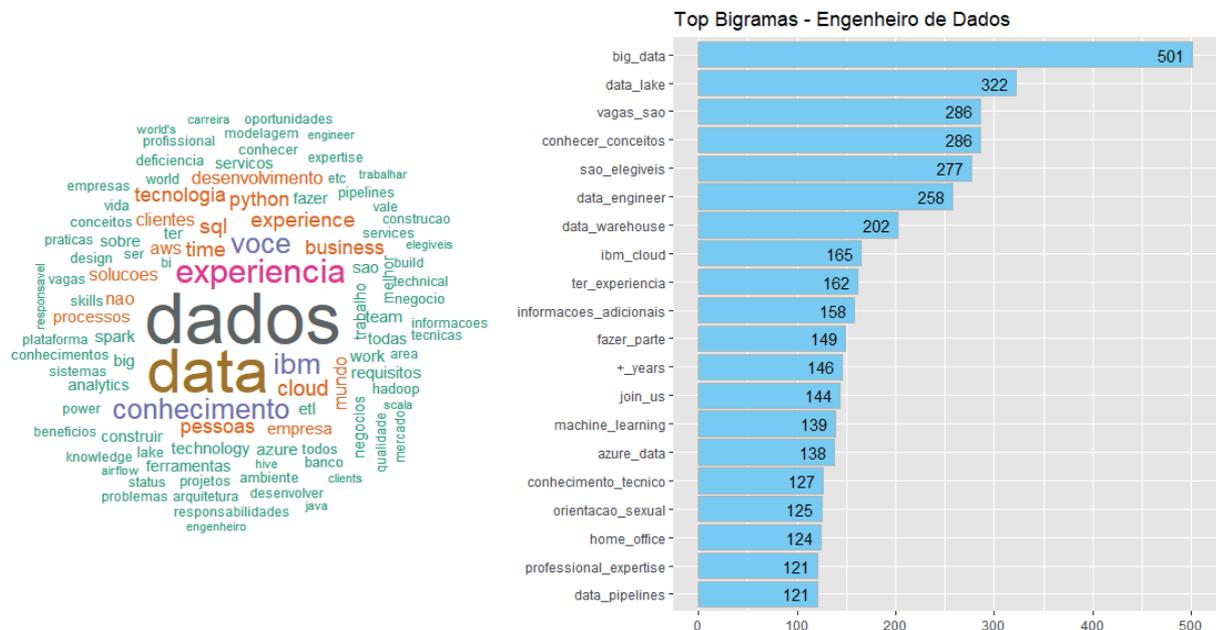


Figura 22. Principais unigramas e bigramas nas vagas de Engenheiro de Dados
 Fonte: Resultados originais da pesquisa

O gerenciamento pessoal nas organizações é uma tarefa complexa, influenciada por diversos fatores internos, como cultura, estratégia e tecnologia, uma vez que as empresas se configuram como sistemas interdependentes, nos quais as pessoas desempenham papel central. Utilizando dados de vagas no

LinkedIn, foi possível identificar as principais habilidades técnicas específicas, além de competências comportamentais que são cruciais em todas as funções da área de TI. Habilidades como Python e SQL destacaram-se como essenciais para qualquer carreira na área, informação que pode auxiliar profissionais e estudantes a se prepararem para o mercado de trabalho.

No entanto, observou-se que ainda há uma carência de definição clara sobre cargos e atribuições, fator importante para definir funções, ajudar os colaboradores a compreenderem seu papel dentro da empresa e direcionar candidatos na aquisição de competências demandadas. O principal objetivo deste estudo foi oferecer *insights* iniciais sobre o tema e, a partir disso, propor a possibilidade de pesquisas futuras explorarem tendências de mercado e perfis profissionais na área de TI, por ser, este, um cenário em constante evolução.

REFERÊNCIAS

- [1] LinkedIn. 2020. Emerging Jobs Report Brazil. Disponível em: https://business.linkedin.com/content/dam/me/business/en-us/talent-solutions/emerging-jobs-report/Emerging_Jobs_Report_Brazil.pdf. Acesso em: 10 mar. 2022.
- [2] Fayyad, U.; Hamutcu, H. 2020. Toward Foundations for Data Science and Analytics: A Knowledge Framework for Professional Standards. *Harvard Data Science Review*, 2(2).
- [3] Persaud, A. 2021. Key competencies for big data analytics professions: a multimethod study. *Information Technology & People* 34(1). 178-203. <https://doi.org/10.1108/ITP-06-2019-0290>.
- [4] Dubey, R.; Gunasekaran, A. 2015. Education and training for successful career in big data and business analytics. *Industrial and Commercial Training* 47(4), 174-181. <https://doi.org/10.1108/ICT-08-2014-0059>.
- [5] Verma, A.; Yurov, K.; Lane, P.; Yurova, Y. 2019. An investigation of skill requirements for business and data analytics positions: A content analysis of job advertisements. *Journal of Education for Business* 94. 1-8. <https://doi.org/10.1080/08832323.2018.1520685>.
- [6] Pellaes, A. 2021. A adoção de ferramentas de atração e recrutamento online pelas organizações no Brasil: um estudo sobre motivações e impactos. *Brazilian Journal of Business* 3(4), 3192-3207. <https://doi.org/10.34140/bjbv3n4-027>
- [7] Zhao, B. 2017. Web Scraping. In: Schintler, L., McNeely, C. (eds) *Encyclopedia of Big Data*. Springer, Cham. https://doi.org/10.1007/978-3-319-32001-4_483-1
- [8] Reis, E.A.; Reis, I.A. 2002. Análise Descritiva de Dados. Relatório Técnico do Departamento de Estatística da UFMG.
- [9] Chicco, D.; Jurman, G. 2020. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*. 21(6). <https://doi.org/10.1186/s12864-019-6413-7>