

UNIVERSIDADE DO VALE DO RIO DOS SINOS - UNISINOS
UNIDADE ACADÊMICA DE GRADUAÇÃO
CURSO DE ADMINISTRAÇÃO

NAYARA SILVA DAS NEVES

CIÊNCIA DE DADOS:
Análise do Perfil do Cidadão Cientista de Dados

Porto Alegre
2021

NAYARA SILVA DAS NEVES

CIÊNCIA DE DADOS:

Análise do Perfil do Cidadão Cientista de Dados

Trabalho de Conclusão de Curso apresentado como requisito parcial para obtenção do título de Bacharel em Administração, pelo Curso de Administração da Universidade do Vale do Rio dos Sinos - UNISINOS

Orientadora: Profa. Dra. Josiane Brietzke Porto

Porto Alegre

2021

À Nayara de ontem, que se dedicou, insistiu, acreditou e não desistiu. Mesmo nos momentos mais difíceis, seguiu confiando na sua própria capacidade para que eu pudesse realizar esta conquista.

AGRADECIMENTOS

Agradeço à Deus, que nesse período tão difícil me concedeu saúde para que eu fosse capaz de trabalhar, estudar e manter um teto sobre a minha cabeça, em momento em que muitos não tiveram a mesma benção.

Ao meu amado companheiro, Evandro Santos, que esteve ao meu lado, em cada passo desta caminhada e em muitos momentos assumiu total reponsabilidade sobre a nossa vida, para que eu pudesse me dedicar totalmente à conclusão do meu curso.

Aos meus incríveis pais, Cleide Ferreira e Rosivaldo Neves, que me ensinaram que não devemos medir esforços para que possamos alcançar nossos objetivos e sempre foram meus exemplos de luta, perseverança e trabalho duro.

À minha avó, Maria de Nazaré Ferreira, a matriarca da nossa família. A mulher mais corajosa e destemida que eu conheço, sinônimo de força, luta e fé, sempre sendo a rocha da nossa família. A pessoa que, em nossos momentos de fraqueza, sempre nos aconselhou a 'vestir a nossa roupa de mulher', o que para ela significa, seja forte.

À minha orientadora, Josiane Porto, pela dedicação e paciência na condução deste trabalho.

“Não é o mais forte que sobrevive, nem o mais inteligente, mas o que melhor se adapta às mudanças” (MEGGINSON, 1963).

RESUMO

Compreender o perfil do Cidadão Cientista de Dados (CCD) é tão desafiador quanto emergente. Deve-se partir do entendimento do que é um CCD e quais são as competências necessárias ao seu perfil. No momento, a literatura relacionada aponta como alguém que preenche a lacuna entre os *insights* do Cientista de Dados e a atividade de análise a ser realizada pelo usuário do negócio. Alguém com experiência e conhecimento específico na área do negócio e uma habilidade técnica considerável na área de Ciência de Dados. Mas, na prática, quais são as competências (conhecimentos, habilidades, atitudes) do perfil de um CCD? A experiência e o conhecimento no negócio são elementos que parecem ser fundamentais para esse papel, que surgiu nas organizações com o intuito de preencher uma necessidade atual do mercado. Este trabalho contribui para uma melhor compreensão, a partir da identificação do perfil atual, tendo como base a percepção de 46 profissionais com experiência em projetos nessa área por meio de uma *survey* eletrônica. A partir da percepção desses profissionais da amostra obteve-se um conjunto das principais competências, consideradas presentes e importantes para o desenvolvimento do papel de CCD.

Palavras-chave: Cidadão cientista de dados. Perfil. Competências. Ciência de Dados.

LISTA DE FIGURAS

Figura 1 - Diagrama de Venn da Ciência de Dados.	18
Figura 2 - Diagrama de Venn da Ciência de Dados, modificado.....	25

LISTA DE GRÁFICOS

Gráfico 1 - Gênero dos respondentes	47
Gráfico 2 - Faixa etária dos respondentes.....	48
Gráfico 3 - Tempo de Experiência	48
Gráfico 4 - Segmentos de Mercado.....	49
Gráfico 5 - Formação Acadêmica.....	50
Gráfico 6 - Posição Hierárquica.....	51
Gráfico 7 - Nível de Formação Acadêmica.....	51

LISTA DE QUADROS

Quadro 1 - Competências do Papel do Cidadão Cientista de Dados.....	40
Quadro 2 - Caracterização dos especialistas participante do pré-teste	44
Quadro 3 - Grau de Importância das Conhecimentos	52
Quadro 4 - Grau de Importância das Habilidades	53
Quadro 5 - Grau de Importância das Habilidades	54
Quadro 6 - Grau de Domínio Prévio.....	55

LISTA DE SIGLAS

CCD	Cidadão Cientista de Dados
CD	Cientista de Dados
BI	<i>Business Intelligence</i>
DSR	<i>Design Science Research</i>
CHA	Conhecimentos, Habilidades e Atitudes
MO	Moda
ME	Média
DP	Desvio Padrão

SUMÁRIO

1 INTRODUÇÃO	12
1.2 MOTIVAÇÃO E JUSTIFICATIVA.....	13
1.3 DELIMITAÇÃO DO TEMA E QUESTÃO DE PESQUISA	14
1.4 OBJETIVO GERAL.....	14
1.4.1 Objetivos Específicos	15
1.5 ORGANIZAÇÃO DO TRABALHO	15
2 FUNDAMENTAÇÃO TEORICA	16
2.1 A CIÊNCIA DE DADOS.....	16
2.2 O Cidadão Cientista de Dados	19
2.2.1 O Cientista de Dados no Brasil e no Cenário Internacional.	21
2.2.2 O Significado de Cidadão Cientista de Dados	23
2.2.3 Desenvolvimento do Papel de Cidadão Cientista de Dados	27
2.3 TRABALHOS ANTERIORES RELACIONADOS	30
2.4 COMPETÊNCIAS DO CIDADÃO CIENTISTA DE DADOS	35
3 METODOLOGIA	41
3.1 DELINEAMENTO DA PESQUISA.....	41
3.2 POPULAÇÃO E AMOSTRA	42
3.3 TÉCNICA PARA COLETA DE DADOS	43
3.3.1 Pré-testes do Instrumento de Pesquisa	44
3.3.2 Aplicação do Instrumento de Pesquisa	45
3.4 ANÁLISE DE DADOS.....	46
4 ANÁLISE DOS RESULTADOS	46
4.1 CARACTERIZAÇÃO DO PERFIL DOS RESPONDENTES	47
4.2 ANÁLISE DAS COMPETÊNCIAS DO PAPEL CIDADÃO CIENTISTA DE DADOS	52
4.3 DISCUSSÃO DOS RESULTADOS	57
5 CONCLUSÃO	60

5.1 CONSIDERAÇÕES FINAIS	60
5.2 CONTRIBUIÇÕES DA PESQUISA	62
5.3 LIMITAÇÕES DA PESQUISA	63
5.4 SUGESTÕES DE PESQUISAS FUTURAS.....	63
REFERÊNCIAS.....	65
APÊNDICE A – COMPETÊNCIAS DO CIDADÃO CIENTISTAS DE DADOS	70

1 INTRODUÇÃO

Toda pessoa com um dispositivo e conexão à internet é um potencial gerador de dados. Pode não ser perceptível aos usuários, mas a cada pesquisa feita em um navegador, cada palavra digitada em redes sociais, cada transação financeira realizada em meios de pagamentos dados associados é gerada e captada por organizações. Por meio da aplicação de conhecimentos, de técnicas e de ferramentas especializadas a tais conjuntos de dados pode-se tanto explicar quanto inferir algo sobre uma dada pessoa ou comportamento e tendências de um modo geral.

Nos últimos anos, as empresas têm reconhecido o valor desses dados e como eles podem se tornar, estrategicamente, um ativo organizacional, através do conhecimento aplicado à informação. As áreas de negócio vêm utilizando esses dados de forma prescritiva e preditiva e para tal necessidade, contratam os Cientistas de Dados. Este profissional de alto desempenho combina aprofundados conhecimentos e habilidades em Estatística, Ciência da Computação, Matemática, entre outros para extrair conhecimento de um grande volume de dados, em projetos na área de ciência de dados, normalmente, constituídos por recursos humanos multidisciplinares.

Considerada a profissão mais requisitada do século 21, a necessidade por Cientistas de Dados cresce mais rápido que a capacidade de formar novos profissionais para preencher essa demanda de mercado atual, surgindo assim o conceito e o papel de Cidadão Cientista de Dados (CCD). (HBR, 2012).

Originalmente, esse conceito foi cunhado por Gartner (2019), como um papel a ser desempenhado por um profissional com habilidades estatísticas e analíticas não tão aprofundadas e avançadas quanto às do Cientista de Dados profissional, mas com a vantagem de ter conhecimentos e larga experiência aplicada na área de atuação específica, no domínio do negócio.

1.2 MOTIVAÇÃO E JUSTIFICATIVA

Em princípio, o papel do CCD veio para sanar uma necessidade relevante do mercado, por meio de profissionais que já atuavam na área do negócio, capacitando-os com conhecimentos analíticos e estatísticos, não tão profundos quanto dos profissionais Cientistas de Dados, mas aprofundados o suficiente para manusear *softwares* especializados e fazer análises preditivas e descritivas preliminares, que proporcionem *insights* para a sua área de atuação.

Considerando que a necessidade da profissão surge no mercado, geralmente, antes da regulamentação de órgão complementares e da formalização da profissão ser efetuada de modo abrangente nas instituições de ensino, na área de Ciência de Dados, isso não está sendo diferente, sendo que nos últimos anos, o mercado de trabalho vem necessitando cada vez mais desses profissionais.

No entanto, tal formação acadêmica e continuada requerem tempo e interessados para tal propósito, por isso é necessário, mais uma vez, uma adaptação do mercado, abrindo-se a oportunidades transitórias (ou não) para o papel de CCD. Futuramente, esse recurso humano poderá se tornar um Cientista de Dados profissional, se assim optar por estudar mais e voltar a sua carreira nessa direção ao longo do tempo, aproveitando-se das experiências e das práticas vivenciadas em oportunidades de projetos nessa área, em que se já encontra envolvido no mercado de trabalho. Ou, então, seguirá atuando por meio desse papel ao lado de profissionais Cientista de Dados numa equipe multidisciplinar, suportando-os com as informações e o valioso ponto de vista do negócio.

Observa-se que o CCD está em procura alta nas empresas e segundo Gartner (2019) estima-se que mais de 40% das tarefas de Ciência de Dados serão automatizadas, resultando em uma maior produtividade e em uso mais amplo de dados e de análises por CCD (tradução livre).

Para atender essa demanda do mercado, um caminho alternativo de solução é que as organizações possam utilizar as habilidades atuais de seus profissionais de

negócio, proporcionando a eles, o conhecimento teórico e prático da área de Ciência de Dados, que sejam suficientes para extrair análises avançadas, minimizando a falta de um profissional Cientista de Dados, no cotidiano do negócio. Assim, as empresas passam a necessitar de um profissional que tenha o conhecimento técnico e do negócio, os *power users*, usuários que poderão realizar tarefas analíticas simples e moderadamente sofisticadas, que anteriormente exigiam mais experiência (CARLIE, 2018).

1.3 DELIMITAÇÃO DO TEMA E QUESTÃO DE PESQUISA

A falta de profissionais especializados de Ciência de Dados no mercado em face da alta demanda atual por esse profissional, levou as organizações a capacitarem rapidamente os seus próprios profissionais com experiência no negócio desenvolvendo as competências técnicas mínimas necessárias. Por outro lado, não é todo profissional da empresa que poderá se tornar um CCD, pois para Carlie (2018) esses *power users* são aqueles com habilidades estatísticas e em análise de dados, além do conhecimento atual no negócio. Então, nesse contexto surge a questão norteadora presente da pesquisa: Para desempenhar o papel de CCD, quais são as competências presentes e importantes para o perfil, na percepção desses profissionais?

1.4 OBJETIVO GERAL

O objetivo geral desse trabalho é analisar as competências necessárias do perfil do CCD para o desempenho de atividades desse papel, segundo a percepção dos profissionais atuantes nessa área no cenário nacional.

1.4.1 Objetivos Específicos

Para atingir esse objetivo geral, os seguintes objetivos específicos foram definidos: (i) identificar quais as competências necessárias para o desempenho de atividades atribuídas ao papel de CCD; (ii) realizar uma pesquisa *survey*, com profissionais que desempenham o papel de CCD; (iii) analisar os resultados da amostra da pesquisa, de acordo com a percepção dos profissionais que desempenham o papel em comparação às informações levantadas na bibliografia.

1.5 ORGANIZAÇÃO DO TRABALHO

Este trabalho está organizado em cinco capítulos. No Capítulo 1 é apresentada a introdução, com a motivação e justificativa para o desenvolvimento deste estudo, seus objetivos, delimitações do tema e como ele está organizado.

O capítulo 2 apresenta a fundamentação teórica dos temas relacionados a este trabalho. Conceitua a Ciência de Dados e o Cidadão Cientista de dados, apresenta o Cientista de Dados em cenário nacional, fala sobre o desenvolvimento do papel do CCD, os trabalhos relacionados, além das competências do CCD.

No Capítulo 3 é apresentada a metodologia de pesquisa utilizada, destacando o método de pesquisa *survey* e os demais aspectos metodológicos.

No Capítulo 4 são analisados e discutidos os resultados da pesquisa, de acordo com a percepção dos respondentes. E por fim, no capítulo 5 são apresentadas considerações finais, contribuições da pesquisa, limitações do estudo e sugestões para pesquisa futura.

2 FUNDAMENTAÇÃO TEORICA

Neste capítulo são apresentados os conceitos de dados, informação, conhecimento e sabedoria, além do contexto histórico e o surgimento da área de Ciência de Dados. Por fim apresenta-se um panorama sobre o papel CCD.

2.1 A CIÊNCIA DE DADOS

A Ciência de Dados é uma área de estudos que tem como objetivo extrair *insights* para auxiliar na tomada de decisões em uma organização, a partir de dados brutos. Historicamente, o termo surgiu em 1960 e era usado como sinônimo de Ciência da Computação. Depois, nos anos noventa, com o surgimento do *Data Warehouse* começou-se a difundir o uso de informação de forma mais intensiva, assim como a preocupação de usar (FAROL, 2017). Dessa forma, a informação para fins analíticos, começou há quase trinta anos.

Antes do uso da informação, a preocupação estava voltada para a informatização. Após a popularização da informatização, os dados passaram a estar disponíveis em maior abundância, no entanto, se tratava de dados separados em diversos sistemas, não possibilitando uma visão ampla. Assim, o *Data Warehouse* surgiu como a missão de reunir os fragmentos de dados em um único repositório.

O *data warehouse* consolida e padroniza as informações oriundas de diferentes bancos de dados operacionais, de modo que elas possam ser usadas por toda a empresa para análise gerencial e tomada de decisões". (LAUDON et al, 2007, p.149)

Já nos anos dois mil, as empresas identificaram a necessidade de dinâmica maior das informações para fins analíticos. Passou-se a ter a compreensão de que a produção de informação para fins analíticos, voltada para as áreas do negócio, precisava ser mais ágil (OLIVEIRA, 2009). Então, surgiu o conceito de *Business*

Intelligence (BI) e com ele, ferramentas que precisavam de pessoas técnicas para manipulação, porém com capacidade de desenvolver relatórios para uso pela área do negócio (NATIVA, 2020).

A função do *Business Intelligence* é usar ferramentas que possam coletar, analisar e extrair dados transformando em conhecimento, que será utilizado para auxiliar a tomada de decisão das empresas, com o objetivo de gerar vantagens competitivas. (OLIVEIRA, 2009, p.14).

Com a necessidade de se analisar e extrair informações úteis desse grande volume de dados, armazenados pelas organizações surgiu o conceito de Ciência de Dados, que pode ser considerado como uma versão mais recente do BI (NATIVA, 2020). Mas, apesar das semelhanças, esses dois conceitos possuem funções e abordagens diferentes. Enquanto o BI utiliza uma análise de dados descritiva ou retrospectiva, a Ciência de Dados vai além e é mais abrangente, ao fazer uso das análises preditivas e descritivas (NATIVA, 2020)

Entende-se Ciência de Dados como uma área de conhecimento interdisciplinar, que faz uso de métodos, processos, ferramentas, algoritmos e sistemas científicos para extrair conhecimento ou *insights* de grande volume de dados. O profissional dessa área combina uma série de competências, incluindo Estatística, Ciência da Computação, Matemática, Segurança da Informação e Administração, para analisar dados coletados de diversos formatos e fontes como- *web*, *smartphones*, clientes, sensores entre outras.

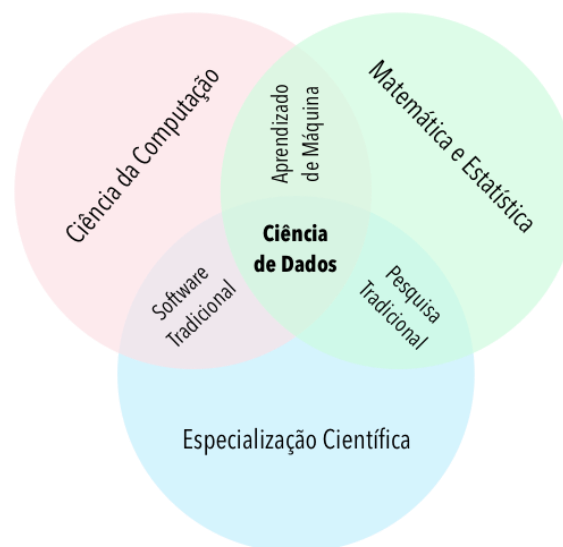
Para o efetivo sucesso das soluções de Ciência de Dados, o conhecimento do domínio do problema deve ser disponível e amplamente utilizado no processo de tomada de decisão. Neste sentido, as soluções de Ciência de Dados são voltadas para a formulação de hipóteses e a aquisição de informação aderente como insumo no processo decisório (CARMO, RAUNTENBERG, 2019, p.59)

Essa área incorpora elementos variados e se baseia em técnicas e teorias oriundas de campos básicos em engenharia e em ciências básicas, sendo assim

intimamente ligada com muitas disciplinas tradicionais bem estabelecidas, porém viabilizando uma nova área altamente interdisciplinar. Dessa forma, na origem e fundamentalmente, a Ciência de Dados está associada a este espírito de aplicação interdisciplinar, (PORTO e ZIVIANI, 2014), conforme ilustrado na Figura 1.

Na Figura 1 é apresentado um Diagrama de Venn, que representa as áreas do conhecimento da área de Ciência de Dados e conseqüentemente, da profissão de Cientista de Dados. Tanto a profissão quanto a área do conhecimento propriamente dita são conceitos considerados novos, ainda que a forma de manusear dados já seja utilizada há bastante tempo.

Figura 1 - Diagrama de Venn da Ciência de Dados.



Fonte: BAUDISCH, 2016

Historicamente, Loh (2014, p.17) exemplifica que:

A identificação de padrões é parte da nossa vida. A descoberta de padrões iniciou há milhares de anos atrás. Nossos antepassados conseguiram prever as variações do tempo, as estações, os ciclos das plantações, as fases lunar e eclipses, e até mesmo o surgimento de reis. E hoje em dia não é diferente.

Apesar de os dados estarem mudando e transformando a forma de fazer, gerir e prospectar negócios, o profissional com capacidade para lidar com estes dados e dar valor de mercado ainda é escasso, embora, se observe um leve conhecimento apropriado nessa área.

Como visto, a Ciência de Dados tem como principal objetivo a extração de conhecimento através de dados brutos e de grandes volumes de dados armazenados. tendo o dado e seus conceitos norteadores como principais ferramentas para o desenvolvimento do trabalho associado. Corresponde a uma habilidade de usar os dados para saber mais do que eles mostram, um modelo que permite representar, explicar e completar um cenário (LOH, 2014). Esta ciência está localizada entre explicar o que aconteceu e prever o que vem pela frente, ela não mais responde “qual foi o site mais acessado no passado?” Ela explica o porquê determinado site foi o mais acessado e prevê qual será aquele mais visitado no futuro, por essa razão agrega valor e inteligência aos negócios com base em dados estruturados e não estruturados.

[...]Servem para direcionar nossas decisões e ações. Por isto, usamos muitas vezes o termo "modelo de predição", porque usando modelos podemos "prever" o futuro (ou tentar, pelo menos). Modelos são utilizados para previsão do tempo, para previsão de colheitas, de níveis de vendas[...] (LOH, 2014, p.18-19)

Cabe ressaltar que há diferenças entre prever e análise preditiva, a previsão é baseada em modelos tradicionais de Estatística e Matemática, a análise preditiva utiliza de algoritmos de *machine learning*, conceitos de inteligência artificial e outras formas avançadas de análise, como *deep learning* (BANKER, 2018).

2.2 O Cidadão Cientista de Dados

De acordo com Varian (2009), a habilidade de obter, entender, processar, extrair valor, visualizar e se comunicar com dados vai ser extremamente importante nas próximas décadas, não só no nível profissional, mas também na educação, a partir

da educação infantil ao ensino superior, pois basicamente dados são gerados a todo momento e existem em todos os cantos, gratuitamente.

Em alguns setores, a previsão pode ter um funcionamento satisfatório. Por exemplo, os aplicativos de cadeia de suprimentos existentes fazem um bom trabalho ao fornecer boas análises em uma ampla variedade de indicadores-chave de desempenho, desde o desempenho no prazo até o custo de separação de caixas em um depósito (BANKER, 2018). Mas, para Banker (2018), têm necessidades de informação que as organizações desejam ou deveriam querer saber, sendo que esses aplicativos e ferramentas não são capazes de produzir.

Com o intuito de elevar a maturidade em análise de seus dados e assim, melhorar a tomada de decisões, tornando-a mais eficaz e eficiente os processos produtivos e de negócios, as organizações têm estabelecidos novas ferramentas para a elaboração dessas análises (FLATH; STEIN, 2017).

Como já dito antes, as previsões ajudam as empresas no seu planejamento e no seu dia a dia. Uma empresa que trabalhe com estoques que consiga prever quanto vai vender nos próximos dias, pode produzir ou comprar somente a quantidade que irá vender. Estoque parado é prejuízo porque a empresa precisa pagar infraestrutura para armazenar (local, pessoas, climatização etc.) e se não vender o produto pode deteriorar (perder prazo de validade, estragar por condições climáticas adversas etc.). (LOH, p. 140)

Com o avanço da capacidade tecnológica, atividades que antes não estavam sendo monitoradas, passaram a ser importantes fontes de dados, ajudando na tomada de decisões e na geração de conhecimento (RIBEIRO, 2014). Para sanar a necessidade de análises preditivas são necessárias soluções que ajudem a desenvolver melhor os negócios. Mas, essas soluções requerem um profissional específico, o Cientista de Dados.

Entretanto, tal profissional não tem uma formação rápida e devido à alta e crescente demanda de mercado, ele se tornou caro e escasso. Portanto, a necessidade atual de um profissional Cidadão Cientista de Dados (CCD). Com um recurso humano designado ao papel de CCD, a ideia é que as organizações possam

aproveitar as competências internas para aproveitar o conhecimento básico da área de Ciência de Dados. em análises preliminares ou um pouco mais avançadas das atuais.

Desse modo, organizações provedoras de soluções de análise de dados também estão trabalhando para desenvolver ferramentas intuitivas e de fácil utilização para apoiar o trabalho de CCD. Elas estão criando ferramentas de visualização de dados, onde fluxos de dados podem ser adicionados, conectados e analisados por meio de métodos e recursos de soltar e arrastar elementos. Essas ferramentas são acionadas por um mecanismo analítico subjacente, portanto, é mais fácil para o CCD criar mais momentos de *insights* usando dados, algoritmos e modelos, com uma curva menor de aprendizado (BANKER, 2018).

2.2.1 O Cientista de Dados no Brasil e no Cenário Internacional.

Primeiramente, para a realização deste trabalho foram analisadas diversas publicações, em português e em inglês, nos últimos 10 anos, no *Google Academic*. Quando se pesquisa pelo profissional Cientista de Dados no contexto internacional, é possível perceber que há uma diferença em relação ao que é considerado um Cientista de Dados no Brasil e o que é considerado como um Cientista de Dados em países onde a profissão não é tão recente.

No cenário internacional, é possível observar que são profissionais que cursaram a graduação em *Data Science*, que têm formação nas áreas do conhecimento relacionadas as funções desenvolvidas, como Engenharia, Ciência da Computação e Estatística e fizeram curso em nível de mestrado ou doutorado, na área de Ciência de Dados.

Após a criação do primeiro curso de Mestrado em Ciências em Análise do mundo criado na *North Carolina State University* em 2007 e o primeiro *Master of Analytics by Research* e *PhD in Analytics* lançado na *Universidade of Technology Sydney* em 2011, mais de 150 universidades e as instituições já

criaram ou estão planejando cursos em ciência de dados, grandes dados e análises (tradução livre) (CAO, 2017).

No Reino Unido, em um levantamento sobre os profissionais atuantes no mercado como cientista de dados e foi identificado que, “a maioria dos cientistas de dados que participaram da pesquisa online para este estudo tem mestrado e cerca de um terço tem outra qualificação relevante de pós-graduação (tradução livre)” (BROW, SWAN, 2008, p. 15).

Para os autores, além dos conhecimentos acadêmicos avançados nos campos supracitados, o estudo também ressaltou a dificuldade de realização de seus trabalhos atuais, sem uma compreensão detalhada da área de assunto com a qual os cientistas de dados estão trabalhando. Esta visão é baseada no requisito de compreender as características particulares das informações associadas as áreas específicas de estudo, mas também no fato de que ter um conhecimento profundo de uma área de assunto facilita a comunicação entre o Cientista de Dados e os outros membros do grupo de trabalho (BROW, SWAN, 2008).

No Brasil, por sua vez, ainda não há um consenso do que seria um Cientista de Dados, nem a clareza da necessidade de uma formação específica para se tornar um profissional. No geral, são necessários conhecimentos nas áreas específicas como Matemática, Estatística e base de programação, em vezes, um curso de curta duração com certificação de Ciência de Dados, não havendo necessariamente a exigência por um curso de ensino superior nessa área para atuar profissionalmente. Se forem consideradas as características da profissão Cientista de Dados nos Estados Unidos, por exemplo, em nível nacional, o profissional que foi descrito anteriormente caracteriza-se como um CCD.

Em outros países, assim como no Brasil, ainda não há regulamentação da profissão, porém, a literatura é unânime em enfatizar que os Cientistas de Dados devem apresentar domínio estatístico e computacional para a programação e uso de sistemas capazes de processar grandes volumes de dados (CHATIFELD et al., 2014;

GRANVILLE, 2014) e ter capacidade de explorar a inteligibilidade em dados que a princípio são desestruturados e sem sentido (VAN DER AALST, 2014)

Com a crescente abertura de cursos específicos a nível de pós-graduação e de graduação, voltados para formar academicamente profissionais Cientistas de Dados no Brasil, a tendência é que em médio e em longo prazo, os significados dos conceitos de Cientista de Dado profissional e de CCD sejam mais claros, em cenário nacional. Conforme consulta pública no E-mec (MEC, 2021), por cursos de Ciência de Dados de Graduação e Pós-graduação, no Brasil, a instituição pioneira foi a Universidade de São Paulo, em 2009, mas a maioria dos cursos de bacharelado em Ciência de dados nos Brasil, foram iniciados em 2020.

Atualmente os cursos de formação de Ciência de Dados no Brasil são: 17 cursos de bacharelado no total, sendo 3 deles na modalidade a distância e 14 presenciais. Desse total, 1 está localizado na região Sul, 2 na região Nordeste, 3 na região Centro-Oeste e 8 cursos, na região Sudeste. São 28 cursos tecnólogos no total, sendo 18 na modalidade a distância e 10 presenciais. Desse total, 2 estão na região Nordeste, 3 na região Sudeste e 5 na região Sul. Em nível de especialização, o primeiro curso iniciou em 2016 e são 99 cursos, no total, sendo 58 na modalidade EAD e 41 presenciais. Desse total de cursos de especialização, 5 deles estão localizados na região Centro-Oeste, 5 na região Norte, 8 na região Nordeste, 11 na região Sul e 12 na região Sudeste.

De toda forma, dado o tempo de preparação para a conclusão desses cursos formais e devido à falta atual de profissionais com uma formação específica concluída, torna-se mais evidente o crescimento do papel do CCD, em organizações brasileiras em curto prazo.

2.2.2 O Significado de Cidadão Cientista de Dados

No presente estudo, o significado de CCD faz referência a um profissional que possui avançadas habilidades em análise de dados, e usa dessa habilidade,

juntamente com o conhecimento na área específica do negócio, para gerar análises preditivas e *insights* (DATAROBOT, 2020; GARTNER, 2019).

Nesse cenário as empresas estão capacitando seus profissionais atuais de análise de dados e outros especialistas em informações de domínio do negócio, com as ferramentas e o suporte de que precisam para se tornarem CCD. Normalmente, já são profissionais de uma linha de negócios como vendas, *marketing*, finanças ou recursos humanos, que possuem um profundo conhecimento dos desafios, que seu respectivo departamento enfrenta no cotidiano.

Fazendo uso de específicos e poderosos *softwares*, os CCDs podem realizar análises diagnósticas detalhadas e criar modelos de aprendizado de máquina para complementar e substituir o trabalho, que exigiria um Cientista de Dados profissional (DATAROBOT, 2020). Dessa maneira, o papel desempenhado pelo CCD, preenche a lacuna entre aqueles que fazem análises de autoatendimento, como usuários de negócios e aqueles que fazem análises avançadas, como os Cientistas de Dados.

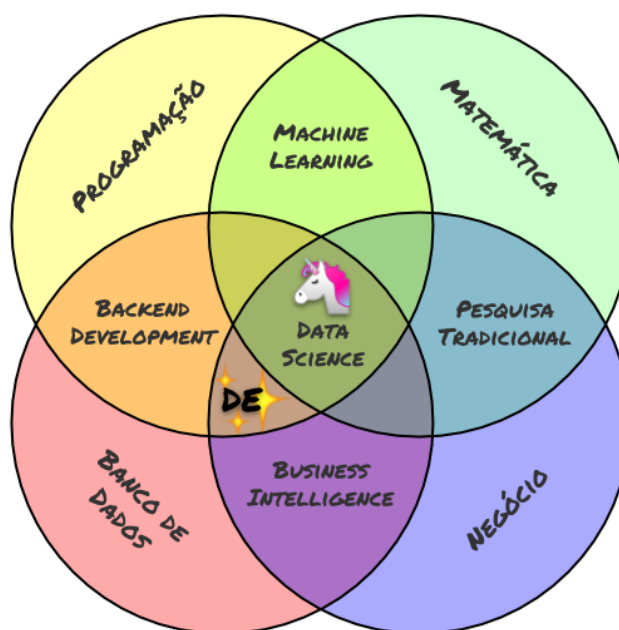
Embora, pareça haver um consenso em torno das habilidades necessárias e de tarefas a serem concluídas, por se tratar de um novo conceito, isso ainda não está claramente definido. Em um artigo publicado na Forbes (BANKER, 2018), o CCD é caracterizado como um profissional que detém habilidades semelhantes à de Ciência de Dados, como Estatística por exemplo, mas não tão avançadas quanto um Cientista de Dados. Nesse mesmo artigo, Banker (2018) descreve as responsabilidades desse papel como a de se aprofundar em novos dados, melhorar os modelos existentes e criar e implantar modelos adicionais para obter uma visão adicional acerca dos dados.

Atualmente, os CCDs desempenham um papel complementar aos Cientistas de Dados, que são profissionais com formação e especialistas nessa área do conhecimento. Cabe ressaltar que os CCDs não substituem os profissionais especialistas, pois não possuem tal a formação específica e avançada em Ciência de Dados, para fazê-lo. Por outro lado, certamente trazem a sua própria experiência e as habilidades únicas para o processo (IDOINE, 2018).

Desse modo, o CCD é um usuário com profundas habilidades analíticas, mas não tão avançadas quanto de um Cientista de Dados profissional. Assim, em geral, o diferencial de um CCD em relação a um Cientista de Dados seria o seu conhecimento e larga experiência na área do negócio, mas isso não consiste em uma vantagem significativa o suficiente a ponto de tornar-se o Cientista de Dados profissional, um recurso humano desnecessário numa organização orientada a dados. Levando estes pontos de atenção em consideração, não seria mais simples contratar apenas Cientistas de Dados profissionais, que também tivessem conhecimento na área do negócio?

Na Figura 2, o Diagrama de Venn não está representado de uma forma tradicional. Esse diagrama modificado representa o fato de que atualmente o Cientista Dados profissional é tão escasso no mercado de trabalho, em virtude da adição da necessidade de conhecimento no negócio, tornando-o um profissional tão raro quanto um unicórnio (CASIMIRO, 2019). Assim, se explica o surgimento do papel do CCD que aconteceu justamente a partir da relação entre a crescente demanda do mercado por mais profissionais de Ciências de Dados e a demorada formação de novos profissionais nessa área. Embora seja importante, mais simples e vantajoso a contratação de um Cientista de Dados, esse profissional é difícil de ser recrutado no mercado atual de trabalho e sua contratação se torna mais restrita, quando se requer conhecimentos específicos do negócio da organização contratante.

Figura 2 - Diagrama de Venn da Ciência de Dados, modificado.



Fonte: Adaptado de CASIMIRO, 2019.

Mesmo que a função CCD tenha sido uma adaptação provisória do mercado, ela não tende a desaparecer com o tempo, pois já demonstrou a importância do seu papel. Geralmente, os profissionais de análise de dados se concentram em encontrar tendências em dados de eventos que já ocorreram, na análise descritiva e expressam seus resultados em painéis visuais, relatórios estáticos e gráficos (DATAROBOT, 2020).

Devido ao avanço das ferramentas profissionais, os CCDs têm sido capazes de resolver problemas de negócios, usando dados com maior profundidade e amplitude do que seus colegas de análise de dados. Eles podem construir modelos e realizar análise preditiva para recomendar por exemplo, o próximo melhor curso de ação, para identificar os clientes em potencial com maior probabilidade de comprar um produto, para determinar quais empréstimos têm maior probabilidade de inadimplência, entre outros (DATAROBOT, 2020).

Embora, um CCD não seja capaz de programar plenamente ou construir modelos estatísticos avançados, ele pode usar ferramentas analíticas para gerar

relatórios de rotina, criar e personalizar visualizações de dados, analisar dados para tomar decisões de negócios, entre outros (REILLY, 2019).

O CCD tem maior flexibilidade e autonomia para tomar decisões e fazer escolhas estratégicas baseadas em dados nos projetos corporativos em que trabalha, porque pode utilizar os dados brutos e a modelagem a seu favor, além de ter conhecimento específico daquela função no domínio do negócio (CONTENT, 2020).

2.2.3 Desenvolvimento do Papel de Cidadão Cientista de Dados

Várias vantagens podem ser identificadas no desenvolvimento de competências em recursos humanos designados para atuação no papel de CCD dentro das organizações, entre elas estão a análise preditiva sendo realizada por um indivíduo que tem experiência na área do negócio. Assim este profissional tende a ter mais autonomia para colocar *insights* em prática, o que refletirá em uma economia de capital e tempo para a empresa, num primeiro momento.

Profissionais das próprias organizações no papel de CCD têm conhecimento na área do negócio e podem gerar *insights* mais precisos e uteis, apresentando uma visão mais ampla, assertiva e competitiva do negócio, colocando a organização à frente no mercado, num menor prazo.

Outra vantagem da designação de um recurso humano previamente contratado para outra função no papel do CCD está no rompimento de silos organizacionais. Silos organizacionais acontecem quando os indivíduos têm dificuldade de colaborar entre setores, áreas ou times diferentes (BASTOS, 2018).

Uma série de fatores podem levar a formação de silos, tais como (BASTOS, 2018): estruturas organizacionais funcionais, divididas por áreas; incentivos financeiros tendem a fortalecer os silos; localização e posição geográfica, diferentes departamentos negociando pela alocação de um recurso escasso; cadeia de comando como uma peça-chave para o surgimento do silo organizacional.

Para Zeenea (2020), a Ciência de Dados em si pode ser um motivador dos silos organizacionais. Todas as funções são importantes na estratégia de gerenciamento de dados de uma organização e todas expressaram interesse em aprender sobre essa área e suas habilidades como a de aprendizado de máquina (ZEENEA, 2020). Ainda para Zeenea (2018), a maior parte do conhecimento de Ciência de Dados, Inteligência Artificial e algoritmos de aprendizado de máquina está isolada num departamento relacionado à essas áreas ou em funções específicas, podendo ter como resultados, esforços de ciência de dados invalidados e não alavancados frequentemente.

Ao estabelecer uma comunidade de prática de funções de negócios e de TI que fornece diretrizes ou recursos detalhados, pode permitir que as organizações capacitem internamente os CCDs (ZEENEA, 2020). Para Zeenea (2018) é importante que as organizações encorajem o compartilhamento de esforços de Ciência de Dados em toda a organização e, assim, podem vir a quebrar eventuais ocorrências de silos.

Mesmo com a democratização do acesso às ferramentas que permitem aos usuários não Cientista da Dados, realizarem a extração de informações necessárias para a realização de análises prescritivas e preditivas, ainda há a necessidades de habilidades específicas para a utilização, sendo que alguns profissionais dentro da própria organização podem possuir essas habilidades.

Considerando as habilidades analíticas que alguns profissionais possuem ou podem vir a adquirir, Idione (2018) trata esses indivíduos como *power users* (usuários avançados). Para a autora os CCDs são exemplos de *power users*, que podem realizar tarefas analíticas, simples e moderadamente, sofisticadas que anteriormente exigiam mais experiência.

Estar ciente de que um CCD não poderá substituir o Cientista de Dados é um ponto importante para a distinção do papel dentro das organizações. O CCD é um papel que evoluiu como extensão de outras funções ou papeis dentro da organização. Zeenea (2020), pontua que um CCD em potencial varia com base em suas habilidades

seu interesse em Ciência de Dados e em conhecimentos relacionados como Inteligência Artificial, *Machine Learning*, *Deep Learning*, entre outros.

Os papéis nas organizações que podem possuir o perfil requerido e se enquadrar ou serem designados para atuar no papel de CCD podem incluir (ZEENEA, 2020): analistas de negócios, analistas ou desenvolvedores de *BI*, analistas de dados, engenheiros de dados, desenvolvedores de aplicativos, gerente de linha de negócios, entre outras linhas de frente do negócio. Assim, para se tornar um CCD numa organização, seriam necessárias algumas habilidades e características requeridas no perfil desse papel, conforme descreve no item 2.3.

Como já mencionado anteriormente, o CCD não substituirá o profissional de Ciência de Dados dentro das organizações, nem tem tendência de desaparecer do mercado em curto prazo. A visão que se tem, é que ambos os papéis coexistirão no desenvolvimento do negócio por algum tempo ainda (BARKER, 2018).

Barker (2018) aponta que um Cientista de Dados tem como responsabilidades selecionar os algoritmos, criar modelos de dados avançados, treinar e implantar uma solução técnica. Já, um CCD tem habilidades semelhantes às da Ciência de Dados, como Estatística, mas não tão avançadas quanto um Cientista de Dados. O CCD é um papel intermediário, que investiga novos dados e apoia a melhoria de modelos já existentes ou criados por Cientistas de Dados. Ele cria e implanta modelos básicos adicionais, para obter uma visão agregada e de valor ao negócio.

Todavia, se faz necessário reconhecer que para o crescimento do papel de CCD é preciso a obtenção de competências em automatização da Ciência de Dados no intuito de criar instrumentos mais intuitivos e simplificados de visualização de dados, adição de fluxo de dados para análise e ações facilitadas, por meio de manipulação de ícones ativos dentro de uma ferramenta de trabalho virtual, *drag and drop* (CONTENT, 2020). Cabe ressaltar que o *drag and drop* ajuda a movimentação dos dados de forma mais simplificada e intuitiva para analistas que não são, necessariamente, programadores e Cientistas de Dados (CONTENT, 2020).

Vários fatores têm contribuído com o crescimento do papel de CCD, sendo que um dos motivos para o aumento está relacionado a necessidade das organizações em encontrar soluções para implementar melhorias em sua operação e a necessidade de profissionais adeptos a metodologia como os CCD. Com isso é notável um movimento crescente para desenvolver funcionários adaptados ao novo papel, com treinamentos baseados em ciência de dados. Além de desenvolver profissionais internos, a contratação de novos talentos com experiência prévia pode trazer impactos positivos.

A geração e o armazenamento de volumes cada vez maiores de dados têm impactado direto no crescimento desse papel, pois gerar dados, mas não coletar informação e não colocar *insights* em ação, não terão utilidade estratégica (CONTENT, 2020). Portanto, se torna perceptível que desenvolver habilidades de CCD está em forte ascensão para as mais diversas áreas profissionais. Pode-se dizer que estar capacitado para inovar com base em dados e diversificar processos com tecnologia é essencial para impactar positivamente os resultados de qualquer empresa (CONTENT, 2020).

2.3 TRABALHOS ANTERIORES RELACIONADOS

Este item destaca a lacuna de investigação identificada e o que está sendo proposto de diferencial neste trabalho de forma complementar os trabalhos anteriores relacionados ao assunto e que foram encontrados ao longo do desenvolvimento dessa pesquisa. Vale ressaltar que a busca de publicações foi feita nas plataformas *Scopus*, *IEEE*, *Springer* e *ACM*, utilizando as *Keywords* “Cidadão Cientista de Dados”, “Cientista de dados”, “*Data Scientists*” e “*Citizen Data Scientists*”, considerou-se as publicações a partir de 2016, em Inglês e em Português.

De acordo os trabalhos encontrados, é possível perceber que o papel CCD tem ganhado destaque em pesquisas e publicações, no entanto, o presente trabalho vem a complementá-los por considerar um enfoque diferente na pesquisa e levantar a

percepção de profissionais atuantes, para que se possa estudar o fenômeno, cooperando com uma investigação de natureza quantitativa, adotando o método de pesquisa *survey* para analisar o papel do CCD atuantes em cenário nacional.

Esse trabalho tem como objetivo agregar, e por consequência se diferenciar dos anteriores por contemplar um escopo mais amplos e apresentar um instrumento de pesquisa para que se possa mensurar a importância e apresentar as competências e responsabilidades ligadas a função de cidadão cientista de dados. Esse instrumento tem como fundamentação uma revisão de literatura sobre o assunto em questão, e por não ter sido identificada uma escala valida e aplicada para tal, em estudos anteriores. Colabora com o levantamento de opiniões dos profissionais atuantes na área, sobre as principais competência para o desempenho do papel.

Sandes, 2019, caracterizou as competências necessárias para um profissional Cientista de Dados, destacando além das competências tradicionais de um profissional que irá trabalhar com dados, atribuições que o mercado passa a buscar nestes profissionais.

[...] um conjunto de habilidades consideradas essenciais pelas organizações na prática da gestão do conhecimento. [...] essas funções estão relacionadas com a execução de atividades na área de classificação das fontes informacionais, acesso, recuperação e análise da informação, desenvolvimento de produtos e serviços a partir da informação, união do conhecimento com a experiência das pessoas dentro da organização. (SANDES, 2019)

Para o autor, além do que ele chama de ‘espinha dorsal das competências exigidas’, não se pode esquecer a ênfase aos conhecimentos administrativos do negócio. Em seu trabalho, Sandes (2019) destaca também a interdisciplinaridade da Ciência de Dados:

Além disso, por ser também multidisciplinar, a ciência de dados possibilita que diversas áreas trabalhem e estudem em conjunto. A principal razão por ser tão atraente para as organizações é sua associação ao processo, análise, interpretação de dados. (SANDES, 2019)

Stadelmann, Stockinger, Burki e Braschelr (2019), em seu livro *Applied Data Science*, abordaram sobre os aspectos da profissão, formação e habilidades de um cientista de dados. Além dos aspectos tradicionais da ciência da computação como as capacidades analíticas, de programação e gestão de dados, também destacam o *Business Thinking*:

Pensar economicamente ajuda a ter um objetivo claro em mente, em vários níveis: contribui para não perder de vista todo o processo de desenvolvimento de um dado produto ao se concentrar, por exemplo, nos desafios analíticos em questão; também ajuda na alocação de recursos para as várias etapas de um projeto e pesar as opções a fim de produzir resultados relevantes para os negócios. Isso acabará por impulsionar o sucesso da análise esforços, uma vez que a maioria das partes interessadas (em empresas, pesquisa ou sociedade) será convencida não pela frieza da engenharia, mas pelos resultados realmente úteis entregues no prazo e no orçamento. (tradução livre) (STADELMANN, STOCKINGER, BURKI E BRASCHEL, 2019 p.34)

Os autores também citam que as habilidades profissionais de um cientista de dados exigidas pelo mercado, caracterizam uma espécie de 'profissional unicórnio'. Levanta também o debate sobre dois tipos de cientistas de dados, o tipo I e o tipo II:

Os cientistas de dados "Tipo II" são gerentes, preocupado em contratar e liderar profissionais de dados e ter uma visão de alto nível de potencialidades e funcionamento das ciências de dados. Por outro lado, os cientistas de dados "Tipo I" sabem como para "fazer as coisas" tecnicamente. Isso abre o caminho para currículos combinados para o tipo de gerente pessoas e pessoas com orientação técnica (funções diferentes), sem comprometer o perfil interdisciplinar de qualquer um deles. Por outro lado, a tentativa de isolar subtipos de um "Tipo I" se reduz a apenas renomear cargos tradicionais, como estatístico, analista de negócios, especialista em BI, minerador de dados, engenheiro de banco de dados. (tradução livre) (STADELMANN, STOCKINGER, E BRASCHEL, 2019, p37)

Mullarkey, Hevner, Gill e Gutta (2019) descreveram a capacitação dos colaboradores de uma empresa de manufatura avançada global, para atuarem como CCD, identificado a necessidade, crescente de profissionais capacitados, para extrair a informação de dados:

À medida que big data e ferramentas analíticas de dados se tornam cada vez mais disponíveis, um fator crítico, a escassez de cientistas de dados treinados tornou-se evidente [...] O problema é particularmente agudo em áreas como manufatura, onde a disponibilidade de dados é enorme, mas também carece de grandes dias de pagamento em potencial associado a serviços financeiros, que atualmente emprega a maior parte dos treinados cientistas de dados (tradução livre). (MULLARKEY, HEVNER, GILL E GUTTA, 2019, p. 192)

Uma vez identificado a dificuldade de contratação de Cientistas de Dados treinados, a empresa optou por treinar seus profissionais já disponíveis, aproveitando-se de seu conhecimento prévio das necessidades da empresa:

Um aspecto significativo do desafio de adquirir cientistas de dados adequados deriva do fato de que a análise eficaz de *big data* envolve duas áreas distintas de especialização: especialização nas ferramentas e técnicas de ciência de dados, e especialização no domínio de qual deve ser aplicado. Enfrentando essa escassez, GCM1 - uma manufatura multinacional empresa localizada na Flórida, EUA, produzindo produtos como placas de circuito, com receita anual de mais de US \$ 18 bilhões - decidiu virar o problema. Em vez de contratar cientistas de dados e, em seguida, ensinando-lhes o negócio, a empresa pegaria funcionários que já conhecem seus respectivos domínios e os treinam em dados relevantes técnicas de ciência. (tradução livre). (MULLARKEY, HEVNER, GILL E GUTTA, 2019, p. 192)

Para a capacitação de seus colaboradores, a empresa contou com a ajuda de pesquisadores treinados em *Design Science Research (DSR)* que é um método de pesquisa que operacionaliza e baseia-se no paradigma de *Design Science*, originário da área de Engenharia de Produção (SILVA, SARTORI, HIRATA E BERKENBROCK, 2016). Bax (2015) caracteriza a DSR como uma metateórica, que auxilia o pesquisador de forma a criar conhecimento teórico durante os processos de concepção de artefatos, justificando como tais processos podem ser significativos para a comunidade científica. Após dois anos de aplicação do DSR, os pesquisadores obtiveram *insights* significativos sobre a maneira pela qual a teoria e prática de DSR.

Foi possível identificar pelos profissionais da empresa e os pesquisadores, que as principais razões para desenvolver CCD podem ser destacadas (MULLARKEY, HEVNER, GILL E GUTTA, 2019): (i) o fato de que o acesso a dados interessantes estava muito mais disponível em todo o empreendimento; (ii) as ferramentas de

Ciência de Dados (SQL, NoSQL, estatísticas, programação, visualização etc.) possuem interfaces de usuário que os tornam acessíveis a "cidadãos" menos técnicos dentro da empresa; (iii) o poder de processamento está muito mais disponível no desktop e em toda a empresa precisava analisar e manipular dados massivos; (iv) os especialistas de domínio possuíam a capacidade técnica necessária para aprender as ferramentas e estavam muito melhor posicionados para identificar as principais áreas de investigação que provavelmente levam a percepções interessantes; (v) os especialistas do domínio tiveram a capacidade de interpretar os resultados de maneiras que podem levar a vantagens competitivas e suporte para uma melhor tomada de decisão na empresa.

Prayaga (2019) projetou um conjunto de ferramentas e interfaces para capacitar os CCDs e os auxiliar a extrair *insights*, a partir de conjuntos de dados gerados por seus dispositivos e/ou outras plataformas de mídia social, afim obter *insights* valiosos. O objetivo do projeto foi capacitar indivíduos com ferramentas fáceis de usar e se tornarem CCDs. Essas ferramentas foram projetadas para encorajar o CCD a contribuir para o campo da Ciência de Dados e fazer uso dos avanços no campo para beneficiar sua tomada de decisão, resultando em melhor retorno sobre os investimentos das corporações onde eles atuam.

Para os autores, uma vez equipados com as ferramentas adequadas, os CCDs podem realizar análises diagnósticas complexas e criar modelos que alavancam análises preditivas ou prescritivas. Além disso, a pesquisa apoia o fato de que vários profissionais estão usando ferramentas analíticas de dados e novos formatos na forma de novos algoritmos ou visualizações para melhor compreender os dados e explorar dados de diferentes tipos.

A partir dos trabalhos supracitados, é possível inferir que a Ciência de Dados tem sido objeto de pesquisas e publicações, caracterizando-se como um campo de pesquisa multidisciplinar, mas ganhando maior destaque nas ciências exatas. É possível destacar a importância do profissional da ciência de dados, ter conhecimento

na área do negócio e estar capacitado para o manuseio de ferramentas específicas e extração de conhecimento a partir dos dados.

Frente necessidade crescente do profissional de Ciência de Dados e diante de sua escassez no mercado de trabalho, há a necessidade de as empresas utilizarem de seus profissionais atuantes em outras áreas e capacitá-los para o papel de CCD. Este trabalho se diferencia e complementar os trabalhos anteriores, ao identificar as competências dos profissionais atuantes no papel de CDD nas empresas, objetivando tornar esse papel mais claro e buscando delimitar as competências necessárias, para que os profissionais das mais diversas áreas, possam desenvolver e atuar no papel de CCD.

2.4 COMPETÊNCIAS DO CIDADÃO CIENTISTA DE DADOS

Este item aborda as competências do CCD. Entretanto, inicialmente, é necessário a compreensão do que é competência. O significado do conceito abrange não apenas habilidades, mas também conhecimentos e atitudes que possibilitam a realização das tarefas de forma mais eficaz. Muitas vezes, o conceito de competência é confundido com traços de personalidade, no entanto, as competências se destacam sendo mais óbvias e palpáveis, podendo ser desenvolvidas por meio da prática e do aprendizado (FERREIRA, 2010).

A revisão bibliográfica analisa o conceito de competência em várias definições. Na Administração, o conceito de competências têm duas abordagens a serem consideradas, nas dimensões estratégicas e individuais. Na dimensão estratégica, a visão de competência associa-se a estratégia competitiva da organização (PORTO, 2013). Barney (2001) abordou a dimensão estratégica na Visão Baseada em Recursos, na qual as principais preposições são: *(i)* foco da análise dos recursos internos em detrimento da análise do ambiente; *(ii)* ênfase dada aos diferentes tipos de recursos na elaboração das estratégias competitivas; *(iii)* exploração dos recursos

de maior valor, raros, não imitáveis e insubstituíveis, de forma mais efetiva que os concorrentes.

Considerando o proposto por Prahalad e Hamel (1990; 1993) o conceito está associado a definição e a implementação da estratégia competitiva não deve se fundamentar apenas no ambiente externo, mas de competências essenciais internas da organização, gerando um diferencial extraordinário, explorando e alavancando os recursos existentes (PORTO, 2013).

Para ser considerada uma competência essencial, uma competência organizacional deve satisfazer simultaneamente aos seguintes critérios: *(i)* contribuir de forma significativa para a agregação de valor aos produtos e serviços percebidos pelos clientes da empresa; *(ii)* oferecer acesso potencial a uma ampla variedade de mercados, negócios e produtos; deve ser de difícil imitação (PRAHALAD; HAMEL, 1990).

Apoiando a noção de competência nessa dimensão, Ruas (et al. 2010, p. 4) afirma que “as competências organizacionais são competências coletivas da organização, que asseguram a realização da missão, da visão e/ou da estratégia da empresa e resultam da combinação de recursos, capacidades, tecnologias e sistemas”.

Na dimensão individual, o termo surgiu dentro do contexto empresarial, na década de 1970, onde era utilizada para designar aqueles que realizavam as tarefas de forma mais eficiente (ZABALA, ARNAU, 2010). As competências individuais podem ser compreendidas como a soma de conhecimentos, habilidades e atitudes.

Para Behar (2013), esses elementos podem ser compreendidos em: *(i)* Conhecimento (C): corresponde ao saber do sujeito, que é construído e reconstruído constantemente, a partir de relações com o meio e diferentes situações enfrentadas; *(ii)* Habilidade (H): corresponde ao saber fazer e está relacionada à aplicação produtiva do conhecimento, num caráter prático, técnico ou procedimental; *(iii)* Atitude (A): corresponde ao elemento pelo qual a competência se concretiza, significando a

prontidão para agir, além de revelar o modelo mental, os valores e as crenças do indivíduo.

No que tange às competências do papel CCD, se observou que precisa ter a compreensão do contexto organizacional, em uma visão ampla, além de compreender a visão, a missão, as necessidades e como os dados poderiam ajudar no tratamento dessas necessidades e impulsionar os objetivos. Este profissional apresenta um pensamento divergente, sendo o ideal pensar fora dos padrões, criando modelos de dados e conexões, que vão além do que um leigo conceituaria (Watts, 2020).

Watts (2020) continua a destacar que um CCD precisa de fortes habilidades analíticas, sendo que isso deve ser identificado como uma marca registrada do papel. Ser capaz de realizar análises de dados bastante complexas faz parte do seu trabalho, bem como a capacidade de avaliar as informações de uma forma significativa, no contexto do negócio.

Para um CCD torna-se importante, não apenas avaliar os dados de forma lógica, mas também tirar conclusões e fazer interpretações significativas deles, conclusões que um usuário iniciante não teria capacidade de identificar. Deve-se enfatizar o valor comercial na análise dos melhores candidatos internos a um papel de CCD, pois trabalham em uma área adjacente à de Ciência de Dados, algo como atividades de processos matemáticos e analíticos. Os desenvolvedores de software e os engenheiros podem ser bons candidatos para a este papel, por exemplo.

Originalmente, o CCD pode ser um profissional advindo de diversas áreas do negócio, com competências que permitem desempenhar as atividades desse papel. Por esse motivo é necessário identificar quais são os conhecimentos, habilidades e atitudes importantes e mais comuns do perfil desse profissional. Na revisão de literatura efetuada na presente pesquisa e conforme conteúdo dos itens anteriores dessa seção, algumas competências já foram mencionadas. Entretanto, se observou que poucas publicações fazem menção diretamente ao perfil de CCD, sendo essa uma das principais contribuições desse estudo.

As tarefas desenvolvidas por esse papel vão além da filtragem de dados, sendo relacionadas à análise de dados com base em técnicas estatísticas e utilização de ferramentas especializadas (SANDES, 2019). Ao se considerar trabalhos anteriores realizados acerca de características dos profissionais, que desenvolvem análises preditivas dentro das corporações é possível elencar conhecimentos para o papel do CCD, como conhecimento multidisciplinar (CUNHA, 2000) a capacidade analítica e ter uma visão gerencial, servindo como aporte no momento de decisão (TARGINO, 2000). Ser criativo, para buscar novas soluções para problemas antigos e a capacidade de constante atualização, ligada a formação continuada e aprendizado de novas tecnologias (TARGINO, 2000).

Competências ligadas a obtenção de conhecimento técnico-científicos, para o desenvolvimento e execução de atividades relacionadas ao tratamento dos dados (VALENTIM, 2020). Habilidades relacionadas à facilidade de trabalhar em grupos, capacidade de utilizar e operar softwares específicos da área, gerenciar pessoas, liderança e ambiente organizacional, uma vez que cada indivíduo possui suas competências e capacidades diante das diversas situações vivenciadas (FERREIRA, 2003).

Belluzzo (2011) aborda cinco grupos de habilidades consideradas fundamentais para o trabalho com análise de dados: *(i)* O Grupo Informação: as habilidades que todos os profissionais devem possuir; *(ii)* O grupo Tecnologia: diz respeito a utilização de instrumentos; *(iii)* O grupo Comunicação: interdependência entre as noções de informação e comunicação; *(iv)* O grupo Gestão: está ligado à gestão da informação e a administração da informação que é gerida; *(v)* O grupo Outros Saberes; que destaca a importância do profissional que trabalha com informação, ser versado em áreas conexas. Mais uma vez a exigência de que o profissional possua conhecimentos da área do negócio (SANDES, 2019).

Para Schutt e O'Neil (2014), o profissional que desempenhará funções relacionadas à Ciência de Dados pode vir de diversos campos de atuação, além disso

esse profissional deve estar alinhado à estratégia da organização em que está inserido, resolvendo desde questões técnicas, até preocupações acerca de tomada de decisões dos gestores.

Na prática, os papéis do Cientista de Dados e do CCD, na prática, têm grande proximidade e limites estreitos e isso pode explicar a pouca quantidade de publicações acerca das competências desse papel, além do fato de ser um tema emergente. A principal distinção entre identificadas entre eles está no grau ou no quão aprofundado são os conhecimentos dos técnicos requeridos.

Enquanto, o Cientista de Dados apresenta um profundo conhecimento em Estatística, Ciência da Computação e outras áreas de conhecimentos afins, o mesmo, não conhece tão profundamente, o domínio do negócio. Por outro lado, o CCD apresenta habilidades minimamente suficientes para a realização de análises preditivas e identificação de *insights*, fazendo uso de ferramentas específicas, com um profundo conhecimento na área do negócio em que está inserido.

No cotidiano, as tarefas de responsabilidade, para ambos os papéis, podem ser listadas com diferentes níveis de contribuição, mas complementares do ponto de vista colaborativo (SANDES, 2019): realizar as perguntas corretas e necessárias para alinhar os resultados das soluções e análise com os objetivos do negócio em questão, identificar os dados que agregam valor, selecionar as melhores tecnologias e ferramentas e ajudar no processo de tomada de decisão.

No Quadro 1, são apresentados de modo resumido um conjunto requerido de conhecimentos, habilidades e atitudes para o perfil de CCD, obtidos a partir da interpretação e da revisão de literatura realizada na presente pesquisa. Nesse quadro é possível notar que cada autor revisado tem a sua visão do que é necessário em termos de competências para que um profissional esteja bem-preparado para gerar ativos e executar atividades sob responsabilidades desse papel. Em contrapartida pode-se identificar como padrão, que os autores destacam a capacidade de gestão, tomada de decisão, conhecimento da área do negócio. Para vários dos autores

supracitados é possível notar que essas competências podem ser aprendidas por profissionais de diversas áreas do conhecimento, ratificando a natureza interdisciplinar da função CCD.

Quadro 1 - Competências do Papel do Cidadão Cientista de Dados

		Descrição	Fontes
Competências (CHA)	Conhecimentos	<ul style="list-style-type: none"> Ter conhecimento multidisciplinar: dominar diferentes disciplinas e áreas de conhecimento relacionadas (Estatística, C. Computação, Matemática etc.) 	Schutt e O’Neil (2014); Belluzo (2011)
		<ul style="list-style-type: none"> Ter conhecimento na área do negócio: união de conhecimento e de experiência na área do negócio; conhecimento administrativo no negócio, incluindo conhecimento administrativo no negócio. 	Ferreira (2003); Sandes (2019); Mullarkey, Hevner, Gill, Gutta (2019)
		<ul style="list-style-type: none"> Ter conhecimentos de <i>softwares</i> específicos de extração e análise de dados. 	Ferreira (2003);
		<ul style="list-style-type: none"> Ter conhecimentos básicos relacionados à Ciência de Dados: conhecimentos relacionados à Inteligência Artificial, <i>Deep Learning</i>, entre outros. 	Zeenea (2020)
		<ul style="list-style-type: none"> Ter conhecimentos técnicos suficientes para uso das ferramentas específicas de Ciências de Dados. 	Mullarkey, Hevner, Gill, Gutta (2019)
		<ul style="list-style-type: none"> Ter visão sistêmica e de valor agregado ao negócio. 	Baker (2018)
		<ul style="list-style-type: none"> Ter conhecimento da organização: envolve entender como a organização funciona, como faz para gerar lucro e como atinge suas metas (Zadra e Porto, 2015). 	Schutt e O’Neil (2014)
Habilidade	<ul style="list-style-type: none"> Saber se comunicar: fazer com que todos os envolvidos consigam entender você (Zadra e Porto, 2015). 	Belluzo (2011)	

Atitudes	<ul style="list-style-type: none"> • Gerir e liderar pessoas: contatar e liderar profissionais de dados. 	Ferreira (2003); Stadelman, Stockinger e Braschehr (2019)
	<ul style="list-style-type: none"> • Prestar apoio a melhoria de modelos matemáticos já existentes. 	Baker (2018)
	<ul style="list-style-type: none"> • Realizar análises de dados bastante completas: precisa ter fortes habilidades analíticas, sendo capaz de analisar as informações de uma forma significativa, no contexto do negócio. 	Watts (2020) e Belluzo (2011)
	<ul style="list-style-type: none"> • Unir processos de negócio, análise e interpretação de dados. 	Sandes (2019)
	<ul style="list-style-type: none"> • Ser criativo; propor soluções criativas, práticas, racionais e econômicas (Zadra e Porto, 2015). 	Targino (2000)
	<ul style="list-style-type: none"> • Ser analítico: capaz de analisar e avaliar informações de forma significativa. 	Watts (2020); (Cunha, 2000)
	<ul style="list-style-type: none"> • Agir de forma alinhada à estratégia da organização. 	Schutt e O'Neil (2014)
	<ul style="list-style-type: none"> • Ser crítico: manter pensamento crítico, organização pessoal e responsabilidade (Zadra e Porto, 2015). 	Targino (2000)

Fonte: Elaborado pela autora.

3 METODOLOGIA

Esse capítulo apresentar aspectos do método de pesquisa adotado, a forma como foi delimitada a pesquisa e as técnicas para coleta e a análise dos resultados.

3.1 DELINEAMENTO DA PESQUISA

Essa pesquisa caracteriza-se pela natureza aplicada e pela abordagem quantitativa. Na pesquisa quantitativa, os resultados da pesquisa podem ser quantificados, podendo ser caracterizada pela utilização objetiva de medidas, na

maioria dos casos numéricas, com objetivo de comparar os resultados e criar comparações estatísticas, como esclarece Fonseca (2002, p. 20):

os resultados da pesquisa quantitativa podem ser quantificados. Como as amostras geralmente são grandes e consideradas representativas da população, os resultados são tomados como se constituíssem um retrato real de toda a população alvo da pesquisa. A pesquisa quantitativa se centra na objetividade. Influenciada pelo positivismo, considera que a realidade só pode ser compreendida com base na análise de dados brutos, recolhidos com o auxílio de instrumentos padronizados e neutros. A pesquisa quantitativa recorre à linguagem matemática para descrever as causas de um fenômeno, as relações entre variáveis etc.

Adota questionário eletrônico, como técnica de coleta de dados, para avaliar opiniões de pessoas (WAINER, 2007), por meio do método de pesquisa *survey*. Para Fonseca (2002, p.33), “a pesquisa com *survey* pode ser referida como sendo a obtenção de dados ou informações sobre as características ou as opiniões de determinado grupo de pessoas, indicado como representante de uma população-alvo, utilizando um questionário como instrumento de pesquisa”. Quanto ao seu propósito, essa *survey* é exploratória, pois o objetivo é familiarizar-se com o tópico, dar ênfase na determinação de quais conceitos devem ser medidos e como devem ser medidos, buscando descobrir novas possibilidades e dimensões da população de interesse, obtida por conveniência (KRAEMER, 1993).

No que tange ao tempo emprega-se corte-transversal (*cross-sectional*) pois a coleta dos dados ocorreu em um só momento, pretendendo descrever e analisar o estado de variáveis em um dado momento (FREITAS, OLIVEIRA, SACCOL e MOSCAROLA, 2000).

3.2 POPULAÇÃO E AMOSTRA

A representação de uma população, ou parte dela é a amostra (FINK, 1995). O grau de erro contido em uma amostra é reflexo da sua imperfeição e nesse caso é

importante traçar objetivos na realização do survey. Fink (1995), afirma que o tamanho da amostra é referente ao número de respondentes necessário para que os resultados obtidos sejam confiáveis. Assim, o aumento do tamanho da amostra diminui o erro, no entanto a partir de certa quantidade, não se tem mais uma forte contribuição agregada por coletar-se maior número de questionários.

O tamanho da amostra deve ser estabelecido considerando alguns aspectos (FINK 1995): se o universo é finito ou infinito; o nível de confiança estabelecido, usualmente 95% e o erro permitido, normalmente não ultrapassando 5%; e a proporção em que a característica foco da pesquisa se manifesta na população.

Para esta *survey*, o público-alvo foram profissionais brasileiros, que desenvolvem ou já desenvolveram as atividades de CCD em organizações. Para obtenção de uma amostra não probabilística e por conveniência foram convidados por meio eletrônico a responder o instrumento de coleta de dados, criado no *software Forms* da Microsoft, via rede de contatos da pesquisadora, em convidados por grupos de discussão e nas redes sociais.

3.3 TÉCNICA PARA COLETA DE DADOS

A coleta de dados de corte- transversal, foi realizada por meio de um questionário eletrônico, conforme o Apêndice A, para que o próprio respondente respondesse a pesquisa, caracterizando-a como *survey* eletrônica. (HAIR et al., 2005). O propósito desta etapa foi a identificação das competências importantes do perfil de CCD, primeiramente baseadas na revisão da literatura pertinente ao fenômeno (vide o Quadro 1) e, posteriormente, validadas por procedimento de pré-teste e levantadas em campo, por meio da opinião de profissionais envolvidos em atividades ou projetos nessa área.

3.3.1 Pré-testes do Instrumento de Pesquisa

Para Freitas, Oliveira, Sacool e Moscarola (2000), os enunciados que compõem o instrumento de pesquisa tomam como base a revisão da literatura pertinente ao fenômeno. Esse tipo de validação pressupõe que todos os aspectos do atributo medido sejam considerados pelo instrumento, ou seja, que os indicadores construídos são uma boa representação do fenômeno a ser estudado.

Tendo como base a revisão teórica sobre o tema foram aplicados a procedimentos de pré-teste na versão inicial do instrumento de pesquisa, tendo como objetivo a identificação de ajustes necessários, a partir dos pré-testes, que divididos em duas etapas: (i) pré-teste de especialistas (ii) pré-teste de respondentes. O pré-teste com especialistas foi realizado no período de 12 a 28 de maio de 2021. O mini currículo dos especialistas participantes está no Quadro 2, cujas informações foram obtidas na Plataforma Lattes. Essa validação ocorreu de forma individual, por e-mail, em mais de dois ciclos de revisão até a obtenção do consenso, junto aos especialistas.

Quadro 2 - Caracterização dos especialistas participante do pré-teste

Especialista 1: Doutora e Mestre em Administração. Especialista em Melhoria de Processos de Software. Bacharel em Ciência da Computação. Experiência de mais de 13 anos no mercado de Tecnologia da Informação (TI), em projetos de desenvolvimento de software, de inovação e de qualidade. Experiência de mais de 10 anos como docente de graduação presencial, nas áreas de TI e Administração. Experiência de mais de 9 anos como professora e coordenadora de cursos de especialização a distância nas áreas de Qualidade de Software, Governança de TI, *Big Data*, *Data Science* e *Data Analytics*, Gestão de Projetos e Gestão Estratégica em Banking.

Especialista 2: Mestrando em Ciência da Computação pela PUCRS desde 2021. Graduado em Ciência da Computação pela PUCRS em 2020. Certificado em Ciência de Dados pela PUCRS, em 2020. Pesquisador no laboratório de Confiabilidade e Segurança de Sistemas, vinculado ao Instituto Nacional de Ciência e Tecnologia Florence.

Fonte: Adaptado da *Plataforma Lattes* (2021).

Após as considerações dos especialistas, foram feitas as seguintes alterações julgadas como pertinentes: (i) alteração na ordem, no enunciado e nas opções de

resposta de algumas perguntas, para facilitar a compreensão; (ii) alteração na instrução para a resposta do participante; (iii) inclusão de novas questões para competência identificadas na literatura; (iv) alteração e diminuição do texto de abertura do formulário, a fim de torná-lo mais claro e conciso para o respondente.

Na segunda etapa, o pré-teste do instrumento de pesquisa de respondentes, ocorreu no período de 31 de maio a 03 de junho, contando com a participação de 5 profissionais, com proximidade às atividades de CCD. Foram convidados via e-mail e telefone de forma individual. Ao final de preenchimento do formulário foi solicitado que o participante enviasse um e-mail com as suas impressões e *feedbacks*. Conforme as considerações recebidas dos respondentes, foram feitas as seguintes alterações: (i) alteração do enunciado da segunda e terceira questões; (ii) correções de escrita. Conforme as alterações descritas, o instrumento foi melhorado de forma evolutiva até a sua versão final (Apêndice A), para a coleta propriamente dita.

3.3.2 Aplicação do Instrumento de Pesquisa

O perfil do qual são coletadas informações e servindo de base para a análise dessa pesquisa são profissionais que desempenham ou já desempenharam atividades do papel de CCD. A amostra da pesquisa foi selecionada por conveniência e envolve profissionais brasileiros, conhecidos da pesquisadora ou convidados e que aceitaram participar dessa pesquisa, respondendo ao questionário eletrônico, conforme estratégias de divulgação empregadas.

A coleta de dados aconteceu de 03 a 07 de junho de 2021, sendo utilizados vários canais de divulgação: (i) envio de mensagem eletrônica para rede de contatos de profissionais, que trabalham ou já trabalharam no papel de CCD; (ii) publicação em grupos de discussão e de pesquisas na área em universidades e empresas; (iii) publicação de mensagens nas páginas oficiais de comunidades de Ciência de Dados em redes sociais como *Facebook*, *Linkedin* e *Whatsapp*; (vi) publicação no perfil de redes sociais; (v) publicação entre alunos de especialização a distância na área de

Ciência de dados, ingressantes em março de 2021; (vi) publicação entre interessados e alunos de graduação presencial e a distância, na área de administração. A coleta de dados da *survey*, foi realizada por meio do endereço eletrônico: <https://forms.office.com/r/6g9xkffzQV>. Foi obtido um total de 46 respostas iniciadas, sendo todas concluídas até o final e consideradas válidas para a análise dos dados.

3.4 ANÁLISE DE DADOS

O objetivo desta etapa foi a análise de dados, com objetivo de organizar, sintetizar e padronizar os dados, de forma que permitam conclusões e respostas ao problema e objetivo da pesquisa (HOPPEN et al., 1996; GIL, 2009). Para dar suporte a esta etapa foram utilizadas as próprias ferramentas do *Forms* e o *software Microsoft Excel*.

Finalizada a coleta de dados da *survey* foi realizada uma depuração dos dados, considerando os questionários iniciados, concluídos, válidos e a taxa de resposta. Na sequência, foram realizadas estatísticas descritivas nas variáveis de dados gerais dos respondentes, com o objetivo de gerar um sumário do perfil dos respondentes.

Também foram obtidos valores de moda, média e desvio padrão, para as variáveis, que mediram as competências do papel. A moda é o valor que detém o maior número de observações, o que é apresentado com mais frequência, considerando um conjunto de dados, não sendo necessariamente única, pois ao contrário da média, pode apresentar-se da forma (GIL, 2009): (i) Bimodal, possui dois valores modais; (ii) Amodal, não possui moda (iii) Multimodal, possui mais que dois valores modais.

4 ANÁLISE DOS RESULTADOS

Nesse capítulo aborda-se a análise e interpretação dos dados obtidos, onde são relacionados os resultados com os objetivos da pesquisa. Para tal, os dados são

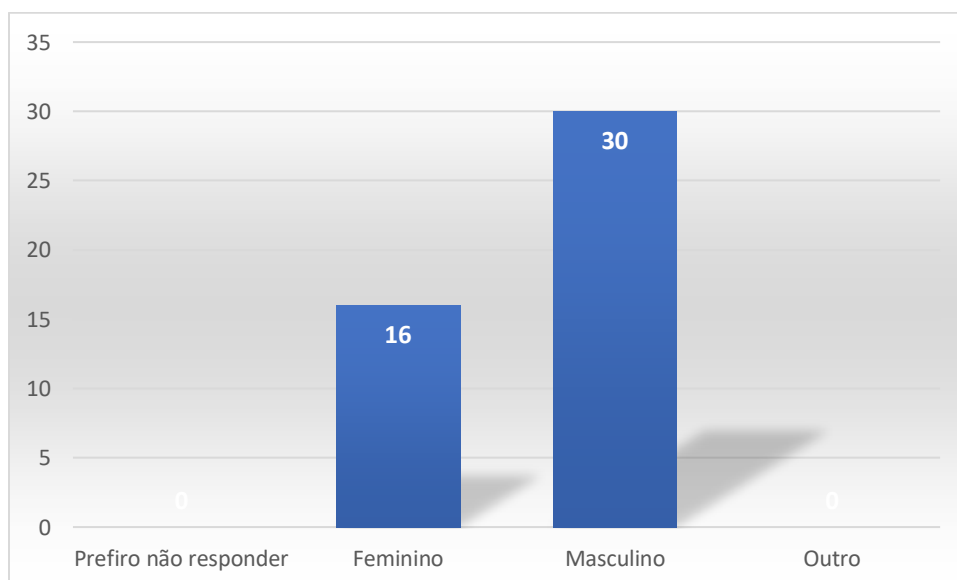
apresentados de acordo com sua análise estatística, incorporando no texto quadros e gráficos necessários à compreensão do raciocínio.

4.1 CARACTERIZAÇÃO DO PERFIL DOS RESPONDENTES

Os respondentes dessa pesquisa foram caracterizados de acordo com o gênero, idade, tempo de atuação no papel de CCD, segmentos de mercado que possuem experiência no papel, nível de formação acadêmica e área de formação acadêmica.

Em relação ao gênero dos respondentes, os respondentes do gênero masculino representam quase o dobro, em relação às respondentes do gênero feminino, conforme apresentado no Gráfico 1.

Gráfico 1 - Gênero dos respondentes

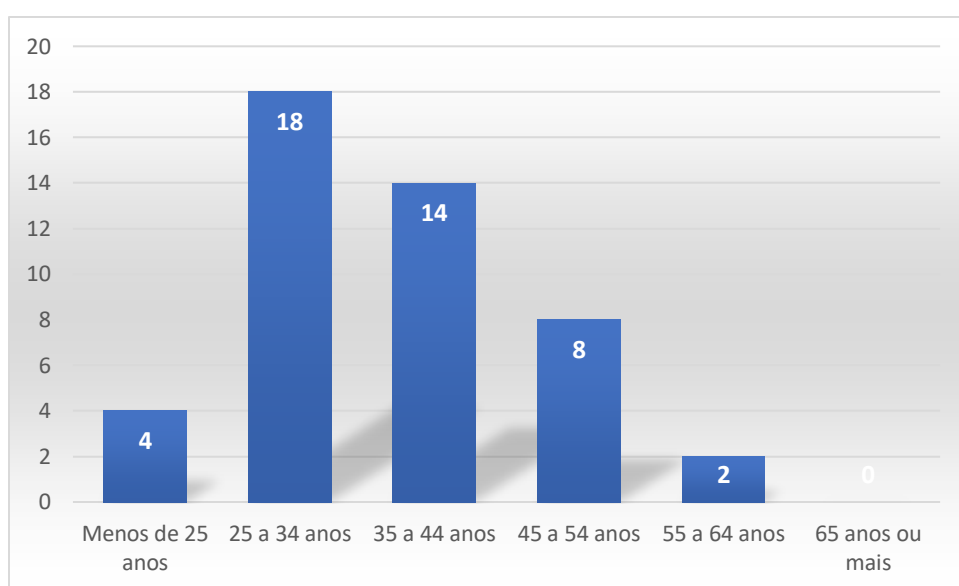


Fonte: Dados da Pesquisa.

De acordo com o critério de idade, os respondentes possuem, na sua maioria, a idade entre 25 e 34 anos e seguido da faixa entre 35 e 44 anos. Os respondentes

mais jovens da amostra possuíam menos de 25 anos e, os mais velhos se encontram na faixa dos 55 a 64 anos, conforme Gráfico 2.

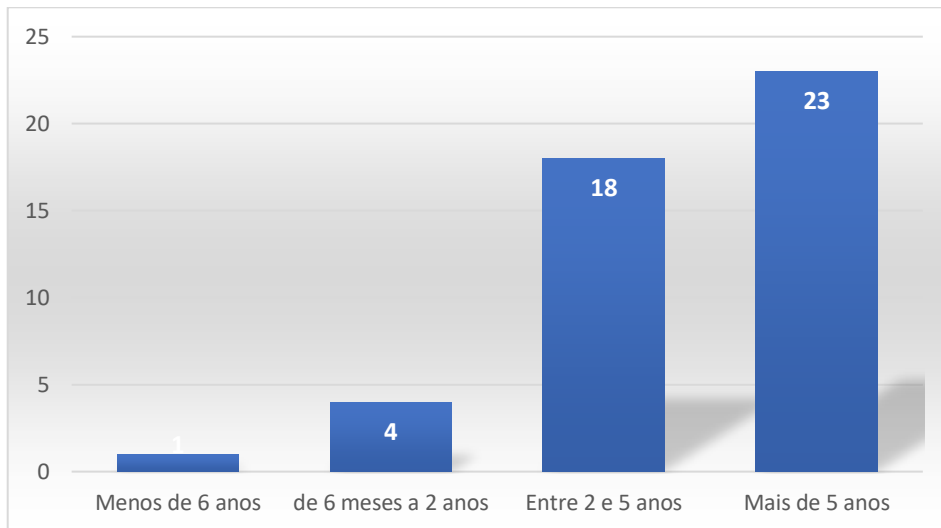
Gráfico 2 - Faixa etária dos respondentes



Fonte: Dados da Pesquisa

Ao analisar o critério tempo de experiência como CCD dos respondentes, os resultados da amostra mostram a maioria com mais de 5 anos de experiência e somente 1 possui menos de 6 meses atuando no papel de CCD, conforme Gráfico 3.

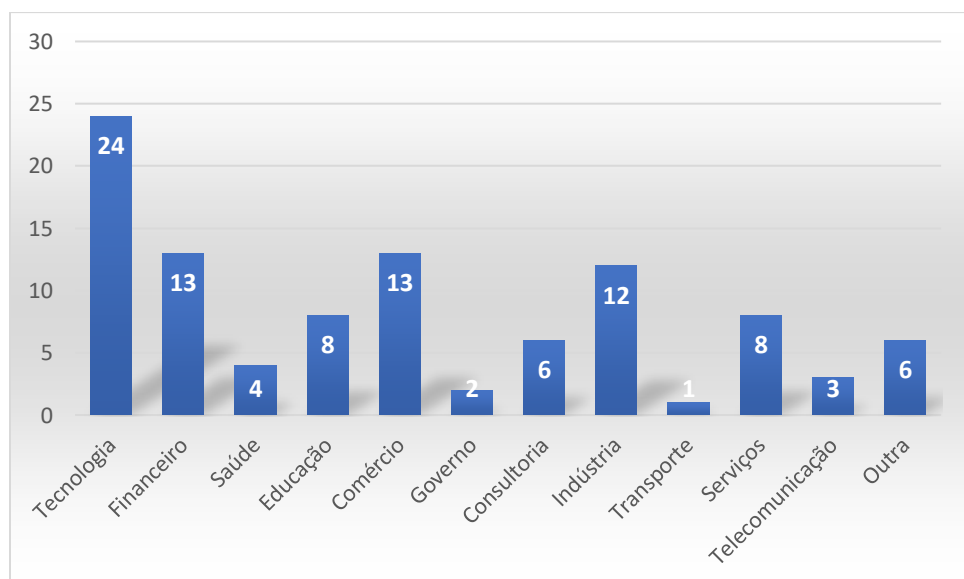
Gráfico 3 - Tempo de Experiência



Fonte: Dados da pesquisa

Em relação ao segmento de mercado com os quais os respondentes possuem experiência no papel de CCD, os respondentes poderiam considerar mais de uma alternativa, a maioria respondeu o segmento da Tecnologia, seguido por Financeiro e Comércio, ambos com 13 respondentes, conforme mostra o Gráfico 4.

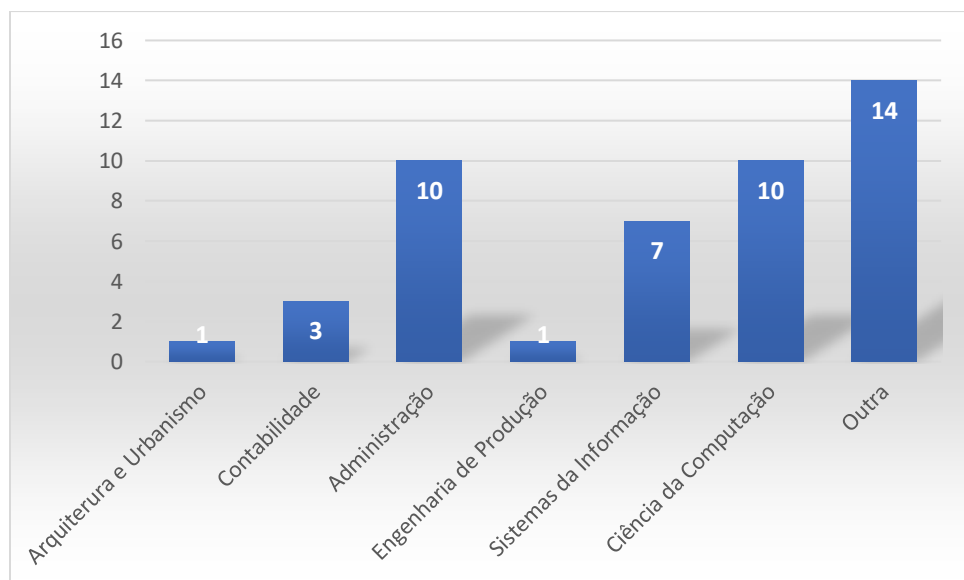
Gráfico 4 - Segmentos de Mercado



Fonte: Dados da pesquisa

Quanto à área de formação acadêmica, a maioria possui formação nas áreas de Administração e Ciência da Computação, com 10 respondentes cada, conforme se observa no Gráfico 5. Isso destaca a importância de conhecimentos relacionados à Ciência de Dados serem tão necessários, quanto os conhecimentos administrativos para os profissionais que exercem o papel de CCD (Mullarkey, Hevner, Gill e Gutta, 2019). Além de evidenciar a característica multidisciplinar da área, pois um total de 14 respondentes escolheu a opção “Outra”, sendo estas: Engenharia de Computação (4), Economia (3), Engenharia Elétrica (3), Engenharia de Controle e Automação (2), Estatística (1) e Matemática Aplicada (1).

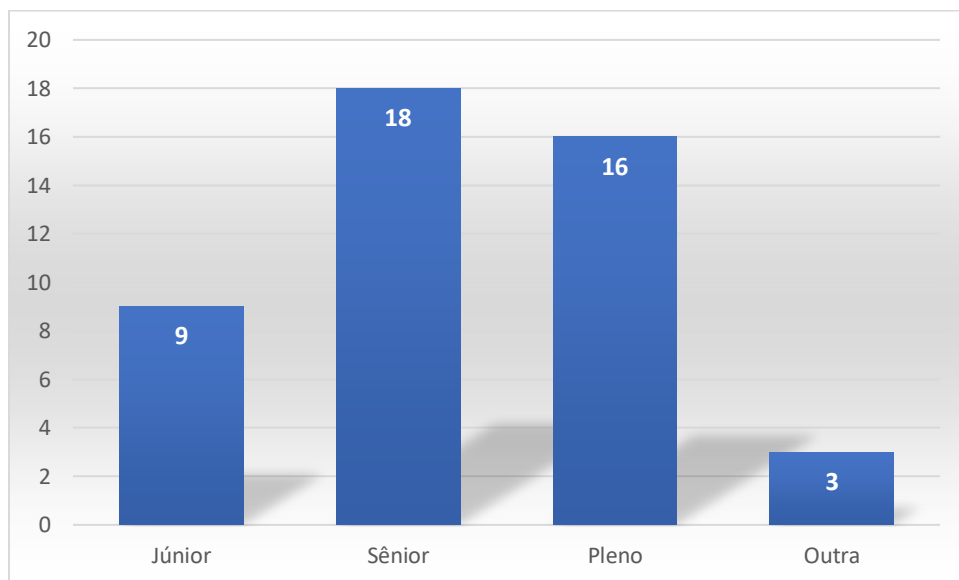
Gráfico 5 - Formação Acadêmica



Fonte: Dados da Pesquisa

Outras perguntas foram levantadas para avaliar o perfil dos respondentes, como a posição hierárquica na organização e o nível de formação acadêmica. Quanto à posição hierárquica, os respondentes, em sua maioria, ocupam a posição de *sênior*, conforme Gráfico 6.

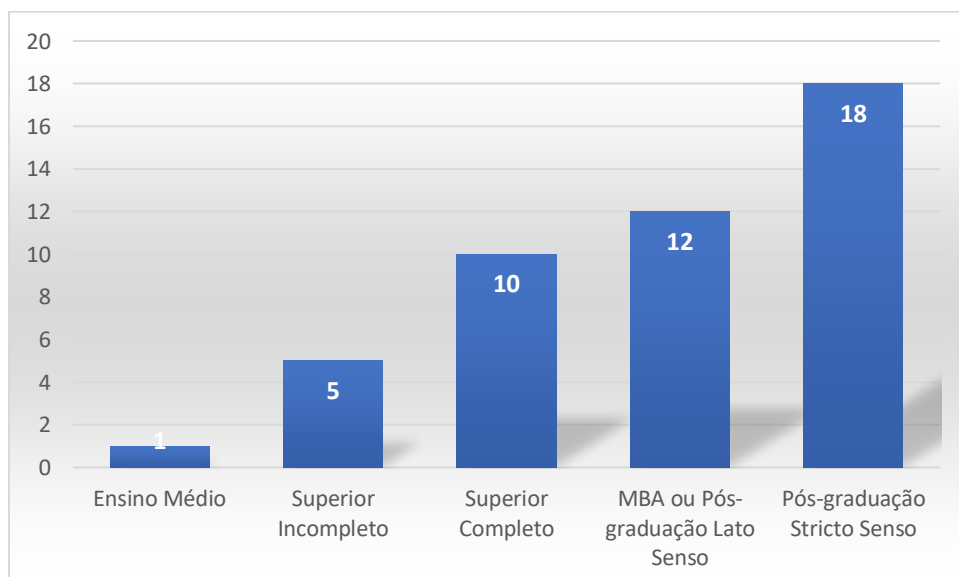
Gráfico 6 - Posição Hierárquica



Fonte: Dados da Pesquisa

Em relação ao nível de formação acadêmica, representada no Gráfico 7, a Pós-graduação *Stricto Senso* foi a maioria entre os respondentes (18), seguida por formação em nível de *Lato Senso* (12).

Gráfico 7 - Nível de Formação Acadêmica



Fonte: Dados da pesquisa

De modo geral, se pode constatar que o perfil dos respondentes dessa amostra se caracteriza em sua maioria pelo gênero masculino e por profissionais experientes, na faixa etária de 25 a 34 anos, graduados em Administração ou em Ciência da Computação. Dos respondentes, mais da metade têm mais de 5 anos de experiência no papel de CCD e com um nível de formação em pós-graduação *Stricto Senso* ou *Lato Senso*.

4.2 ANÁLISE DAS COMPETÊNCIAS DO PAPEL CIDADÃO CIENTISTA DE DADOS

Neste item é analisado o papel do CCD, em termos de competências consideradas importantes, segundo a percepção dos respondentes dessa amostra. No Quadro 3 são apresentados os resultados obtidos em relação à importância dos conhecimentos para atuar nesse papel de segundo a percepção dos respondentes da amostra dessa pesquisa, com a moda (MO), média (ME) e o desvio padrão (DP) para cada uma das variáveis.

Quadro 3 - Grau de Importância das Conhecimentos

Competências- Conhecimentos		Grau de Importância								
		0	1	2	3	4	5	MO	ME	DP
C1	Ter conhecimento multidisciplinar: dominar diferentes disciplinas e áreas de conhecimento relacionadas (Estatística, C. Computação, Matemática etc.)	0	0	2	6	15	23	5	4,283	0,766
C2	Ter conhecimento na área do negócio: união de conhecimento e de experiência na área do negócio; incluindo conhecimento administrativo no negócio.	0	0	0	13	16	17	5	4,087	0,755

C3	Ter conhecimentos de software, específicos de extração e análise de dados.	0	1	1	11	19	14	4	3,957	0,747
C4	Ter conhecimentos básicos relacionados à Ciência de Dados: conhecimentos relacionados à Inteligência Artificial, <i>Deep Learning</i> , entre outros.	0	0	1	14	16	15	4	3,978	0,749
C5	Ter conhecimentos técnicos suficientes para uso das ferramentas específicas de Ciências de Dados.	0	2	2	8	19	15	4	3,935	0,746
C6	Ter visão sistêmica e de valor agregado ao negócio.	0	0	2	6	16	22	5	4,261	0,765
C7	Ter conhecimento da organização: envolve entender como a organização funciona, como faz para gerar lucro e como atinge suas metas.	0	1	0	6	19	20	5	4,239	0,764

Fonte: Dados da Pesquisa

Os respondentes consideraram que quatro conhecimentos se destacaram com as maiores modas (5) na avaliação por grau de importância. A maioria dos respondentes (23) considerou a variável C1 com maior grau de importância, porém do total, três têm como base o conhecimento do negócio no qual estão atuando, que são as variáveis C6 (22), C7 (20) e C2 (17). Nenhum dos conhecimentos propostos foi considerável dispensável, para os participantes. O desvio padrão não teve muita variação no seu resultado, desta forma, é possível notar que há semelhança e heterogeneidade na percepção dos respondentes.

No Quadro 4, são apresentados os resultados obtidos em relação ao grau de importância das habilidades no papel de CCD, segundo a percepção dos respondentes da amostra dessa pesquisa, com a MO, ME e DP para cada uma das variáveis.

Quadro 4 - Grau de Importância das Habilidades

Competências - Habilidades	Grau de Importância
----------------------------	---------------------

		0	1	2	3	4	5	MO	ME	DP
H1	Saber se comunicar: fazer com que todos os envolvidos consigam entender você.	0	0	1	8	14	23	5	4,283	0,766
H2	Gerir e liderar pessoas: contatar e liderar profissionais de dados.	0	3	1	16	17	9	4	3,609	0,723
H3	Prestar apoio a melhoria de modelos matemáticos já existentes.	2	1	1	15	19	8	4	3,565	0,720
H4	Realizar análise de dados bastante completas, precisa ter fortes habilidades analíticas, sendo capaz de analisar as informações de uma forma significativa, no contexto do negócio.	0	0	0	3	20	23	5	4,435	0,775
H5	Unir processos de negócio, análise e interpretação de dados.	0	1	0	1	19	25	5	4,457	0,776

Fonte: Dados da Pesquisa

Desta vez, respondentes consideraram que três habilidades se destacaram com as maiores modas (5) na avaliação por grau de importância. A maioria dos respondentes (25) considerou a variável H5, seguida das variáveis H1 e H4, ambas com 23 respondentes. Para dois respondentes, a habilidade “H3 - Prestar apoio a melhoria de modelos matemáticos já existentes”, não demonstra grau de importância para o desenvolvimento do papel de CCD.

No Quadro 5, são expressos os resultados obtidos em relação ao grau de importância das atitudes necessárias no papel de CCD, segundo a percepção dos respondentes da amostra dessa pesquisa, com a MO, ME e o DP para cada uma das variáveis.

Quadro 5 - Grau de Importância das Atitudes

Competências - Atitudes		Grau de Importância								
		0	1	2	3	4	5	MO	ME	DP
A1	Ser criativo; propor soluções criativas, práticas, racionais e econômicas.	0	0	1	3	22	20	4	4,326	0,769

A2	Ser analítico: capaz de analisar e avaliar informações de forma significativa.	0	0	0	2	21	23	5	4,457	0,776
A3	Agir de forma alinhada à estratégia da organização.	0	0	1	2	24	19	4	4,326	0,769
A4	Ser crítico: manter pensamento crítico, organização pessoal e responsabilidade.	0	0	0	8	13	25	5	4,370	0,771
A5	Estar alinhado as estratégias da organização.	0	0	0	3	20	23	5	4,435	0,775

Fonte: Dados da Pesquisa

Em relação às atitudes, se pode observar que a maioria das variáveis, obtiveram modas máxima (5) na avaliação por grau de importância, para os respondentes, Ser crítico (A4), ser analítico (A2) e estar alinhado as estratégias da empresa (A5) tem alto grau de importância para o papel de CCD.

No Quadro 6, buscou-se compreender, utilizando os mesmos Conhecimentos, Habilidades e Atitudes (CHA), dos Quadros 3, 4 e 5, quais os graus de presença de domínio prévio que dos participantes, antes de iniciarem a sua atuação em atividades no papel CCD.

Quadro 6 - Grau de Domínio Prévio.

Conhecimentos, Habilidades e Atitudes		Grau de Domínio Prévio								
		0	1	2	3	4	5	MO	ME	DP
CHA1	Ter conhecimento multidisciplinar: dominar diferentes disciplinas e áreas de conhecimento relacionadas (Estatística, C. Computação, Matemática etc.)	0	0	8	14	11	13	3	3,630	0,725
CHA2	Ter conhecimento na área do negócio: união de conhecimento e de experiência na área do negócio; incluindo conhecimento administrativo no negócio.	1	2	7	12	14	10	4	3,435	0,709

CHA3	Ter conhecimentos de software, específicos de extração e análise de dados.	2	1	7	15	14	7	3	3,283	0,695
CHA4	Ter conhecimentos básicos relacionados à Ciência de Dados: conhecimentos relacionados à Inteligência Artificial, <i>Deep Learning</i> , entre outros.	2	2	13	10	15	4	4	3,000	0,667
CHA5	Ter conhecimentos técnicos suficientes para uso das ferramentas específicas de Ciências de Dados.	1	2	7	10	15	11	4	3,500	0,714
CHA6	Ter visão sistêmica e de valor agregado ao negócio.	3	1	5	11	15	11	4	3,457	0,711
CHA7	Ter conhecimento da organização: envolve entender como a organização funciona, como faz para gerar lucro e como atinge suas metas.	2	1	8	10	16	8	4	3,356	0,702
CHA8	Saber se comunicar: fazer com que todos os envolvidos consigam entender você.	0	0	9	8	15	14	4	3,739	0,733
CHA9	Conhecimentos administrativos	2	2	8	14	12	8	3	3,217	0,689
CHA10	Capacidade de utilizar e operar ferramentas específicas	2	2	4	17	14	7	3	3,304	0,697
CHA11	Gerir e liderar pessoas: contatar e liderar profissionais de dados.	2	1	8	14	15	6	4	3,239	0,691
CHA12	Prestar apoio a melhoria de modelos matemáticos já existentes.	2	1	11	14	9	9	3	3,174	0,685
CHA13	Realizar análise de dados bastante completas, precisa ter fortes habilidades analíticas, sendo capaz de analisar as informações de uma forma significativa, no contexto do negócio.	2	0	4	5	23	8	4	3,690	0,729
CHA14	Unir processos de negócio, análise e interpretação de dados.	2	1	3	8	19	9	4	3,619	0,724
CHA15	Ser criativo; propor soluções criativas, práticas, racionais e econômicas.	0	0	3	10	19	10	4	3,857	0,741

CHA16	Ser analítico: capaz de analisar e avaliar informações de forma significativa.	0	1	1	9	20	11	4	3,929	0,745
CHA17	Agir de forma alinhada à estratégia da organização.	2	0	4	10	18	8	4	3,571	0,720
CHA18	Ser crítico: manter pensamento crítico, organização pessoal e responsabilidade.	0	1	1	7	22	11	4	3,976	0,749

Fonte: Dados da Pesquisa

Segundo os respondentes, eles não apresentavam a maior moda em nenhuma das variáveis do CHA, aproximadamente 54% respondentes, consideraram que não apresentavam nenhum domínio em 13 das 18 variáveis apresentadas e 78% dos respondentes consideraram que apresentavam pouco ou muito pouco domínio nas variáveis apresentadas. A CHA13 foi a que apresentou, em números absolutos, maior número de respondentes, onde 23 dos participantes consideraram possuir alto domínio em “Realizar análise de dados bastante completas”.

4.3 DISCUSSÃO DOS RESULTADOS

Conforme apresentado por Sandes (2019), a análise de dados demanda, além dos conhecimentos técnicos da Ciência de Dados, também conhecimentos administrativos do negócio, dado este que foi apoiado unanimemente entre os respondentes, sendo que mais de 70% deles consideraram como importante ou muito importante.

Considerando os dados levantados na *survey*, o conhecimento multidisciplinar, compreendido como o domínio de diferentes disciplinas e áreas de conhecimento relacionadas, ganha destaque entre os respondentes em relação ao seu grau de importância, uma vez que mais de 80% consideraram esta competência como importante ou muito importante no desenvolvimento do papel de CCD. Este resultado

concorda com o levantamento literário, uma vez que a Ciência de Dados foi apresentada como uma área interdisciplinar possibilitando que diversas áreas trabalhem e estudem em conjunto (SANDES, 2019).

As capacidades analíticas, de programação e de gestão de dados são conhecimentos importantes para o desenvolvimento do papel (*powers users*) (STADELMANN, STOCKINGER, BURKI E BRASCHEL, 2019). Para 71% dos respondentes, ter conhecimento sobre *software* de extração e análise de dados, conhecimentos técnicos para uso de ferramentas para Ciência de Dados e conhecimentos em Inteligência Artificial e *Deep Learning*, foi considerado como importante ou muito importante.

Stadelmann, Stockinger, Burki e Braschelr (2019), destacam que além dos conhecimentos técnicos há importância também do *business thinking* (pensar economicamente), pois ajuda a ter um objetivo empresarial. Em convergência com esses autores, 82% dos respondentes consideraram como importante ou muito importante a variável “C6 - Ter visão sistêmica e de valor agregado ao negócio” e, ainda, 84% consideraram como importante ou muito importante “C7 - Ter conhecimento da organização”.

Mesmos que os conhecimentos técnicos tenham sido considerados com um alto grau de importância pela maioria dos respondentes, os mesmos, quando perguntados sobre quais os conhecimentos prévios possuíam, apresentaram na variável “C4 - Ter conhecimentos básicos relacionados à Ciência de Dados: conhecimentos relacionados à Inteligência Artificial, *Deep Learning*, entre outros”, o maior número de respostas como nenhum, pouco ou muito pouco domínio. Esse resultado, pode ser justificado pelo fato de que as ferramentas de Ciência de Dados (SQL, NoSQL, estatísticas, programação, visualização etc.) estão ganhando interfaces de usuário que se tornam acessíveis aos "cidadãos" menos técnicos dentro da empresa (MULLARKEY, HEVNER, GILL E GUTTA, 2019).

Prayaga (2019) considera que uma vez equipados com as ferramentas adequadas, os CCDs terão facilidade em realizar análises diagnósticas complexas e criar modelos de análises preditivas ou prescritivas. Vários profissionais já estão usando ferramentas analíticas de dados e novos formatos na forma de novos algoritmos ou visualizações, o que corrobora para melhor compreender os dados e explorar dados de diferentes tipos. Isso também reflete a percepção dos profissionais respondentes da *survey*, pois as habilidades relacionadas a analisar análise de dados bastante completas e significativas foram consideradas como importante ou muito importantes para mais de 90%.

Uma vez que o negócio é orientado aos dados, o acesso aos dados interessantes se torna mais disponível em todo o empreendimento (MULLARKEY, HEVNER, GILL E GUTTA, 2019). Desta forma, cresce a importância de profissionais com atitude analítica e crítica. Na coleta de dados, as variáveis “A4 - Ser crítico: manter pensamento crítico, organização pessoal” e “A2 - Ser analítico: capaz de analisar e avaliar informações de forma significativa.”, foram consideradas como importante ou muito importante para mais de 95% dos respondentes.

A variável “A1 - Ser criativo: propor soluções criativas, práticas, racionais e econômicas” apresentou mais de 95% respostas para importante ou muito importante. Essa representação, relaciona-se ao fato de que uma vez que profissionais da empresa tenham sido capacitados para interpretar dados, as práticas resultam em maneiras de obter vantagens competitivas e suporte para uma melhor tomada de decisão na empresa (MULLARKEY, HEVNER, GILL E GUTTA, 2019).

Embora, seja reconhecida, a importância de diversos conhecimentos, habilidades e atitudes, por praticamente a totalidade dos respondentes, a maioria também informou que, ao iniciar sua atuação no papel de CCD, não possuía boa parte dos CHAs, vindo a adquiri-los durante sua atuação na prática profissional. Os próprios profissionais reconhecem esses conhecimentos como essenciais, pois em certo momento precisarão aplicá-los. Para Mullarkey, Hevner, Gill e Gutta (2019), os CCDs

qualificados possuem a capacidade técnica necessária para aprender as ferramentas e estarem mais bem posicionados, para identificar as principais áreas de investigação que provavelmente levarão às percepções interessantes.

5 CONCLUSÃO

Esse capítulo tem o intuito de responder ao problema inicial e aos objetivos deste trabalho, através das considerações finais extraídas dos resultados obtidos com a pesquisa *survey* e sugeridas ao longo do desenvolvimento deste trabalho. Destacará-se as suas principais contribuições, suas limitações e, por fim, as sugestões de pesquisas futuras.

5.1 CONSIDERAÇÕES FINAIS

No início deste trabalho foram propostos três objetivos específicos, que ajudaram a alcançar o objetivo geral e encontrar uma possível resposta para a questão de pesquisa. Esses objetivos são: *(i)* identificar quais as competências necessárias para o desempenho de atividades atribuídas ao papel de CCD; *(ii)* realizar uma pesquisa *survey*, com profissionais que desempenham o papel de CCD; *(iii)* analisar os resultados da amostra da pesquisa, de acordo com a percepção dos profissionais que desempenham o papel em comparação às informações levantadas na bibliografia.

Tendo em vista o objetivo de “*(i)* identificar quais as competências necessárias para o desempenho de atividades atribuídas ao papel de CCD”, partiu-se para a revisão literária em busca da compreensão do papel do CCD e identificação do conjunto de CHAs. Viu-se que este papel surgiu recentemente e vem ganhando destaque devido a uma necessidade atual das corporações.

Durante o desenvolvimento da pesquisa, compreendeu-se como esse profissional possui avançadas habilidades em análise de dados e usa dessa

habilidade, juntamente com o conhecimento na área específica do negócio, para gerar análises preditivas e *insights*. Observou-se que os CCDs são *power users*, que possuem conhecimentos técnicos avançados o suficiente para a análise de dados e conhecimentos da área do negócio, onde desempenha o seu papel.

Considerando a pesquisa *survey* realizada, 23 dos respondentes da amostra coletada atribuem alto grau de importância ao conhecimento multidisciplinar no desenvolvimento do papel do CCD e 42 consideram alto grau de importância, os conhecimentos relacionados a ter visão sistemática de valor agregado ao negócio ou ter conhecimento da organização, destacando a importância do conhecimento na área do negócio e domínio de conhecimentos multidisciplinar para o CCD.

Para o objetivo específico “(ii) compreensão do papel do CCD, com profissionais que desempenham o papel de CCD em nível nacional”, foram coletadas 46 respostas de profissionais que atuam ou já atuaram, no papel de CCD. Foram feitas perguntas para caracterizar o perfil dos respondentes e sobre o grau de importância das competências, coletadas na revisão literária e com base nas percepções de importância de profissionais atuantes nesse papel.

Por fim, para o objetivo específico “(iii) analisar os resultados da amostra da pesquisa, de acordo com a percepção dos profissionais que desempenham o papel”, a aplicação de estatística descritivas e medidas básicas nos dados da amostra, possibilitaram a identificação e discussão das principais competências do CCD.

Este levantamento de dados foi realizado com profissionais atuantes no mercado, a maioria com mais de 5 anos de experiência e, por estes respondentes foram apontados quais os conhecimentos, habilidades e atitudes dos profissionais de CCD, considerando seu grau de importância para atuação. Porém, quando questionado quais CHAs estes mesmos profissionais já possuíam, antes de entrarem nesse papel no mercado de trabalho, foi possível identificar uma série de CHAs ausentes. Isso pode ser justificado pelo fato de as empresas estarem capacitando os seus profissionais atuais, para se tornarem CCD, assim não há exigências de um

profissional com todos os CHAs considerados importantes, plenamente desenvolvidos, antes do início da atuação nesse papel.

Assim concluiu-se que, de acordo com a percepção dos profissionais respondentes da amostra da pesquisa, o papel de CCD está de acordo com as competências levantadas na revisão literária, inicialmente. O papel do CCD existe pela escassez de profissionais de Ciência de Dados capacitados e disponíveis no mercado, frente a alta demanda do mercado por este profissional. O CCD preenche a lacuna entre as análises diagnósticas complexas e a criação de modelos que alavancam análises preditivas ou prescritivas e a aplicação desses *insights* de forma alinhada às estratégias e à realidade da empresa. O CCD é um profissional com habilidades para realizar a análise de dados bastante preliminares e completas e tem fortes habilidades analíticas, capazes de gerar as informações de uma forma significativa, agregando valor, no contexto do negócio.

5.2 CONTRIBUIÇÕES DA PESQUISA

A maior contribuição acadêmica desse trabalho consiste na delimitação das competências necessárias para a atuação profissional no papel de CCD, conforme apresentado no Quadro 1. Quanto à contribuição gerencial, o trabalho destaca a importância do CCD e a possibilidade de planejamento e de desenvolvimento de pessoal interno da própria organização. Além de ser uma necessidade passageira do mercado, um profissional capacitado nesse papel pode ser capaz de diminuir silos e aumentar a comunicação entre análises técnicas e o desenvolvimento do negócio.

Hoje o mercado está começando a observar a importância de capacitar seus profissionais, que possuem a capacidade técnica necessária para aprender as ferramentas específicas da área de Ciência de Dados, que foram adaptadas com interfaces de usuário que as tornam acessíveis aos "cidadãos" menos técnicos dentro da empresa. Assim, empresas orientadas aos dados podem levar vantagens

competitivas no mercado e ainda obter suporte para uma melhor tomada de decisão na empresa.

No que se refere à contribuição científica desse trabalho, se pode destacar a criação do instrumento de pesquisa para mensurar a importância de determinadas competências do papel do CCD, a partir da revisão de literatura efetuada.

5.3 LIMITAÇÕES DA PESQUISA

Entre as limitações têm-se o tamanho da amostra dessa pesquisa, devido ao grande obstáculo do público-alvo ser muito específico e ao curto prazo de coleta. Todavia, diversas estratégias e canais de divulgação foram adotados, visando amenizar tal limitação (redes sociais, envios de e-mail, contatos diretos com conhecidos etc.). Devido ao papel do CCD ainda ser considerado recente no mercado de trabalho, a obtenção de uma amostra maior ainda poderá ser buscada, futuramente, com novas ondas de coleta.

Assim, essa pesquisa está limitada e sujeita à incerteza em relação à representatividade dessa amostra. O tamanho amostral é importante, porque à medida que se amplia a amostra e diminui o grau do erro amostral, assim, tendo uma boa probabilidade de seleção de elementos de dois extremos da distribuição e, portanto, a média amostral tende a estar mais próxima da média populacional (DANCEY; REIDY, 2006).

5.4 SUGESTÕES DE PESQUISAS FUTURAS

Considerando as limitações do estudo supramencionadas e alguns aspectos pouco explorados, essa seção apresenta sugestões de pesquisas, que podem ser feitas por futuros pesquisadores, com o intuito de dar continuidade ou vir a complementar esse trabalho.

Nesse estudo, se investigou as competências do profissional que desenvolve o papel de CCD dentro das organizações, através dos profissionais atuantes. Sugere-se, como um trabalho futuro, delimitar responsabilidades do profissional, por meio de observação direta de pesquisa em profundidade, em uma dada unidade de análise.

No que tange a dar continuidade ao presente trabalho sugere-se um estudo de caso em empresas, selecionando um determinado grupo de profissionais, para a capacitação destes profissionais como CCD, seguindo estratégias adotadas por Mullarkey, Heyner, Grandon e Dutta (2019), para que se possa comparar se a percepção dos respondentes da *survey* e o levantamento bibliográfico deste trabalho e surtirão efeito prático, num contexto real.

REFERÊNCIAS

BANKER, Steve. The Citizen Data Scientist. Forbes. [S.I.] 2018. Disponível em: <https://www.forbes.com/sites/stevebanker/2018/01/19/the-citizen-data-scientist/?sh=20044b872702>. Acesso em: 07 nov. 2020.

BASTOS, Rodrigo. Silos Organizacionais. Tarrget Teal. [S.I.] 2018. Disponível em: <https://targetteal.com/pt/blog/silos-organizacionais/>. Acesso em: 08 nov. 2020.

BAUDISCH, Alfred Reinold. Ciência de Dados é explorar Big Data para fazer perguntas para prever o futuro. Medium. 2016. Disponível em: <https://alfredbaudisch.medium.com/o-que-%C3%A9-ci%C3%Aancia-de-dados-data-science-7af5bdac101a>. Acesso em: 01 jun. 2021.

BAX, M. P. Design science: filosofia da pesquisa em ciência da informação e tecnologia. *Ciência da Informação*, v.42, n.2, 2015.

BELLINGER, Gene; CASTRO, Durval; MILLS, Anthony. Data, Information, Knowledge and Wisdom. The Way of Systems. [S.I.] 2004. Disponível em: <https://www.systems-thinking.org/dikw/dikw.htm>. Acesso em: 07 nov. 2020

BELLUZZO, R. C. B. As competências do profissional da informação nas organizações contemporâneas. *RBBB, Revista Brasileira de Biblioteconomia e Documentação*, São Paulo, v. 7, n. 1, p. 58-73, jan./jun.2011

BRASCHLER, Marti; STADELMANN, Thilo; STOCKINGER, Kurt. *Applied Data Science: Lessons Learned for the Data-Driven Business*. Única edição. Winterthur, Suíça: Springer, 2019.

CAO, Longbing. Data Science: A Comprehensive Overview. *ACM Computing Surveys*, Sydney, Vol. 50, No. 3, Article 43, jun. 2017.

CHATFIELD, A. T. et al. Data Scientists as a game changers in big data environments. In: *PROCEEDINGS OF THE 25TH AUSTRALASIAN CONFERENCE ON INFORMATION SYSTEMS (ACIS)*, Anais..., Auckland: Auckland University of Technology, 2014. p.1-11.

CONTENT, Rock. O que faz os Citizen Data Scientists. Rock Content. [S.I.] 2020. Disponível em: <https://inteligencia.rockcontent.com/citizen-data-scientist/>. Acesso em: 02 nov. 2020.

CUNHA, M. V. O profissional da informação e o mercado de trabalho. Informação e Sociedade: Estudos. João Pessoa, v. 10, n. 1, p. 1-5, 2000.

DANCEY, C. P. e REIDY. Estatística sem matemática para psicologia: usando SPSS para Windows. Porto Alegre: Artmed, 2006.

DATAROBOT. Citizen Data Scientist. DataRobot. [S.I.] 2020. Disponível em: <https://www.datarobot.com/wiki/citizen-data-scientist/>. Acesso em: 02 nov. 2020.

DAVENPORT, H. Thomas Zeenea. Patil, D.J. Data Scientist: The Sexiest Job of the 21st Century. Harvard Business Review. Brighton, MA, 2012. Disponível em: <https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century>. Acesso em: 02 nov. 2020.

de dados. Tradução de Adonai Schlup Sant'Anna; Anselmo Chaves Neto. 5. ed. Porto Alegre: Bookman, 2005.

CADASTRO Nacional de Cursos e Instituições de Educação Superior. e-MEC. Disponível em: <<https://emec.mec.gov.br/>>. Acesso em: 31 de maio de 2021

FAROL. Ciência de dados – Entenda o que é e como colabora com sua empresa. FAROL. [S.I.] 2017. <https://farolbi.com.br/entenda-o-que-e-e-como-a-ciencia-de-dados-colabora-com-sua-empr>. Acesso em: 01 nov. 2020.

FERREIRA, D. T. Profissional da informação: perfil de habilidades demandadas pelo mercado de trabalho. Ciência da Informação, Brasília, v. 32, v n. 1, p. 42-49, jan./abr. 2003

FINK, Arlene. The survey handbook. Thousand Oaks, Sage, 1995a. [The Survey Kit, v.1]

FLATH, C.M.; STEIN. N. Towards a data Science toolbox for industrial analytics applications. Computers in industry, Wurzburg, German: Julius Maximilians University, n 94, 2017. Referências 27

FONSECA, J. J. S. Metodologia da pesquisa científica. Fortaleza: UEC, 2002. Apostila. GIL, A.C. Como elaborar projetos de pesquisa. 3ªed. São Paulo, Atlas, 1991. 159p.

FREITAS (H.), OLIVEIRA (M.), SACCOL (A.Z.) e MOSCAROLA (J.). O método de Gil, A. Como elaborar projetos de pesquisa. 4ªed. São Paulo: Atlas , 2009.

Ferreira, P. M. G. A liderança exercida pelo Conselho Geral: o papel do líder em regime de autonomia, gestão e administração escolar: um estudo de caso. Master's thesis, Universidade de Évora, 2010. Disponível em:

https://dspace.uevora.pt/rdpc/bitstream/10174/11848/1/Disserta%c3%a7%c3%a3o_Pedro_Manuel_Gon%c3%a7alves_Ferreira.pdf

HAIR, J. F.; TATHAM, R. L.; ANDERSON, R. E.; BLACK, W. C. Análise multivariada

Idoine, Carlie. Citizen Data Scientists and Why They Matter. Gartner. [S.I.]

2018. Disponível em: <https://blogs.gartner.com/carlie-idoine/2018/05/13/citizen-data-scientists-and-why-they-matter/>. Acesso em: 17 out. 2020

LAUDO, Kenneth C.; LAUDON, Jane P. Sistemas de informações. 7. Ed. São Paulo: Pearson Prentice Hall, 2007.

LE COADIC, Yves-François. A ciência da informação. 2 ed. Brasília: Brique de Lemos, 2004.

LOH, Stanley. BI na era do big data para cientistas de dados - indo além de cubos e dashboards na busca pelos porquês, explicações e padrões. Porto Alegre, 2014.

MEGGINSON, L. Lessons from Europe for American Business. The Southwestern Social Science Quarterly, v.44, n. 1, p. 3-13, 1963. Disponível em:

<http://www.jstor.org/stable/42866937>. Acesso em: 1 jun 2021.

Mullarkey M.T., Hevner A.R., Grandon Gill T., Dutta K. Citizen Data Scientist: A Design Science Research Method for the Conduct of Data Science Projects. Springer Nature, p. 191–205, 2019.

NATIVA. Data Science – saiba como a ciência de dados pode te ajudar!. Oracle. [S.I.] 2020. Disponível em: <https://navita.com.br/blog/data-science-saiba-como-a-ciencia-de-dados-pode-te-ajudar/>. Acesso em: 17 out. 2020.

ORACLE. Ciência de Dados. Oracle. [S.I.] 2020. Disponível em:

<https://www.oracle.com/br/data-science/what-is-data-science.html>. Acesso em: 17 out. 2020.

OLIVEIRA, Djalma de Pinho Rebouças. Estratégia Empresarial & Vantagem Competitiva. 6. ed. São Paulo: Atlas, 2009.

PRAHALAD, C.K.; HAMEL G. The core competence of the corporation. Harvard Business Review, p. 19-91, may-jun. 1990.

PRAYAGA, Lakshmi. Empowering Citizen Data Scientists-Cyber Science Social Media track. In: 2019 International Conference on Cyber Situational Awareness, Data Analytics And Assessment (Cyber SA). IEEE, 2019. p. 1-2.

pesquisa survey. São Paulo/SP: Revista de Administração da USP, RAUSP, v. 35, n. 3, Jul-Set. 2000, p.105-112

PINSONNEAULT, A. & KRAEMER, K. L. Survey research in management information systems: an assesement. Journal of Management Information System, 1993.

PORTO, Fábio; ZIVIANI, Arthur. "Ciência de Dados". In: 3o. Seminário de Grandes Desafios da Computação no Brasil, SBC, 2014.

RAUTENBERG, Sandro; CARMO, Paulo Ricardo Viviurka do. Big Data e Ciência de Dados: complementariedade conceitual no processo de tomada de decisão. Brazilian Journal of Information Studies: Research Trends. 13:1 p.56-67, 2019.

REILLY, Pete. Citizen Data Scientist | What it Means in 2019. Answer Rocket. [S.I.] 2019. Disponível em: <https://www.answerrocket.com/citizen-data-science/>. Acesso em: 08 nov. 2020.

RIBEIRO, C.J.S. Bid data: os novos desafios para o profissional da informação. Informação e Tecnologia (ITEC), João Pessoa/ Marília, v 1, p. 96-105, jan./jun. 2014.

RUAS, R. et al. Gestão por competências: revisão de trabalhos acadêmicos no Brasil no período 2000 a 2008. In: ENCONTRO DA ANPAD, 34., 2010, Rio de Janeiro. Anais... Rio de Janeiro: ANPAD, 2010. CD-ROM.

SANDES, Jorge. O perfil e o papel do cientista de dados. R. BNDES, Rio de Janeiro, v. 26, n. 52, p. 275-319, dez. 2019.

SILVA, Daniel Maniglia Amancio da; SARTORI, Deivid Felipe; HIRATA, Celso Massaki E BERKENBROCK, Carla Diacui Medeiros. Abordagem utilizando o Design Science Research para o Desenvolvimento de Sistema Colaborativo Assistivo. Revista de Informática Aplicada, v 12, n 1, p75-95, 2016.

SWAN, A.; BROWN, S. The skills, role and careers tructure of data scientists and curators: anassessment of current practice and future needs. Reporttothe Joint Information Systems Committee (JISC). Truro: Key Perspectives for JISC, 2008. 34 p.

APÊNDICE A – COMPETÊNCIAS DO CIDADÃO CIENTISTAS DE DADOS

Olá!

Esta pesquisa é sobre as competências do Cidadão Cientista de Dados (CCD), consideradas importantes pelos próprios profissionais atuantes nesse papel, que surgiu para tratar a falta de profissionais formados e especializados de Ciência de Dados.

Nos últimos anos, em face da alta demanda atual no mercado, as organizações precisaram capacitar rapidamente os seus próprios profissionais experientes no negócio, desenvolvendo competências mínimas necessárias para atuação no papel de CCD, como conhecimentos analíticos e estatísticos, não tão aprofundados quanto dos Cientistas de Dados, mas o suficiente para manusear softwares especializados e fazer análises preditivas e descritivas preliminares, que proporcionem insights para a sua área de atuação.

Portanto, essa pesquisa deve ser respondida por profissionais que atuam ou já atuaram em atividades relacionadas ao papel de CCD.

Cabe destacar que todos os dados serão usados unicamente para fins acadêmicos, protegidos por sigilo ético e os participantes não serão identificados. A participação consistirá em responder a um questionário eletrônico, com tempo médio de preenchimento de 05 minutos.

Os responsáveis por essa pesquisa são a Profa. Dra. Josiane Brietzke Porto e a aluna de graduação Nayara Neves, do curso de Administração, da UNISINOS. Estabelecem o compromisso de esclarecer quaisquer dúvidas e demais informações, que sejam necessárias no momento do preenchimento deste questionário ou posteriormente, através do e-mail nayaraneves91@gmail.com.

Após ter sido devidamente informado(a) dos aspectos relacionados à pesquisa, por favor, para prosseguir, é necessário que você indique a concordância em colaborar com a pesquisa.

1. Diante do exposto acima: *

- Concordo em participar da pesquisa
- Não concordo em participar da pesquisa

2. Qual o grau de importância dos conhecimentos listados para o papel de Cidadão Cientista de Dados?

0: Sem condições de opinar; 1: Sem importância; 2: Pouco importante; 3: Razoavelmente importante; 4: Importante; 5: Muito importante.

	0	1	2	3	4	5
Ter conhecimento multidisciplinar: dominar diferentes disciplinas e áreas de conhecimento relacionadas (Estatística, C. Computação, Matemática etc.)						
Ter conhecimento na área do negócio: união de conhecimento e de experiência na área do negócio; incluindo conhecimento administrativo no negócio.						
Ter conhecimentos de <i>software</i> , específicos de extração e análise de dados.						
Ter conhecimentos básicos relacionados à Ciência de Dados: conhecimentos relacionados à Inteligência Artificial, <i>Deep Learning</i> , entre outros.						
Ter conhecimentos técnicos suficientes para uso das ferramentas específicas de Ciências de Dados.						

Ter visão sistêmica e de valor agregado ao negócio.						
---	--	--	--	--	--	--

3. Qual o grau de importância dos habilidades listados para o papel de Cidadão Cientista de Dados?

0: Sem condições de opinar; 1: Sem importância; 2:Pouco importante; 3: Razoavelmente importante; 4: Importante; 5: Muito importante.

	0	1	2	3	4	5
Saber se comunicar: fazer com que todos os envolvidos consigam entender você.						
Gerir e liderar pessoas: contatar e liderar profissionais de dados.						
Prestar apoio a melhoria de modelos matemáticos já existentes.						
Realizar análise de dados bastante completas, precisa ter fortes habilidades analíticas, sendo capaz de analisar as informações de uma forma significativa, no contexto do negócio.						
Unir processos de negócio, análise e interpretação de dados.						

4. Qual o grau de importância das atitudes listados para o papel de Cidadão Cientista de Dados?

0: Sem condições de opinar; 1: Sem importância; 2:Pouco importante; 3: Razoavelmente importante; 4: Importante; 5: Muito importante.

	0	1	2	3	4	5
Ser criativo; propor soluções criativas, práticas, racionais e econômicas.						

Ser analítico: capaz de analisar e avaliar informações de forma significativa.						
Agir de forma alinhada à estratégia da organização.						
Ser crítico: manter pensamento crítico, organização pessoal e responsabilidade.						
Estar alinhado as estratégias da organização.						

5. Quais conhecimentos, habilidades e/ou atitudes você já possuía antes de iniciar o desenvolvimento do papel de Cidadão Cientista de Dados?

0: Nenhum; 1: Muito pouco; 2: Pouco; 3: Intermediário; 4: Alto; 5: Muito alto

	0	1	2	3	4	5
Saber se comunicar: fazer com que todos os envolvidos consigam entender você.						
Gerir e liderar pessoas: contatar e liderar profissionais de dados.						
Prestar apoio a melhoria de modelos matemáticos já existentes.						
Realizar análise de dados bastante completas, precisa ter fortes habilidades analíticas, sendo capaz de analisar as informações de uma forma significativa, no contexto do negócio.						
Unir processos de negócio, análise e interpretação de dados.						
Saber se comunicar: fazer com que todos os envolvidos consigam entender você.						
Gerir e liderar pessoas: contatar e liderar profissionais de dados.						

Prestar apoio a melhoria de modelos matemáticos já existentes.						
Realizar análise de dados bastante completas, precisa ter fortes habilidades analíticas, sendo capaz de analisar as informações de uma forma significativa, no contexto do negócio.						
Unir processos de negócio, análise e interpretação de dados.						
Ser criativo; propor soluções criativas, práticas, racionais e econômicas.						
Ser analítico: capaz de analisar e avaliar informações de forma significativa.						
Agir de forma alinhada à estratégia da organização.						
Ser crítico: manter pensamento crítico, organização pessoal e responsabilidade.						
Estar alinhado as estratégias da organização.						

6. Quais segmentos de mercado você possui experiência como Cidadão Cientista de Dados?

Tecnologia	
Financeiro	
Seguros	
Saúde	
Educação	
Comércio	
Governo	
Consultoria	
Indústria	
Transporte	

Serviços	
Telecomunicação	
Outra	

7. Qual seu nível de formação acadêmica?

Ensino Médio	
Superior Incompleto	
Superior Completo	
MBA ou Pós-graduação Lato Senso	
Pós-graduação Stricto Senso	

8. Qual a sua área de formação acadêmica?

Pedagogia	
Serviço Social	
Enfermagem	
Nutrição	
Psicologia	
Fisioterapia	
Odontologia	
Farmácia	
Arquitetura e Urbanismo	
Medicina	
Contabilidade	
Direito	
Direito	
Engenharia de Produção	
Engenharia Civil	
Sistemas de Informação	

Ciência da Computação	
Engenharia Mecânica	

9. Atualmente, o papel que você está desempenhando, é na sua área de formação acadêmica?

Sim	
Não	

10. Qual a posição hierárquica que ocupa em sua equipe?

Júnior	
Sênior	
Pleno	

11. Qual seu tempo de experiência no papel desempenhado?

Menos de 6 meses	
De 6 meses a 2 anos	
Entre 2 e 5 anos	
Mais de 5 anos	

12. De acordo com sua experiência, o desenvolvimento da sua papel está subordinada a qual área?

Área de Tecnologia de Informação e Comunicação	
Área de Negócio (Marketing, Logística, Faturamento etc.)	

13. Qual seu gênero?

Prefiro não responder	
Feminino	
Masculino	
Outro	

14. Qual a sua faixa etária?

Menos de 25 anos	
25 a 34 anos	
35 a 44 anos	
45 a 54 anos	
55 a 64 anos	
65 anos ou mais	

15. Por favor, adicione seu nome e e-mail, para que possa ser enviado o resultados desta pesquisa.