



# **CIÊNCIAS EMPRESARIAIS**

ESCOLA SUPERIOR  
POLITÉCNICO SETÚBAL

PATRÍCIA DA  
SILVA ALVES  
SCHÜTZ

## **APLICAÇÃO DA CIÊNCIA DE DADOS À INVESTIGAÇÃO COMPORTAMENTAL EM INQUÉRITOS ORGANIZACIONAIS**

Relatório de Projeto de investigação do Mestrado  
em Ciência de Dados para Empresas

### **ORIENTADORES**

Professor Doutor Pedro Fernandes da Anunciação  
Professora Doutora Ana de Jesus Mendes

Dezembro de 2025

PATRÍCIA DA  
SILVA ALVES  
SCHÜTZ

**APLICAÇÃO DA CIÊNCIA DE  
DADOS À INVESTIGAÇÃO  
COMPORTAMENTAL EM  
INQUÉRITOS ORGANIZACIONAIS**

**JÚRI**

*Presidente:* Prof. Adjunto David Alexandre Mendes Silva Simões

*Orientador:* Prof. Coordenador Pedro Fernandes da Anunciação

*Vogal:* Prof. Coordenador Sandra Cristina Dias Nunes

Dezembro de 2025

## **Abstract**

This study analyses the application of Behavioral Data Science to the field of Diversity, Equity and Inclusion (DEI), based on a survey conducted by the Cloo Behavioral Insights Unit. The questionnaire, administered in 2022 to 1,500 participants from all regions of Brazil, gathered perceptions on inclusion, organizational support, and microaggressions in the workplace context. Using a post-hoc quantitative approach, the latent structure of responses was explored through Principal Component Analysis (PCA) and segmentation via *k*-means clustering, enabling the identification of groups with distinct patterns of perception. The information sources included the survey data and scientific literature in the fields of DEI and data analysis. The results reveal relevant differences in reported experiences across the identified groups, suggesting that variables such as gender and race may be associated with perceptions of inclusion and exclusion. The study resulted in the development of the CLOO-DEI Index, which enables organizations to assess their relative positioning in relation to perceptions of inclusion within the DEI domain. The project also contributes to the creation of strategic recommendations and the redesign of a training offer grounded in empirical evidence.

**Keywords:** Behavioral Data Science, Diversity, Equity and Inclusion (DEI), Behavioral Segmentation, Organisational Inclusion Index

## Sumário

Resumo.....	i
Abstract .....	ii
Lista de Abreviaturas.....	v
Lista de Figuras .....	vi
Introdução.....	1
Capítulo 1 – Revisão de Literatura.....	5
1.1 Ciência de Dados Comportamentais: Fundamentos, Estrutura e Aplicação Organizacional .....	5
1.2 Definição e tipos de dados comportamentais .....	5
1.3 Desafios metodológicos na análise de dados comportamentais.....	6
1.4 Análise de componentes principais (ACP) na identificação de padrões latentes.....	8
1.5 Segmentação e definição de grupos .....	10
1.6 O Método <i>K-Means</i> .....	10
1.7 Integração entre ACP e segmentação .....	11
1.8 Implicações para a Personalização de Produtos e Serviços .....	11
Capítulo 2 – Âmbito e Metodologia.....	13
2.1 Âmbito.....	13
2.2 Metodologia .....	14
2.2.1 Preparação dos Dados .....	15
2.2.4 Modelação: Análise de Componentes Principais (ACP).....	16
2.2.4.1 Composição dos Componentes Principais .....	17
2.2.5 Modelação: segmentação por <i>K-Means</i> .....	18
2.2.6 Avaliação da qualidade da segmentação .....	20
2.5 Considerações éticas e reprodutibilidade .....	22
Capítulo 3 – Análise dos Resultados .....	23

3.1 Descrição demográfica da amostra .....	23
3.2 Descrição demográfica dos agrupamentos formados.....	24
3.3 Resultados da segmentação a partir da estrutura do questionário .....	25
3.4 Análise dos grupos a partir dos componentes principais (CP).....	29
3.4 Recomendações metodológicas para futuros inquéritos .....	32
3.4.1 Avaliação da consistência interna e validação da estrutura latente .....	32
3.4.2 Modularização do inquérito e normalização das escalas.....	33
3.5 Redefinição de Produtos Oferecidos e Posicionamento.....	33
3.5.1 Redefinição da oferta formativa: formações orientadas por evidências.....	33
3.5.2 Produtos analíticos e leitura integrada (medir, comparar, melhorar) .....	34
3.5.3 Posicionamento e referência estratégica e científica.....	35
3.6 Limitações e caminhos para investigações futuras .....	35
3.6.1 Estrutura e qualidade dos dados .....	35
3.6.2 Redução da dimensionalidade e segmentação.....	36
3.6.3 Representatividade e generalização.....	36
3.6.4 Avaliação psicométrica do instrumento.....	36
Conclusão e Investigação Futura .....	38
Referências Bibliográficas .....	40
Apêndice I - Descrição das questões originais e dimensões .....	45
Apêndice II - Tabela ANOVA questões mais determinantes para a formação dos grupos.....	477
Apêndice III Distribuição por género × etnia nas quatro dimensões .....	499

## **Lista de Abreviaturas**

**AA** – Análise Arquetípica

**ACP** – Análise de Componentes Principais

**CP** – Componentes Principais

**CRISP-DM** – *Cross-Industry Standard Process for Data Mining*

**DEI** – Diversidade, equidade e Inclusão

**QIQFI** – Questionário de Investigação Quantitativa sobre Fatores de Inclusão

## Lista de Figuras

Figura 1 – <i>Curva do Cotovelo</i> vs Número de <i>Clusters</i> (k).....	19
Figura 2 – Curva <i>Silhouette Score</i> .....	19
Figura 3 – Espaço bidimensional das suas primeiras componentes principais .....	20
Figura 4 – Distribuição das respostas do índice microagressões por grupo .....	27
Figura 5 – Indicadores estratégicos por perfil de experiência .....	32

## Introdução

A diversidade nas organizações tem sido amplamente reconhecida como um fator estratégico, tanto pela sua importância ética e social (Mor-Barak, 2014), como pelo seu potencial contributivo para vantagens competitivas e a capacidade de inovação (Cox & Blake, 1991; Mor-Barak, 2014; Shore et al., 2011). Contudo, transformar dados comportamentais e informações sobre diversidade organizacional em recomendações concretas continua a apresentar desafios. Neste contexto, a Ciência de Dados revela-se particularmente promissora, pois permite identificar padrões latentes e gerar inferências a partir de dados empíricos, recorrendo a técnicas estatísticas avançadas e com suporte interdisciplinar (Defossez et al., 2020; Iodice D’Enza et al., 2024).

O presente relatório insere-se no âmbito do Mestrado em Ciência de Dados para Empresas do Instituto Politécnico de Setúbal e resulta de uma colaboração com a CLOO *Behavioral Insights Unit*, uma consultora fundada em 2017 no Porto que aplica metodologias de inteligência científica e ciências comportamentais em empresas e políticas públicas, desenvolvendo soluções baseadas em evidências para promover mudanças sustentadas de comportamento (Matos et al., 2021).

Num cenário organizacional em que a diversidade é amplamente promovida como vector de inovação e vantagem competitiva (Ketemaw et al., 2024; Saxena et al., 2023, Cox & Blake, 1991), persistem desafios na tradução desses valores em programas de formação e na prática de consultoria. A literatura em Ética Empresarial evidencia um impasse entre a perspetiva utilitarista, que justifica a diversidade pelos seus ganhos económicos imediatos, e a abordagem deontológica, que a defende como imperativo moral, o que pode comprometer a sustentabilidade das iniciativas de diversidade, equidade e inclusão (DEI) (van Dijk et al., 2012). Como alternativa, propõe-se uma abordagem baseada em virtudes tais como empatia, coragem e justiça, bem como em valores organizacionais partilhados, capazes de consolidar compromissos com a inclusão no longo prazo (van Dijk et al., 2012). Para colmatar esta lacuna entre os princípios éticos e sua implementação prática, este estudo recorre ao Questionário de Investigação Quantitativa sobre Fatores de Inclusão (QIQFI), com o objetivo de identificar *insights* estratégicos que possam orientar uma reformulação da oferta formativa e definição de recomendações estratégicas baseada em dados.

A empresa disponibilizou uma base de dados recolhida em 2022, em parceria com uma organização cliente. A base é composta por respostas de 1.500 participantes provenientes de todas as regiões do Brasil. O inquérito tinha como objetivo original identificar fatores que dificultam ou facilitam a inserção, inclusão e retenção de pessoas negras em cargos de liderança. Apesar do tema de fundo ser a diversidade, equidade e inclusão (DEI), o foco deste projeto não reside na análise temática da desigualdade, objeto de análise já realizada pela empresa oportunamente, mas sim na forma como a Ciência de Dados pode ser aplicada para gerar *insights* práticos e valor estratégico para a organização a nível da avaliação da sua oferta de serviço relacionado ao tema de DEI.

A pergunta de partida que norteia este projeto é: **De que forma os padrões emergentes de microagressões, inclusão e suporte podem orientar as estratégias de consultoria da CLOO e o respetivo redesenho da oferta formativa?**

### **Objetivo Geral**

Desenvolver diretrizes para o redesenho da oferta formativa e recomendações estratégicas que tornem os serviços de consultoria da CLOO mais eficazes, com base nos padrões emergentes de microagressões, inclusão e suporte identificados no Questionário de Investigação Quantitativa sobre Fatores de Inclusão (QIQFI)

### **Objetivos Específicos**

1. Explorar, de forma aprofundada, as respostas ao QIQFI relativas a microagressões, inclusão e suporte, de modo a identificar informações relevantes que fundamentem o redesenho da oferta formativa.
2. Avaliar criticamente o QIQFI com base na estrutura e consistência das variáveis recolhidas;
3. Apresentar recomendações práticas para futuras investigações e intervenções da empresa, baseadas nos resultados obtidos;
4. Evidenciar a aplicabilidade das técnicas de Ciência de Dados Comportamentais no contexto de problemas organizacionais complexos.

O projeto seguiu uma abordagem quantitativa, retrospectiva e exploratória, fundamentada em métodos de ciência de dados, orientada pelo modelo CRISP-DM (*Cross-Industry Standard*

*Process for Data Mining*), amplamente reconhecido pela sua flexibilidade e adaptabilidade a projetos aplicados (Shearer C, 2000). A adoção deste modelo visou assegurar rigor metodológico, alinhamento com as boas práticas da ciência de dados e transparência no processo analítico.

A natureza dos dados, principalmente oriundos de escalas tipo *Likert*, favoreceu a aplicação de técnicas estatísticas que permitissem tanto a redução da dimensionalidade quanto a identificação de padrões subjacentes. Assim, optou-se pela utilização da Análise de Componentes Principais (ACP), com o objetivo de sintetizar a variabilidade dos dados em um número menor de dimensões, e da segmentação por *k-means*, como estratégia de segmentação dos participantes com base em perfis comportamentais latentes. Esta escolha metodológica não só está alinhada com a literatura em Ciência de Dados, como também permite uma exploração mais aprofundada dos dados já disponíveis, respondendo ao desafio prático de extrair valor estratégico de um *corpus* previamente recolhido.

A análise foi conduzida segundo o modelo CRISP-DM, englobando as fases de compreensão do negócio e dos dados, preparação dos dados (anonimização, codificação e padronização), modelação, avaliação interpretativa dos segmentos identificados e comunicação dos resultados através de recomendações práticas direcionadas aos produtos e processos da CLOO.

O relatório encontra-se estruturado em cinco capítulos centrais. Após esta Introdução, segue-se o capítulo 1, revisão de literatura, onde são abordados conceitos-chave relacionados com a Ciência de Dados Comportamentais, as principais características dos dados comportamentais provenientes de escalas tipo *Likert*, e uma revisão das técnicas estatísticas usadas neste estudo (nomeadamente ACP e segmentação por *k-means*), bem como uma análise da relação entre diversidade, inclusão e *performance* organizacional. No capítulo 3, Âmbito e Metodologia, é detalhada a abordagem metodológica adotada, desde a apresentação conceptual do modelo CRISP-DM até às especificidades do tratamento analítico, incluindo preparação dos dados, métodos estatísticos utilizados, critérios de validação dos resultados e preocupações éticas e de reprodutibilidade. O capítulo 4, resultados e discussão, apresenta e interpreta os resultados obtidos através das análises quantitativas realizadas. Descrevem-se os perfis comportamentais identificados pela segmentação e discutem-se as suas implicações diretas para o redesenho da oferta formativa e para a construção de recomendações estratégicas para a CLOO. Ainda neste capítulo, são abordadas limitações inerentes ao estudo e indicadas perspectivas para pesquisas futuras. Finalmente, o capítulo de conclusões e investigação futura sintetiza as principais

contribuições do projeto, reforça o papel estratégico da Ciência de Dados para as organizações e indica caminhos para implementação prática e futuras investigações.

Para além da contribuição aplicada à realidade da CLOO, este relatório pretende evidenciar o papel do cientista de dados como parceiro estratégico na análise e desenvolvimento de pesquisas comportamentais. O cruzamento entre métodos quantitativos, análise exploratória e sensibilidade contextual permite extrair valor de dados que, por vezes, permanecem subutilizados em abordagens tradicionais. Essa abordagem não apenas valoriza os investimentos prévios em recolha de dados, como também demonstra a capacidade dos cientistas de dados de atuarem de forma transversal em equipas interdisciplinares.

## Capítulo 1 – Revisão de Literatura

O presente capítulo tem como propósito apresentar o enquadramento teórico e metodológico que sustenta esta investigação, explorando os principais conceitos, autores e abordagens que integram o campo da Ciência de Dados Comportamentais. Procura-se discutir de que modo a convergência entre as ciências de dados e as ciências comportamentais tem permitido novas formas de compreender, modelar e prever o comportamento humano em contextos organizacionais. A revisão contempla, ainda, uma análise das principais metodologias utilizadas para a identificação de padrões latentes e para a segmentação de grupos comportamentais, culminando na reflexão sobre as suas implicações práticas para a personalização de produtos e serviços.

### 1.1 Ciência de Dados Comportamentais: Fundamentos, Estrutura e Aplicação Organizacional

A Ciência de Dados Comportamentais constitui um campo emergente cujo objetivo geral é a análise de dados relativos a diferentes aspetos do comportamento humano. Este empreendimento é simultaneamente valioso e desafiador. Contrariamente a uma visão simplificada, que poderia reduzi-la a uma mera aplicação de métodos estatísticos a dados de comportamento, a Ciência de Dados Comportamentais é, por sua natureza, profundamente interdisciplinar. Ela emerge da intersecção de campos tradicionalmente distintos: a Ciência de Dados, que engloba estatística, *machine learning* e ciência da computação, e as Ciências Comportamentais, que variam desde economia e finanças até psicologia, sociologia e comportamentos relacionados com a saúde (Iodice D’Enza et al., 2024).

Esta natureza interdisciplinar impõe um olhar facetado sobre a análise, onde o conhecimento específico do domínio de aplicação não é meramente complementar, mas sim mandatório, devendo guiar a escolha e o desenvolvimento de ferramentas metodológicas. A integração do conhecimento específico do domínio é essencial em todas as fases do processo de análise, desde o pré-processamento e *feature engineering* até à interpretação dos resultados.

### 1.2 Definição e tipos de dados comportamentais

Os dados comportamentais são intrinsecamente diversos, refletindo a multiplicidade de formas pelas quais o comportamento pode ser medido. Podem ser classificados a partir de diferentes perspetivas. Um dos sistemas de classificação utilizado nas ciências cognitivas é o proposto por Raymond B. Cattell (1950), que distingue três tipos principais:

- **Dados Q (Questionários):** Referem-se a dados de autorrelato (*self-report*). Baseiam-se em como os indivíduos avaliam os próprios comportamentos, opiniões ou sentimentos. Frequentemente recolhidos através de questionários ou escalas *Likert*. A validade e utilidade destes dados podem ser condicionadas por fatores como estilos de resposta, intenção de partilhar informação, motivação e capacidade de autorreflexão do indivíduo. Apesar destas limitações, os autorrelatos são considerados ferramentas úteis e preditivas sob certas condições de aplicação. A questão central reside na fiabilidade da autodescrição (Cattell, 1950).
- **Dados T (Testes experimentais):** provêm de paradigmas experimentais, como as tarefas *Stroop* (onde o tempo de reação pode indicar controlo cognitivo) ou *Go/No-go* (onde os erros refletem falhas na inibição de resposta), permitem aceder a medidas indiretas de funcionamento cognitivo e emocional, oferecendo maior controlo experimental e potencial de inferência causal (Cattell, 1950).
- **Dados L (Registos de vida):** englobam dados objetivos extraídos de fontes como histórico académico, registos clínicos, plataformas digitais ou sensores, sendo úteis para análises longitudinais e de comportamento naturalista (Cattell, 1950; Iodice D’Enza et al., 2024).

De forma complementar, Defosse et al. (2020) propõem uma estrutura mais detalhada voltada para dados digitais oriundos de experiências informatizadas. A sua proposta, o *Behavense Data Model* (BDM), organiza os dados comportamentais em níveis:

- **Dados de Evento:** representam comportamentos discretos (ex.: cliques, movimentos, respostas) com marcação temporal.
- **Dados de Tentativa:** agregam eventos por unidade de análise (ex.: cada linha representa uma tentativa de tarefa) com base em critérios experimentais.
- **Dados L1 a L4:** representam diferentes níveis de agregação e transformação dos dados originais, facilitando a normalização e a análise estatística avançada.

Esses sistemas evidenciam a complexidade da estruturação dos dados comportamentais e reforçam a importância de padronização para permitir reprodutibilidade e análise comparável entre estudos.

### 1.3 Desafios metodológicos na análise de dados comportamentais

A análise de dados comportamentais enfrenta desafios que vão além da mera aplicação de algoritmos. Um dos maiores obstáculos é a falta de padrões amplamente adotados para formatar,

nomear, organizar, descrever ou partilhar dados comportamentais (Defossez et al., 2020). Esta inconsistência dificulta a reutilização eficaz dos dados e impede o desenvolvimento de ferramentas analíticas eficientes, além disso, o estado atual da padronização estrutural dos dados comportamentais é considerado muito baixo (níveis 0 a 1 numa escala de 7) e sua interpretação requer uma análise no contexto em que ocorreram, o que implica anotações sobre a tarefa, o estímulo e o estado do indivíduo. No entanto, a falta de clareza conceptual em termos fundamentais como "instruções", "*feedback*" ou "tentativa" complica a conversão desta informação contextual em estruturas de dados utilizáveis (Defossez et al., 2020).

No âmbito da ciência de dados comportamentais ganham relevo a explicabilidade dos modelos e a integração de conhecimento do domínio em todas as etapas do processo analítico. Quando o objetivo é compreender o comportamento humano, modelos de "caixa negra" (*black boxes*) são menos desejáveis ou confiáveis (Iodice, 2024). É o conhecimento substantivo que permite contextualizar eventos, dar significado a registos e interpretar padrões identificados por algoritmos (Iodice D'Enza et al., 2024; Defossez et al., 2020).

Acrescem ainda implicações éticas relevantes, particularmente quando envolvem identidades sociais, vieses algorítmicos e decisões com impacto real sobre pessoas (como recrutamento ou avaliações de desempenho) (Varsha P.S., 2023). Por isso, a importância em garantir não apenas a segurança e anonimização dos dados, mas também o uso responsável e transparente das informações produzidas.

A diversidade de tipos de dados amplia estes desafios: (i) estruturados, organizados em tabelas com atributos uniformes; (ii) semiestruturados, como registos de sensores ou *logs* de utilização; e (iii) não estruturados, como texto, imagem ou vídeo, que exigem pré-tratamento e seleção de atributos antes de integrações analíticas (Kelleher & Tierney, 2018; Asfahani, 2024). Em domínios como gestão de talentos, educação ou comunicação organizacional, a análise integrada destas fontes é necessária para compreender perceções, interações e níveis de conhecimento, e para adaptar intervenções (Iodice D'Enza et al., 2024; Asfahani, 2024).

Para além da heterogeneidade, muitos conjuntos apresentam elevada dimensionalidade (grande número de variáveis, frequentemente correlacionadas), o que dificulta a visualização e a identificação de regularidades, fenómeno associado à maldição da dimensionalidade (James et al., 2023; Xu & Wunsch, 2005). Perante este cenário, torna-se justificada a redução da dimensionalidade, que sintetiza a variabilidade essencial por meio de componentes principais, cujos *scores* correspondem a combinações lineares ortogonais das variáveis originais que, por

ordem, maximizam a variância explicada (Jolliffe & Cadima, 2016; Bro & Smilde, 2014) e ainda a segmentação (*clustering*), que particiona observações em grupos relativamente homogêneos com base em múltiplas variáveis inter-relacionadas (Defossez et al., 2020; Asfahani, 2024). A secção seguinte apresenta a Análise de Componentes Principais (ACP) como etapa de síntese informacional e, subsequentemente, a segmentação (com destaque para o algoritmo *k-means*) como procedimento para revelar padrões e perfis latentes relevantes para a personalização de intervenções.

Nesta dissertação, a análise quantitativa dialoga com enquadramentos conceptuais das ciências comportamentais usados para explicar participação, segurança para falar, clima de inclusão e estruturas de rede em equipas. A literatura indica que a inclusão efetiva resulta da combinação entre pertença e unicidade (modelo de inclusão em grupos) (Shore et al., 2011), da segurança psicológica para assumir riscos interpessoais (Edmondson, 1999), da qualidade da relação líder–membro (LMX), que condiciona acesso a informação, patrocínio e oportunidades (Graen & Uhl-Bien, 1995; Dulebohn et al., 2012), e de processos culturais que alinham valores apregoados e práticas (Schein, 2010/2017). Adicionalmente, a estrutura das redes de trabalho molda o acesso a recursos e a influência em equipas (Borgatti & Halgin, 2011). Estes enquadramentos informam a leitura dos componentes extraídos por ACP e das segmentações obtidas.

#### **1.4 Análise de componentes principais (ACP) na identificação de padrões latentes**

Quando um conjunto de dados contém um grande número de variáveis, diz-se que possui elevada dimensionalidade. Este facto, sobretudo quando as variáveis estão correlacionadas, dificulta a visualização e a interpretação dos padrões subjacentes (James et al., 2023). Para ultrapassar esta limitação, recorrem-se a técnicas de redução de dimensionalidade, que consiste em transformar o conjunto inicial num número menor de variáveis sintéticas, capazes de reter a maior parte da variabilidade estatística presente nos dados originais. Entre diversas técnicas de redução de dimensionalidade, destaca-se a Análise de Componentes Principais (ACP), uma abordagem estatística multivariada e amplamente utilizada. O seu propósito é transformar o conjunto original de variáveis iniciais. Estes componentes sintetizam a informação de modo a reter a maior parte da variabilidade presentes nos dados, permitindo uma representação mais compacta e interpretável do fenómeno estudado (Bro & Smilde, 2014).

Em essência, a ACP transforma um conjunto de variáveis potencialmente correlacionadas num conjunto mais pequeno de variáveis não correlacionadas (os componentes principais), ordenadas pela quantidade de variação que explicam (Bro & Smilde, 2014). Esta capacidade de

resumir a informação presente em múltiplas variáveis inter-relacionadas torna a ACP particularmente útil para identificar padrões ou dimensões latentes nos dados comportamentais ou de perceção. Os resultados de uma análise ACP podem ser visualizados e interpretados através de:

- *Scores* (Pontuações): representam as coordenadas de cada observação (indivíduo, país, etc.) no espaço definido pelos componentes principais (Bro & Smilde, 2014), (James et al., 2013). A análise dos *scores* permite visualizar as relações entre as observações e detetar potenciais agrupamentos ou padrões (Bro & Smilde, 2014). Por exemplo, no estudo dos comportamentos alimentares, as pontuações nos componentes principais poderiam revelar como diferentes países se posicionam em relação aos padrões gerais de aprovisionamento alimentar (Giordani & Kiers, 2023)
- *Loadings* (vetores de peso): representam a contribuição de cada variável original para a formação de cada componente principal. A análise destes coeficientes permite compreender o significado de cada componente, isto é, identificar a combinação linear das variáveis que a define, possibilitando a interpretação das dimensões latentes reveladas pela análise (Bro & Smilde, 2014).
- *Biplots*: Combinam *scores* e *loadings* numa única visualização, facilitando a interpretação conjunta das relações entre observações e variáveis (Bro & Smilde, 2014).

A ACP é uma técnica estatística multivariada que fornece uma descrição mais parcimoniosa da variabilidade dos dados (Bro & Smilde, 2014). Diferentemente dos métodos de *clustering*, que agrupam unidades observadas em função da proximidade a um centro representativo, como os centroides no caso do *k-means*, (Giordani & Kiers, 2023), a ACP procura identificar eixos ortogonais que maximizam a variância dos dados, a denominada *principal axis property* (Bro & Smilde, 2014).

Uma aplicação particularmente relevante da ACP no contexto da segmentação é o seu uso como etapa de pré-processamento para outras técnicas analíticas. Ao reduzir o número de variáveis originais a um conjunto menor de *scores* de componentes principais, que capturam a variação essencial, a ACP gera uma representação mais compacta e menos ruidosa dos dados. Essa redução não só facilita a visualização e interpretação, como também melhora a eficiência computacional e reduz a possibilidade de sobreajustamento em análises subsequentes (Jolliffe & Cadima, 2016). A literatura em análise multivariada mostra que esta abordagem é frequentemente utilizada antes da aplicação de métodos de agrupamento, como o *k-means* ou a

análise hierárquica, permitindo formar *clusters* mais estáveis e interpretáveis (Everitt et al., 2011; Xu & Wunsch, 2005). A integração de métodos evidencia que a ACP não é apenas uma técnica descritiva, mas desempenha igualmente um papel instrumental em processos analíticos multietapas.

### **1.5 Segmentação e definição de grupos**

A segmentação, também designada por análise de agrupamento (*clustering*), corresponde a um conjunto de métodos de aprendizagem não supervisionada que visam identificar subgrupos (*clusters*) latentes num conjunto de dados, sem recurso a uma variável-resposta pré-definida (Hastie, Tibshirani & Friedman, 2009; Everitt et al., 2011). No contexto organizacional, estes métodos são particularmente úteis para compreender padrões de perceção, atitudes ou experiências partilhadas entre indivíduos, apoiando a personalização de intervenções, produtos formativos ou estratégias de comunicação (James et al., 2023).

Em termos técnicos, a segmentação procura particionar observações em grupos mutuamente exclusivos, de forma que as unidades dentro de cada grupo sejam tão semelhantes quanto possível e, simultaneamente, tão distintas quanto possível das pertencentes a outros grupos (Kaufman & Rousseeuw, 2005; Xu & Wunsch, 2005). Essa semelhança ou dissemelhança é geralmente medida através de métricas de distância, como a distância Euclidiana, embora a escolha do critério dependa do domínio de aplicação e da natureza dos dados (Everitt et al., 2011). Ao estruturar dessa forma, a segmentação permite revelar padrões comportamentais, atitudes ou perfis latentes de resposta que dificilmente seriam identificados por meio de análises univariadas ou por inspeção direta dos dados (Kaufman & Rousseeuw, 2005).

### **1.6 O Método *K-Means***

Entre os métodos de segmentação, o algoritmo *K-means* é amplamente utilizado devido à sua relativa simplicidade e eficiência computacional (Jain, 2010; Everitt et al., 2011). O seu objetivo central consiste em minimizar a variabilidade intragrupo, isto é, agrupar as observações de forma que a distância entre cada ponto e o centroide do seu respetivo grupo seja mínima.

O algoritmo pode ser descrito com os seguintes passos:

1. Inicialização: cada observação é atribuída aleatoriamente a um dos *K clusters*;
2. Cálculo dos centroides: determina-se para cada *cluster*, o centroide, correspondente à média das observações em todas as dimensões;

3. Reatribuição: cada observação é associada ao grupo cujo centroide se encontra mais próximo, de acordo com uma métrica de distância (tipicamente Euclidiana);
4. Iteração: Os passos 2 e 3 são repetidos até que as atribuições deixem de mudar, atingindo uma solução de mínimo (Lloyd, 1982; James et al., 2023).

Este processo é sensível à escolha inicial dos centroides, motivo pelo qual se recomenda a execução múltipla do algoritmo com inicializações distintas.

### **1.7 Integração entre ACP e segmentação**

A integração entre Análise de Componentes Principais (ACP) e segmentação visa sintetizar informação distribuída por múltiplas variáveis correlacionadas e, em seguida, identificar perfis latentes num espaço de menor dimensionalidade. A ACP gera *scores* ortogonais que retêm a variabilidade estatisticamente relevante e facilitam a interpretação; a segmentação — neste estudo, por *k-means* — particiona as observações com base nesses *scores*, produzindo grupos legíveis para fins analíticos e de decisão (Jolliffe & Cadima, 2016; Everitt et al., 2011; Xu & Wunsch, 2005; Jain, 2010; James et al., 2023). Esta estratégia é particularmente adequada quando existe elevada dimensionalidade e correlação entre variáveis. Os procedimentos adotados encontram-se detalhados na Metodologia (Secções 2.2.4–2.2.6).

### **1.8 Implicações para a Personalização de Produtos e Serviços**

A identificação de perfis latentes por meio da integração entre ACP e de técnicas de segmentação permite às organizações conceber soluções mais eficazes e ajustadas às necessidades reais dos diferentes segmentos. No contexto de produtos formativos ou estratégias de comunicação:

- Evidencia estruturas latentes: a ACP sintetiza variáveis correlacionadas em componentes ortogonais que captam a variabilidade estatisticamente relevante, permitindo reconhecer eixos interpretáveis que sustentam a formação de perfis (Bro & Smilde, 2014; Jolliffe & Cadima, 2016)
- Melhora a qualidade da segmentação: A segmentação em *scores* de ACP tende a ser mais consistente do que aquela aplicada diretamente aos dados originais de alta dimensionalidade, os quais frequentemente são ruidosos ou apresentam variáveis redundantes. Ao condensar a informação relevante e atenuar parte do ruído, a ACP contribui para a obtenção de *clusters* mais estáveis e interpretáveis (James et al., 2023)

- Fornece perfis interpretáveis: Os *loadings* da ACP fornecem uma descrição do que cada dimensão latente representa. Ao examinar os *scores* médios de cada *cluster* nas dimensões da ACP e, posteriormente, os valores médios das variáveis originais para cada *cluster*, é possível criar perfis detalhados e interpretáveis para cada segmento de público identificado. Estes perfis descrevem as características distintivas de percepção ou comportamento de cada grupo.

Estes perfis sustentam a personalização e em termos de aplicações, esta abordagem tem sido utilizada, entre outras possibilidades, para:

- Agrupar indivíduos com diferentes níveis de conhecimento sobre um tema e desenvolver recomendações personalizadas (Iodice D’Enza et al., 2024);
- Revelar subgrupos de percepções e experiências em contextos organizacionais, fundamentais para o redesenho de programas formativos e estratégias de comunicação (Mor-Barak, 2014).

Além disso, ao associar os resultados estatísticos a conceitos comportamentais (inclusão/voz, segurança psicológica, LMX, justiça informacional e redes) tal personalização da oferta de serviços passa a ser ainda mais relevante.

## Capítulo 2 – Âmbito e Metodologia

Este capítulo tem por objetivo apresentar o enquadramento empírico e metodológico do estudo. Na primeira parte, é delineado o âmbito da investigação, contextualizando a escolha da entidade parceira, a natureza dos dados recolhidos e a pertinência do inquérito analisado. Em seguida, descrevem-se os procedimentos metodológicos adotados para o tratamento e análise dos dados, incluindo as técnicas estatísticas utilizadas e os critérios de validação das etapas analíticas. Ao articular estas duas dimensões — contexto e método — procura-se garantir transparência, rigor e coerência na construção dos resultados que serão apresentados nos capítulos seguintes.

### 2.1 Âmbito

A presente investigação desenvolve-se no contexto de uma colaboração com a CLOO *Behavioral Insights Unit*, uma organização que se distingue pela aplicação das ciências comportamentais à realidade organizacional, aliando o rigor científico à intervenção prática. Com uma atuação orientada para a mudança de comportamentos e a geração de impacto positivo na sociedade, a CLOO promove projetos que abordam problemas reais através de uma leitura sistemática e baseada em evidência do comportamento humano, apoiando empresas e instituições na construção de soluções mais humanas, éticas e eficazes.

O estudo parte da oportunidade de aplicar metodologias de ciência de dados a uma base de dados de natureza comportamental, recolhida pela CLOO no âmbito de um inquérito nacional conduzido no Brasil. Este inquérito visou mapear perceções, práticas e desafios relacionados com a diversidade, equidade e inclusão (DEI) no ambiente de trabalho. A base empírica distingue-se não apenas pela sua expressividade e diversidade amostral, mas também pela natureza sensível e estratégica dos dados, cujo desenho e recolha foram orientados por uma perspetiva científica da experiência vivida por profissionais em diferentes sectores.

A colaboração com a CLOO permitiu acesso aos dados e criou condições para a realização da análise em consonância com uma perspetiva de investigação aplicada. Este enquadramento possibilitou tanto a identificação de padrões e relações comportamentais como a elaboração de recomendações orientadas para a prática organizacional, reforçando a articulação entre investigação e impacto social.

## 2.2 Metodologia

O estudo empírico baseia-se numa base de dados recolhida em 2022, no âmbito de um inquérito desenvolvido pela CLOO *Behavioral Insights Unit*, em colaboração com uma empresa cliente. A amostra inclui 1500 participantes de todas as regiões do Brasil, abrangendo diferentes perfis sociais e profissionais, independentemente da sua situação laboral ou posição hierárquica. O principal objetivo do inquérito consistiu em identificar fatores comportamentais que dificultam a inserção, inclusão, desenvolvimento e retenção de pessoas negras em cargos de liderança (Relatório CLOO, 2023), explorando perceções e experiências vividas no interior das organizações sob a lente da diversidade racial.

A metodologia CRISP-DM (*Cross Industry Standard Process for Data Mining*) constitui um modelo de referência amplamente utilizado para orientar projetos de Ciência de Dados, estruturando-os em seis fases iterativas: compreensão do negócio (*Business Understanding*), compreensão dos dados (*Data Understanding*), preparação dos dados (*Data Preparation*), modelação (*Modeling*), avaliação (*Evaluation*) e implementação (*Deployment*) (Chapman et al., 2000; Wirth & Hipp, 2000). A adoção deste modelo justifica-se pela sua ampla aceitação e pela clareza com que orienta a transição de dados complexos para resultados aplicáveis, o que o torna especialmente adequado a investigações que pretendem gerar conhecimento orientado para a prática.

No enquadramento desta investigação, a secção que se segue corresponde à fase Compreensão do Negócio. O objetivo não é retomar a finalidade substantiva do estudo original, mas sim demonstrar o valor acrescentado que as técnicas de Ciência de Dados podem gerar em termos de *insights* comportamentais acionáveis para organizações. Para esse efeito, foram definidas três metas estratégicas alinhadas com as prioridades comerciais da CLOO:

- Apoiar o desenho e o aperfeiçoamento de produtos formativos e interventivos em DEI, recorrendo a evidência empírica extraída da base de dados existente;
- Priorizar ações de gestão de pessoas nas empresas-cliente, fornecendo segmentações estatisticamente consistentes que permitam alocar recursos de forma mais eficaz;
- Produzir evidência comercial que sustente o posicionamento da CLOO como fornecedor de soluções *data-driven* em DEI.

Com base nestes objetivos, formularam-se hipóteses de trabalho que orientam as fases subsequentes do CRISP-DM. Em particular, avalia-se como a Análise de Componentes Principais (ACP) gerando *scores* que sintetizam a variabilidade relevante e a segmentação *K*-

*means* podem revelar padrões latentes nas respostas e convertê-los em operacionais para o portfólio de produtos e serviços da CLOO em DEI. Estabelecidos os objetivos e as hipóteses, a investigação avança para a fase de Compreensão dos Dados, na qual se examinam a estrutura, a qualidade e a representatividade da base de dados

### 2.2.1 Preparação dos Dados

Inicialmente, o conjunto de dados original foi submetido a um processo de anonimização. Este procedimento visou garantir a total confidencialidade dos participantes e o estrito cumprimento das normativas éticas e de proteção de dados. Posteriormente, procedeu-se à padronização das respostas obtidas através de escalas do tipo *Likert*. Os valores nominais ou textuais foram mapeados para um intervalo numérico ordinal de 1 a 5 pontos. O tratamento das variáveis categóricas constituiu um passo subsequente envolvendo distintas estratégias de codificação, selecionadas em função da natureza intrínseca de cada variável (Kuhn & Johnson, 2013). Para as variáveis binomiais, designadamente aquelas correspondentes a respostas dicotômicas (e.g., "sim/não"), foi empregue a técnica de *one-hot encoding*. Este método consiste na criação de vetores binários (0/1) que representam univocamente cada categoria, mitigando o risco de imposição de uma relação ordinal inexistente entre elas (Hair et al., 2009)

No que concerne a variáveis ordinais, como "classe social" e "nível de escolaridade", que exibem uma hierarquia inerente entre as suas categorias, optou-se pela aplicação de uma codificação ordinal (também designada em alguns contextos por *label encoding*, desde que a atribuição numérica respeite a ordem). Esta técnica atribui um valor numérico sequencial a cada categoria, de forma a preservar a sua relação de ordem natural (e.g., "Ensino Básico" = 1, "Ensino Secundário" = 2, "Ensino Superior" = 3).

Foi adotado um tratamento específico às variáveis categóricas de seleção múltipla, em particular a autoidentificação étnica ('etnia própria') e à etnia dos progenitores. Na primeira etapa procedeu-se à binarização de rótulos múltiplos, (*one-hot encoding*), gerando novas colunas binárias (0 ou 1) para cada opção de resposta étnica disponível. Numa segunda etapa, e com o objetivo de assegurar maior coerência analítica e potencial interpretativo, as categorias étnicas resultantes desta expansão foram submetidas a um reagrupamento estratégico. Esse procedimento foi realizado para manter a coerência com o tratamento realizado pela empresa na exploração inicial dos dados realizadas logo após a execução da investigação na qual agrupou pessoas pretas e pardas em um único grupo designado negros (CLOO, 2023). Esta consolidação teve por objetivo reduzir o número de preditores e aumentar a frequência efetiva

por categoria após a codificação multirrótulo, favorecendo estimativas mais estáveis e uma interpretação substantiva mais consistente (Kuhn & Johnson, 2013).

Finalmente, no contexto da análise posterior orientada para a formação de agrupamentos (*clustering*), com particular enfoque nas variáveis relativas à diversidade e inclusão, procedeu-se à supressão seletiva de determinadas informações do questionário, nomeadamente a etnia dos progenitores. Esta decisão de seleção de atributos (*feature selection*) fundamentou-se na constatação de que as referidas variáveis não introduziam variabilidade explicativa relevante nem contribuía de forma significativa para a discriminação e interpretação dos agrupamentos resultantes. Tal procedimento visou a otimização e a parcimónia do modelo (Kuhn & Johnson, 2013).

#### **2.2.4 Modelação: Análise de Componentes Principais (ACP)**

Com o objetivo de preparar a segmentação dos inqueridos a partir das respetivas respostas, aplicou-se a Análise de Componentes Principais (ACP) com vista a reduzir dimensionalidade dos conjuntos de dados com um grande número de variáveis potencialmente correlacionadas, preservando a máxima variância possível (Giordani et al., 2023; Kuhn & Johnson, 2013). A ACP transforma o vetor original de variáveis num novo sistema de eixos ortogonais (componentes principais). Cada componente é uma combinação linear otimizada para captar, em ordem decrescente, a proporção mais elevada de variância existente, cabendo ao primeiro componente o maior contributo explicativo (Palazzo et al., 2024).

A implementação foi realizada em *Python*, recorrendo às bibliotecas *Pandas* e *NumPy* para manipulação e estruturação dos dados, e a *Scikit-learn* para standardização (*StandardScaler*) e estimação da ACP (*PCA*) sobre a matriz de correlações; *Matplotlib* e *Seaborn* foram utilizadas para inspeções e visualizações exploratórias. Os procedimentos seguiram boas práticas de reprodutibilidade (registo de versões e definição de *seed*). A variável categórica “*existiam\_liderancas\_negras*” foi excluída antes da aplicação da ACP por não ser contínua.

A retenção de componentes baseou-se na combinação do critério de Kaiser (autovalores  $> 1$ ), do *scree plot* e da variância explicada cumulativa: os cinco primeiros CPs explicaram  $\approx 63,8\%$  da variância e foram retidos para a etapa de *clustering*. Testou-se a inclusão de um 6.º CP ( $\approx 65,7\%$  cumulativo), mas a qualidade dos agrupamentos diminuiu (*silhouette*  $\approx 0,25$ ), pelo que se adotou a solução com 5 CPs (Jolliffe & Cadima, 2016; Kuhn & Johnson, 2013).

Do ponto de vista analítico, a ACP produz *loadings* (coeficientes que definem cada componente em termos das variáveis originais) e *scores* (posicionamento das observações nos eixos latentes). Os *scores* das componentes retidas serviram de base à segmentação por *k-means* (Sec. 2.2.5), enquanto a leitura dos *loadings* sustentou a interpretação das dimensões latentes. Esta estratégia condensa informação relevante e atenua redundâncias antes do agrupamento, favorecendo soluções legíveis e compatíveis com a finalidade explicativa do estudo (Everitt et al., 2011; Xu & Wunsch, 2005; James et al., 2023).

#### **2.2.4.1 Composição dos Componentes Principais**

Após a aplicação da Análise de Componentes Principais (ACP), observou-se que os cinco primeiros componentes apresentaram autovalores superiores a 1, são eles: [8,85, 2,35, 1,86, 1,27, 1,17]. Esses autovalores indicam a quantidade de variância que cada componente principal explica no conjunto de dados. De acordo com o critério de Kaiser (Kaiser, 1960), componentes com autovalores maiores que 1 devem ser retidos, pois representam mais variância que uma variável original padronizada. A variância explicada por cada componente foi, respetivamente, 38,80%, 8,66%, 7,42%, 4,66% e 4,26%, totalizando 63,80% da variância dos dados. Esse resultado é considerado aceitável no contexto de estudos com variáveis comportamentais, que frequentemente apresentam múltiplos fatores latentes e interdependentes (Giordani & Kiers, 2023; Palazzo et al., 2023).

##### **Componente 1 – Participação Ativa (38,80 %)**

Este componente é dominado pelas variáveis “*influencia\_convocado\_reunioes*” (0,247), “*influencia\_convocado\_autodesenvolvimento*” (0,247) e “*influencia\_informado\_atividades*” (0,246). Reflete um eixo de participação e inclusão nas dinâmicas organizacionais, evidenciando a abertura da organização à escuta e à presença ativa dos seus membros.

##### **Componente 2 – Perceção Emocional e Microagressões (8,66 %)**

Fortemente influenciado por “*acredito\_submetido\_desconfortaveis*” (0,372), “*percebo\_facil\_branco\_lider*” (0,359) e “*acredito\_fazer\_mais*” (0,363), este eixo traduz sentimentos subjectivos de vigilância, esforço adicional e potenciais microagressões no contexto laboral.

##### **Componente 3 – Crenças de Igualdade e Acolhimento (7,42 %)**

Destacam-se as variáveis “todas\_bemvindas” (0,425), “igualdade\_negras\_branças” (0,391) e “nao\_necessita\_medidas\_inclusivas” (0,193). Este eixo expressa crenças normativas sobre igualdade racial e um ambiente organizacional inclusivo.

#### **Componente 4 – Atitudes face a Políticas Inclusivas (4,66 %)**

Influenciado por “nao\_necessita\_medidas\_inclusivas” (0,363), “lideranças\_sensíveis\_inclusão\_racial” (0,313) e “sempre\_maximo\_reconhecido” (-0,348), este componente reflecte atitudes e juízos sobre mérito, reconhecimento e eficácia das acções afirmativas.

#### **Componente 5 – Suporte e Barreiras Relacionais (4,26 %)**

As variáveis mais relevantes são “percebo\_facil\_branco\_lider” (0,514), “facil\_branças\_integrar” (0,475) e “influencia\_raramente\_convivio” (-0,312). O PC5 capta a dimensão relativa ao suporte racial e às barreiras estruturais à ascensão de pessoas negras em ambientes organizacionais.

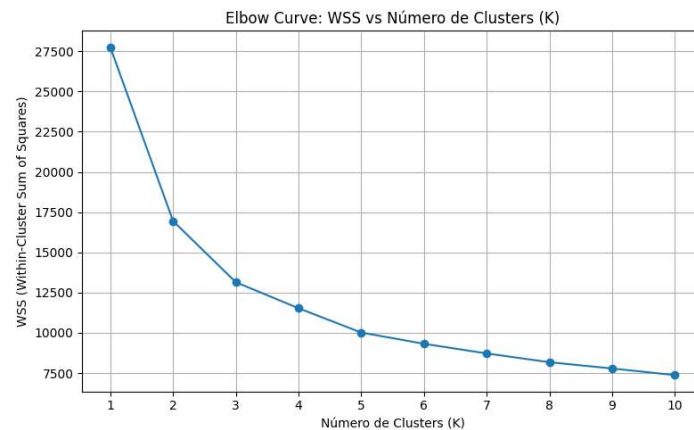
### **2.2.5 Modelação: segmentação por *K-Means***

Com o objetivo de identificar grupos latentes nas percepções dos respondentes, aplicou-se a técnica de *K-means* aos cinco primeiros *scores* da ACP. A implementação foi realizada em *Python (scikit-learn)*, com distância Euclidiana (padrão algoritmo), inicialização *k-means++*, múltiplas inicializações para mitigação de ótimos locais (*n\_init=10*), número máximo de de iterações padrão (*max\_iter=300*) e semente definida para reprodutibilidade (*random\_state=42*) (MacQueen, 1967; Lloyd, 1982; Everitt et al., 2011; Jain, 2010)

A validade interna da segmentação foi avaliada por meio de duas métricas complementares: a curva de cotovelo (*Elbow Method*), baseada na soma das distâncias quadradas intra *cluster* (*WSS – Within-Cluster Sum of Squares*), e o índice de *Silhouette*, que mede a coesão interna e separação entre grupos (Kaufman & Rousseeuw, 2005; James et al., 2023; Lloyd, 1982; Jain, 2010). Para a determinação do número de grupos (*k*), foram utilizados dois critérios complementares:

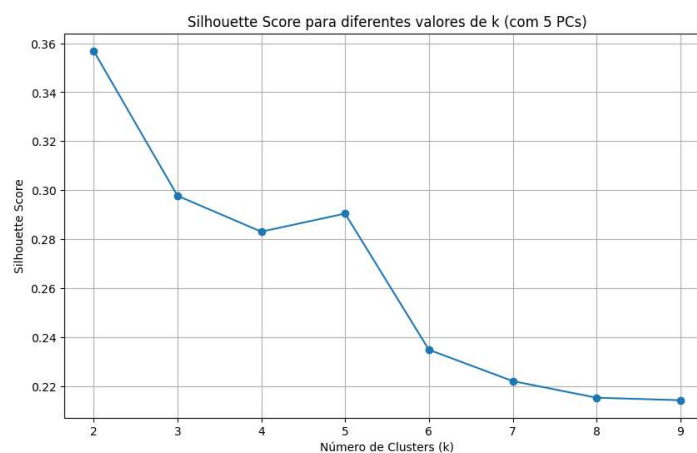
**1. Curva do Cotovelo (WSS):** Calculou-se a *within-cluster sum of squares (WSS/inertia* do *k-means*) para  $k \in [1,10]$ . Observou-se uma queda acentuada da WSS até  $k=3$ , seguida de decréscimos marginais para valores superiores (Figura 1, “Curva do Cotovelo vs Número de

Clusters (k)”. Este ponto de inflexão sugere que adicionar mais grupos traz ganhos pouco substantivos em compacidade.



**Figura 1:** Curva do Cotovelo vs Número de *Clusters* (k)

**2. Índice de *silhouette*.** Varreu-se  $k \in [2,9]$  sobre os **5 PCs**, obtendo-se  $silhouette=0,36$  para  $k=2$  (máximo) e  $silhouette=0,298$  para  $k=3$ , com declínio progressivo a partir daí (Figura 2, painel “*Silhouette*”). O valor em  $k=3$  traduz separação fraca a moderada — padrão frequente em dados comportamentais de elevada dimensionalidade e correlações moderadas (Kaufman & Rousseeuw, 2005; James et al., 2023; Xu & Wunsch, 2005).



**Figura 2:** Curva *Silhouette Score*

A escolha final por  $k=3$  resulta da convergência entre o cotovelo (compacidade) e a interpretabilidade substantiva no espaço dos componentes, aceitando uma separação moderada de acordo com as características do domínio (James et al., 2023; Asfahani, 2024). Como verificação adicional no espaço dos 5 PCs, a razão BSS/TSS foi de  $\approx 52,62\%$ , indicando que mais de metade da variabilidade projetada é explicada pelas diferenças entre segmentos, o que

é compatível com uma estrutura segmentável de magnitude intermédia. Para potenciar a clareza interpretativa, os *clusters* foram projetados no plano definido pelas duas primeiras componentes principais (PC1 x PC2), responsáveis pela maior parcela da variância explicada. A figura 3 mostra que, embora exista alguma sobreposição marginal, a configuração dos agrupamentos permanece discernível, o que é consistente com os indicadores de validade interna e apoia a opção pela segmentação. A caracterização dos centroides no espaço dos *scores* e a interpretação substantiva de cada perfil apoiada nos *loadings* das componentes principais são apresentadas na Sec. 2.2.4.1 e desenvolvidas nos Resultados (Cap. 3).

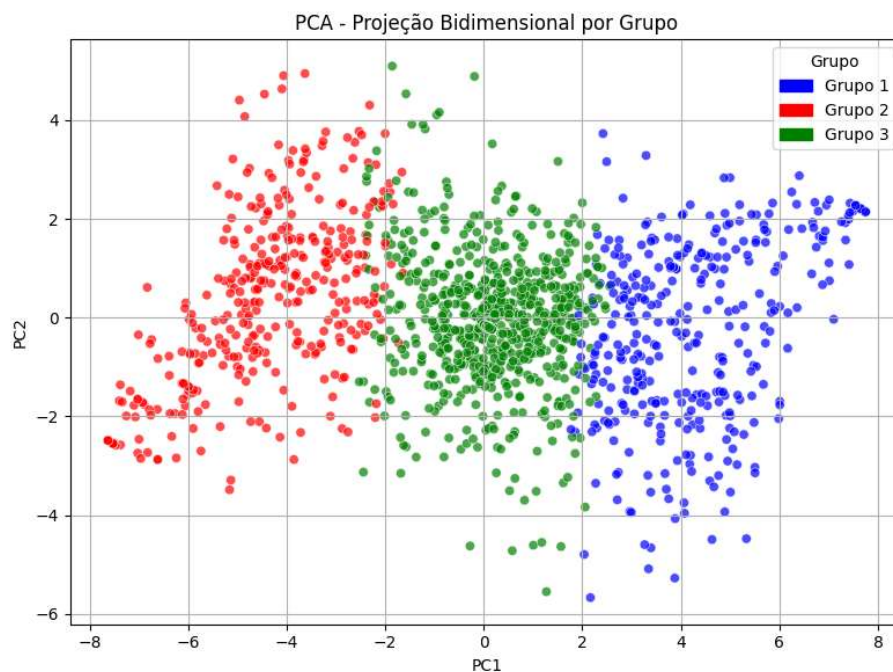


Figura 3: Espaço bidimensional das suas primeiras componentes principais

### 2.2.6 Avaliação da qualidade da segmentação

A avaliação da solução de segmentação visou dois objetivos complementares: estimar a qualidade interna dos agrupamentos obtidos por *k-means* no espaço dos cinco primeiros *scores* da ACP e verificar a diferenciação substantiva entre os grupos nas variáveis que estruturam esse espaço. Optou-se por um conjunto de métricas alinhado com a natureza do algoritmo e com a literatura de referência. Em primeiro lugar, utilizou-se a WSS (*within-cluster sum of squares*) e a correspondente curva do cotovelo para apreciar a compacidade dos grupos à medida que o número de *clusters* aumenta, uma vez que o *k-means* minimiza diretamente essa quantidade (Lloyd, 1982; Jain, 2010; Everitt et al., 2011). Em segundo lugar, recorreu-se ao índice de *silhouette*, que sintetiza, num único *score*, a coesão intra-grupo e a separação inter-grupos, permitindo comparar soluções com diferentes valores de *k*

(Kaufman & Rousseeuw, 2005; James et al., 2023). Como medida agregada adicional, calculou-se a razão BSS/TSS (variância entre grupos sobre a variância total projetada nos componentes), útil como indicador descritivo da fração de variabilidade explicada pela partição (Everitt et al., 2011; Jain, 2010). A escolha deste trio (WSS/cotovelo, *silhouette* e BSS/TSS) privilegia coerência com a função-objetivo do algoritmo, interpretação direta e comunicabilidade dos resultados, evitando redundância de critérios.

No exercício de determinação de  $k$ , a curva do cotovelo mostrou uma redução marcada da WSS até  $k = 3$ , a partir do qual os ganhos adicionais se tornaram marginais, sugerindo um ponto de inflexão nesse valor. Em paralelo, o índice de *silhouette* atingiu o seu máximo em  $k = 2$  ( $\approx 0,36$ ) e decresceu para 0,298 em  $k = 3$ , mantendo uma trajetória descendente nos valores seguintes. Embora  $k = 2$  maximize a separação relativa segundo o *silhouette*, a solução com  $k = 3$  oferece um equilíbrio mais favorável entre compacidade e granularidade interpretativa dos perfis, em linha com a literatura para dados comportamentais de elevada dimensionalidade e correlações moderadas (James et al., 2023; Xu & Wunsch, 2005; Asfahani, 2024). Consistentemente, no espaço dos cinco componentes, a razão BSS/TSS foi de  $\approx 52,62\%$ , indicando que mais de metade da variabilidade projetada é explicada pela partição, magnitude compatível com uma estrutura segmentável de nível intermédio (Everitt et al., 2011; Jain, 2010).

Para além da qualidade interna, avaliou-se a diferenciação entre grupos por meio de ANOVA unifatorial (*one-way*) aplicada, variável a variável, às medidas utilizadas na modelação. As variáveis foram previamente normalizadas e tratadas como contínuas para efeitos de teste. Esta escolha segue a prática recomendada na literatura de análise de *clusters*, que utiliza ANOVA como verificação da diferença de médias entre segmentos nas variáveis que estruturam o espaço de segmentação (Everitt et al., 2011; James et al., 2023). Observou-se diferença estatisticamente significativa ( $p < 0,001$ ) em todas as variáveis analisadas, com estatísticas F particularmente elevadas em “ser informado sobre atividades”, “convite para autodesenvolvimento” e “convite para reuniões”, sugerindo que acesso à informação, participação ativa e apoio hierárquico concentram a maior diferenciação entre segmentos. Reconhece-se, todavia, o potencial de circularidade quando se testam variáveis que também contribuíram para a formação dos *clusters*; por isso, estes resultados são apresentados como evidência descritiva de diferenciação e discutidos em conjunto com a interpretação substantiva baseada nos *loadings* das componentes (Cap. 3), onde se encontram as tabelas detalhadas e as figuras correspondentes.

Do ponto de vista computacional, todos os cálculos foram implementados em *Python*, com *scikit-learn* para o *k-means* e o *silhouette*, *NumPy/Pandas* para agregações e variâncias (WSS,

TSS, BSS) e *SciPy* para a ANOVA, assegurando-se a reprodutibilidade por definição de *seed* e registo de versões do ambiente. Esta combinação de procedimentos e métricas, ancorada em referências consolidadas da área (Lloyd, 1982; Kaufman & Rousseeuw, 2005; Everitt et al., 2011; Jain, 2010; James et al., 2023), oferece uma base suficiente para julgar a adequação da solução de três segmentos no contexto empírico em análise

## **2.5 Considerações éticas e reprodutibilidade**

Os dados foram utilizados no quadro de um acordo com a entidade fornecedora, exclusivamente para fins desta investigação, observando limitação de finalidade e confidencialidade. Antes da análise, identificadores pessoais foram suprimidos e não se efetuou qualquer tentativa de reidentificação. O conjunto original (não anonimizado) permanece em repositório de acesso restrito. A versão anonimizada, sem identificadores diretos, encontra-se disponível publicamente no *GitHub*. Esta política concilia transparência com proteção de dados, em linha com recomendações para gestão responsável de dados comportamentais (Defossez et al., 2020; Varsha P. S., 2023).

Para reprodutibilidade, todo o processo foi documentado de modo a permitir o rastreio integral das etapas: preparação, ACP e segmentação por *k-means*, incluindo semente aleatória, versões de *software* e ficheiros de ambiente (*requirements.txt/environment.yml*). O código e a documentação técnica estão no repositório *GitHub* público do projeto (controlo de versões e auditoria por *commit*). As opções metodológicas e de partilha seguem boas práticas de ciência de dados aplicada e de documentação reprodutível (Kuhn & Johnson, 2013; James et al., 2023).

## Capítulo 3 – Análise dos Resultados

Concluída a etapa metodológica, este capítulo apresenta e interpreta os resultados empíricos obtidos a partir do inquérito de diversidade e inclusão conduzidos pela *CLOO Behavioral Insights Unit*. Procura-se mostrar como a análise retrospectiva de 1 500 respostas evidencia padrões latentes de participação, percepções de inclusão e barreiras estruturais; como permite identificar segmentos com necessidades diferenciadas em matéria de DEI; e de que modo esses resultados se traduzem em indicadores acionáveis para o desenho de produtos formativos e para a orientação estratégica junto das empresas-cliente.

A exposição inicia-se com a caracterização descritiva da amostra e dos agrupamentos formados e prossegue para os resultados da segmentação obtida a partir da estrutura do questionário e com a análise dos grupos no espaço das componentes principais (CP). Em cada conjunto de resultados privilegia-se a interpretação substantiva dos perfis e as suas implicações práticas, assinalando, quando pertinente, limitações analíticas (por exemplo, separações moderadas entre grupos ou potenciais enviesamentos de autosseleção). A interpretação dos resultados apoia-se em quadros comportamentais consolidados (inclusão e voz, segurança psicológica, LMX, justiça informacional e redes/homofilia), de modo a traduzir padrões estatísticos em mecanismos organizacionais e ações de desenho de serviços.

### 3.1 Descrição demográfica da amostra

O inquérito subdivide-se em quatro secções principais: (i) Dados sociodemográficos e de percurso educativo, incluindo género, autoidentificação étnico-racial, região de residência, idade e nível de escolaridade; (ii) Estrato socioeconómico, aferido por rendimento familiar; (iii) Fatores de inclusão organizacional, com 15 itens avaliados em escala de *Likert* de cinco pontos, cobrindo decisão, redes de informação e grau de envolvimento; (iv) Cenários de microagressões e suporte de liderança, onde as afirmações medem, igualmente numa escala *Likert* de cinco pontos, percepções de vigilância, esforço adicional e apoio recebido de líderes negros ou brancos.

A amostra analisada compreende 1 500 respondentes, apresentando uma distribuição geográfica alinhada à densidade populacional brasileira: quase metade reside na Região Sudeste (47,8 %), enquanto Nordeste (23,3 %), Sul (14,8 %), Centro-Oeste (7,6 %) e Norte (6,5 %) surgem em proporções decrescentes. Em termos socioeconómicos, predomina a classe média — estratos C2 (19,3 %), C1 (25,7 %) e B2 (27,9 %) totalizam 72,9 % da amostra — ao passo que os segmentos B1 (12,1 %) e A (15,0 %) perfazem o grupo de maior poder aquisitivo. A escolaridade revela percursos educacionais diversificados: a maioria concluiu o Ensino Médio

(36,5 %), seguida por Ensino Superior completo (22,0 %) e incompleto (14,8 %); graus de pós-graduação (*lato e stricto sensu*) somam 14,2 %, enquanto o Ensino Fundamental não concluído representa menos de 4 %. A composição por género é equilibrada (51,5 % mulheres; 48,5 % homens), tal como a distribuição étnica (47,7 % brancos; 46,2 % negros; 6,1 % outras etnias). Por fim, 280 participantes (18,7 %) declararam ocupar cargos de liderança.

### 3.2 Descrição demográfica dos agrupamentos formados

A análise de agrupamento com base em cinco componentes principais do ACP resultou na formação de três grupos distintos de participantes (Grupo 1, Grupo 2 e Grupo 3), cujas características médias são apresentadas a seguir.

- O **Grupo 1** (n = 358; 48,3% líderes) apresenta uma idade média de 39,09 anos, a mais elevada entre os *clusters*, e a maior média de escolaridade (4,6 equivalente a ensino superior completo). É o grupo com maior percentual de participantes que se autoidentificam como líderes (48%) e com maior proporção de conhecimento em inglês (52%). A distribuição de género é equilibrada (52% mulheres e 48% homens), com uma leve predominância de participantes brancos (53%) em relação aos negros (44%).
- O **Grupo 2** (n = 381; 1,3% líderes) se destaca por apresentar a menor média de escolaridade (3,4 equivalente a ensino médio completo) e o menor percentual de conhecimento de inglês (23%). Embora tenha uma proporção maior de mulheres (58%), é o grupo com menor percentual de líderes autoidentificados (1%). A proporção de pessoas negras (48%) supera ligeiramente a de brancos (46%), sendo o único grupo onde isso ocorre.
- O **Grupo 3** (n = 761; 13,4% líderes) é o mais numeroso entre os *clusters*. Com uma idade média de 35,54 anos, representa o grupo mais jovem. Apresenta uma escolaridade média de 4,1 (equivalente a ensino superior incompleto) e uma proporção mais equilibrada entre brancos (46%) e negros (46%). Apesar de sua grande representatividade numérica, apenas 13% dos participantes se reconhecem como líderes.

De modo geral, observa-se que o perfil de liderança está mais concentrado no Grupo 1, associado a maior escolaridade, domínio do inglês e maior idade. Por outro lado, o Grupo 2 demonstra menor acesso a oportunidades de desenvolvimento profissional, ao passo que o Grupo 3 apresenta um potencial expressivo, embora ainda pouco refletido na presença em cargos de liderança.

### 3.3 Resultados da segmentação a partir da estrutura do questionário

Foi realizada uma leitura diretamente ancorada na estrutura do questionário, agregando os itens em quatro dimensões: inclusão organizacional, percepções de cultura inclusiva, microagressões e apoio das lideranças. Cada dimensão foi sintetizada pela média dos itens respectivos. Na presente análise, ressalta-se que os itens foram mantidos na sua direção original; valores mais altos significam maior inclusão, maior concordância com afirmações pró-inclusão, maior concordância com enunciados de desconforto/esforço e maior apoio percebido. O Figura 4 apresenta as médias por grupo enquanto o Anexo III detalha resultados por subgrupos (género x etnia).

**Inclusão Organizacional:** Avalia o grau de integração formal e informal dos profissionais nos processos de decisão e circulação de informação. Inclui indicadores como: ser convidado para reuniões, ter opinião escutada pelas chefias, participação em atividades de desenvolvimento e acesso a fluxos informacionais. A média destas variáveis gera o índice “inclusao\_organizacional”.

Observa-se um gradiente nítido: o Grupo 1 apresenta médias elevadas ( $\approx 4,0$ ), compatíveis com convites formais, acesso a informação e oportunidades de contribuição; o Grupo 2 exibe médias substancialmente mais baixas ( $\approx 1,3-1,5$ ), sugerindo exclusão processual; o Grupo 3 encontra-se num patamar intermédio ( $\approx 3,0$ ). A leitura por subgrupos confirma o padrão em todos os perfis, com variação residual dentro de cada grupo (ver Anexo III).

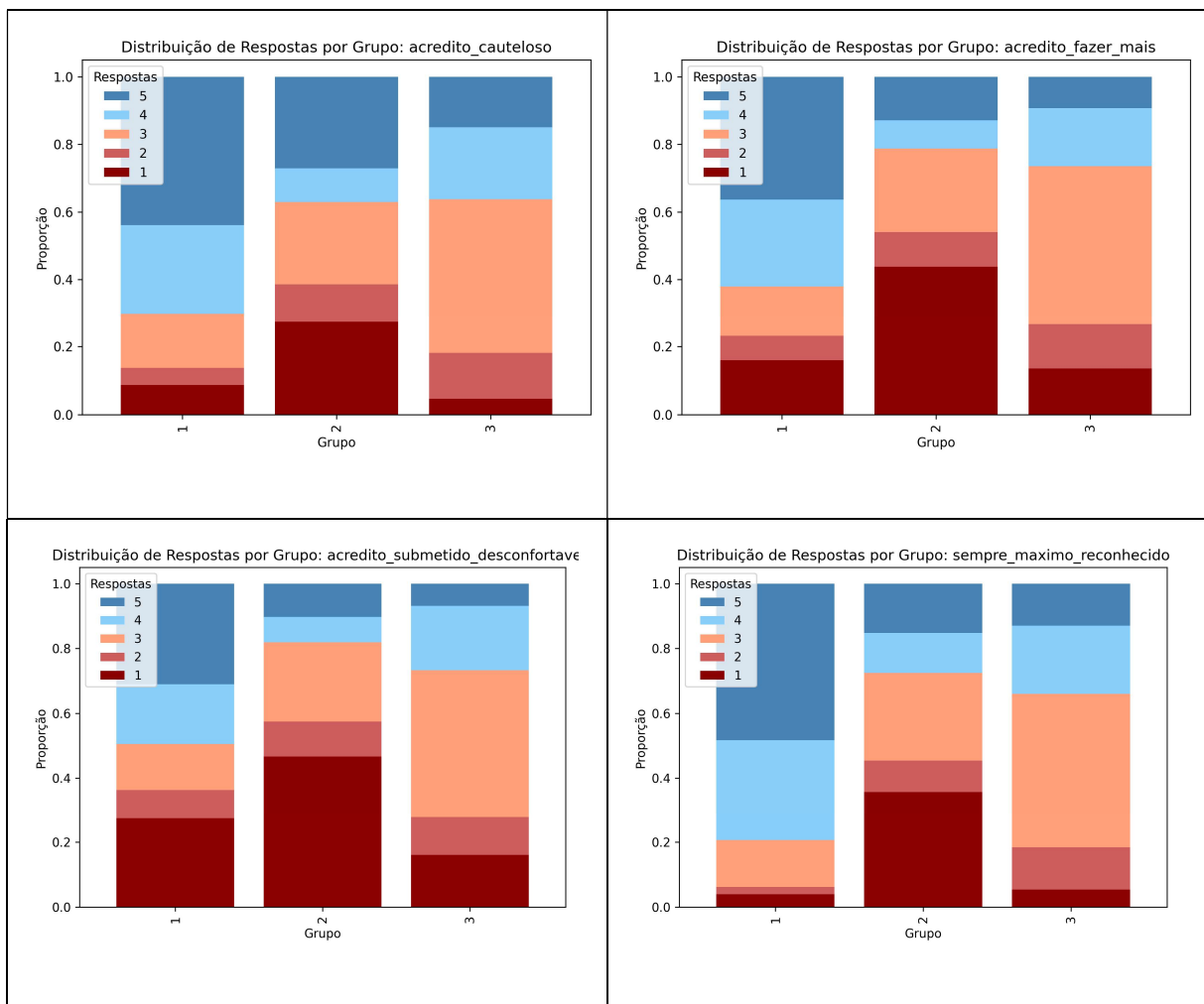
**Percepções de cultura inclusiva:** Reflete crenças normativas acerca da equidade racial e da abertura cultural da instituição. Engloba itens sobre facilidade de exercício de liderança, sentimento de acolhimento (“todas as pessoas são bem-vindas”), avaliação da igualdade racial nas práticas quotidianas e sensibilidade das lideranças à inclusão. Estes itens compõem o índice “percepção\_cultura\_inclusiva”.

As diferenças seguem orientação semelhante: Grupo 1 mais alto ( $\approx 4,0-4,4$ ), Grupo 2 mais baixo ( $\approx 2,1-2,9$ ) e Grupo 3 intermédio ( $\approx 3,0-3,3$ ). Importa sublinhar que esta dimensão capta crenças normativas (acolhimento, igualdade) e, portanto, deve ser cotejada com a inclusão processual: níveis altos de discurso pró-inclusão não garantem, por si, acesso efetivo a informação e decisão (Shore et al., 2011; Schein, 2017).

**Microagressões e Apoio das lideranças:** Para melhor captar as informações desse bloco, foram divididos em duas partes, a de microagressões e a de suporte percebido por terem sentido de valência opostas.

- **Apoio das lideranças:** Esta dimensão capta a percepção de apoio instrumental e patrocínio por parte de chefias diretas e indiretas (ajuda no desenvolvimento, criação de oportunidades), aproximando-se do constructo LMX — Leader–Member Exchange (Graen & Uhl-Bien, 1995). Na literatura, relações de alta qualidade com a liderança estão associadas a maior acesso a informação e recursos, tarefas de maior visibilidade e oportunidades de progressão, bem como a melhores atitudes e desempenho (Gerstner & Day, 1997; Dulebohn et al., 2012). Foi possível observar que no conjunto de dados do inquérito, o padrão repete-se: Grupo 1 com médias elevadas ( $\approx 4,0-4,3$ ), Grupo 2 com valores reduzidos ( $\approx 1,6-2,4$ ) e Grupo 3 em posição intermédia ( $\approx 2,6-3,0$ ). Esta configuração é compatível com a evidência de que maior qualidade relacional tende a ampliar o patrocínio e a abrir oportunidades de desenvolvimento (LMX alto), enquanto níveis baixos associam-se a apoio incipiente e menor circulação de recursos. Importa notar que o item de suporte diferenciado por raça conversa com a literatura sobre diferenciação de LMX dentro das equipas e com debates sobre equidade na distribuição de oportunidades; por isso, recomenda-se monitorização contínua e triangulação com indicadores objetivos de acesso a projetos, mentoria e progressão (Liden & Maslyn, 1998; Dulebohn et al., 2012)

- **Microagressões:** Capta experiências subtis de deslegitimação, vigilância simbólica e esforço desproporcional para reconhecimento. Inclui afirmações sobre ter de "fazer mais" para ser valorizado, submeter-se a situações desconfortáveis e o nível de reconhecimento máximo recebido. Entendem-se microagressões como indignidades quotidianas, verbais, comportamentais ou ambientais, frequentemente não intencionais, que comunicam hostilidade, desqualificação ou desvalorização a membros de grupos socialmente marginalizados (Sue et al., 2007; Mor-Barak, 2014). Por essa relevância substantiva e pelo potencial de ambiguidade interpretativa, este bloco é analisado com maior detalhe e com Figura 4 que contém as distribuições por item.



**Figura 4:** Distribuição das respostas do índice microagressões por grupo

As distribuições evidenciam três domínios relacionados, mas não idênticos: *(i)* desconforto reportado (“submetido a situações desconfortáveis”), *(ii)* hipervigilância (“ser sempre cauteloso”) e *(iii)* lógicas de reconhecimento (os dois itens sobre “fazer o máximo/ser reconhecido”). Em “tenho de fazer mais...” e “ser sempre cauteloso”, o Grupo 1 concentra maiores proporções de respostas 4–5, sugerindo carga subjetiva associada à visibilidade e às expectativas de desempenho. Em “submetido a situações desconfortáveis”, observa-se polarização no Grupo 1 (caudas em 1 e 4–5), discordância mais frequente no Grupo 2 (1–3) e zona intermédia no Grupo 3 (3–4). Finalmente, “sempre fiz o máximo...” apresenta endosso elevado, sobretudo no Grupo 1, indicando uma orientação/estratégia de desempenho que pode coexistir com, mas não equivale automaticamente a, experiências de microagressão. Assim, mesmo mantendo os quatro itens no compósito nesta versão, a interpretação é conduzida item a item e comparativamente entre grupos, evitando confundir adversidade percebida com normas de *performance*.

Duas implicações decorrem desta leitura. Primeiro, no Grupo 1 a elevada participação e o patrocínio coocorrem com hipervigilância e esforço percebido para reconhecimento, dinâmica coerente com contextos de alta visibilidade (Edmondson, 1999). Segundo, a menor concordância do Grupo 2 em alguns enunciados de desconforto pode refletir subrelato/normalização, menor exposição a interações onde microagressões emergem (pela própria exclusão processual) ou a direccionalidade dos enunciados nesta medição. Em trabalhos futuros, recomenda-se consolidar subdomínios (desconforto; hipervigilância; reconhecimento), reorientar sistematicamente itens de semântica negativa e reportar proporções 4-5 com intervalos de confiança, reforçando a precisão conceptual (Sue et al., 2007; Edmondson, 1999).

### **Leitura integrada por grupo:**

- Grupo 1: “Líderes Vigilantes”. Elevada inclusão processual e apoio percebido; crenças pró-inclusão também elevadas. A distribuição no índice de microagressões requer leitura cuidadosa por item, dado coexistirem altas expectativas de desempenho/visibilidade com concordância variável em desconforto.
- Grupo 2: “Potenciais em Resistência”. Menor inclusão e menor apoio, acompanhados de maiores relatos de desconforto/esforço. É o alvo prioritário para ações de mentoria/patrocínio, rotinas de convite e protocolos anti-microagressões.
- Grupo 3: “Trabalhadores Invisibilizados”. Posição intermédia em todas as dimensões, sugerindo integração parcial e atuação institucional ambígua. Pode ser beneficiado com a aplicação de medidas que reduzam a distância entre discurso pró-inclusão e práticas de partilha.

### 3.4 Análise dos grupos a partir dos componentes principais (CP)

A Análise de Componentes Principais (ACP) permitiu reduzir a dimensionalidade dos dados e extrair cinco componentes latentes (CP1-CP5) que, em conjunto explicam 63,9% da variância total (ver Sec. 2.2.4.1). Para leitura aplicada, cada componente foi operacionalizado como sumário exploratório calculado pela média das três variáveis com maior *loading* absoluto. Nota metodológica: nesta versão não houve reorientação de itens com semântica negativa; por isso, os CPs não devem ser lidos como escalas monotónicas. A interpretação privilegia a comparação relativa entre grupos e a semântica dos itens que compõem cada CP. As médias por grupo constam da Tabela 1 e a comparação visual do perfil dos grupos é apresentada no Figura 3. A descrição completa das questões são apresentadas no Anexo I.

**CP1. Participação ativa (38,8%):** captura convites formais, fluxo de informação e envolvimento em rotinas organizacionais.

**Variáveis:** influencia\_convitado\_reunioes, influencia\_convitado\_autodesenvolvimento, influencia\_informado\_atividades

O Grupo 1 apresenta médias elevadas ( $\approx 4,4$ ), consistentes com acesso a espaços de decisão e convites para desenvolvimento; o Grupo 2 fica muito abaixo ( $\approx 1,2$ ), refletindo exclusão processual; o Grupo 3 é intermédio ( $\approx 2,9$ ). Em termos comportamentais, CP1 aproxima-se do eixo pertença + unicidade (*belongingness/uniqueness*) e de voz efetiva na organização (Shore et al., 2011; Mor-Barak, 2014). A distância entre Grupos 1 e 2 sugere que barreiras de acesso (quem é chamado, quem recebe informação) são um ponto de alavanca imediato para políticas de inclusão.

**CP2. Perceção emocional e microagressões (8,40%):** Mede desconforto, hipervigilância e esforço para reconhecimento, domínios associados a microagressões e a climas de menor segurança psicológica (Sue et al., 2007; Edmondson, 1999).

**Variáveis:** acredito\_submetido\_desconfortaveis, acredito\_fazer\_mais, facil\_branças\_integrar

O Grupo 1 tende a valores mais altos ( $\approx 3,5$ ), indicando que visibilidade e expectativas de desempenho podem coexistir com carga subjetiva (necessidade de “fazer mais”, maior vigilância); o Grupo 2 mostra valores mais baixos a moderados ( $\approx 2,5$ ), o que pode refletir sub-relato, normalização ou menor exposição por exclusão de arenas de interação; o Grupo 3 situa-se entre ambos ( $\approx 3,0$ ). Em termos conceptuais, a combinação “desconforto + esforço para reconhecimento” é compatível com ameaça identitária e custos psicossociais de navegação de

ambiente (Sue et al., 2007). Reitera-se que o item “fazer mais para ser reconhecido” não define por si microagressões, mas coocorre com elas; a interpretação aqui é item a item, apoiada nos figuras específicos deste bloco (Sec. 3.2).

**CP3. Crença na cultura inclusiva (7,41%):** Mede crenças normativas sobre acolhimento e igualdade racial. Aproxima-se da narrativa cultural declarada (Schein, 2010/2017).

**Variáveis:** todas\_bemvindas, igualdade\_negras\_brancas, influencia\_supervisor\_nao\_partilha (valência negativa).

O Grupo 1 exhibe médias altas ( $\approx 4,0$ ), denotando confiança no discurso de inclusão; o Grupo 2 é mais cético ( $\approx 2,6$ ) e o Grupo 3 mantém-se intermédio ( $\approx 3,3$ ). A presença do item negativo (“não partilha informação”) no conjunto top-3 relembra que retórica de igualdade deve ser lida em conjunto com sinais processuais de partilha e voz (ligação a CP1). Em termos comportamentais, discrepâncias entre discurso e prática podem gerar incongruência de justiça e ambiguidade de normas (Greenberg, 1990; Schein, 2010/2017).

**CP4. Atitudes sobre políticas inclusivas (4,7%):** Expressa o grau de apoio ou rejeição a ações afirmativas e medidas institucionais de inclusão. Combina normas de mérito, sensibilidade das lideranças e reconhecimento de necessidade de políticas (Ajzen, 1991).

**Variáveis:** nao\_necessita\_medidas\_inclusivas (valência negativa), sempre\_maximo\_reconhecido, lideranças\_sensíveis\_inclusão\_racial

O Grupo 1 apresenta valores mais altos ( $\approx 4,0$ ), compatíveis com confiança no mérito e percepção de sensibilidade das lideranças; o Grupo 2 fica mais abaixo ( $\approx 2,6$ ), sinalizando maior aceitação de políticas e percepção de insuficiência atual; o Grupo 3 mantém posição intermediária ( $\approx 3,1$ ). Em termos de comportamento, atitudes favoráveis a políticas podem depender de crenças de eficácia e normas subjetivas (Ajzen, 1991) e dialogam com percepções de justiça organizacional (Greenberg, 1990).

**CP5. Suporte e barreiras institucionais (4,3%):** Reflete a percepção sobre obstáculos estruturais percebidos e inclusão informal; cruza expectativas de maior facilidade para brancos com sinais de exclusão relacional (por ex., convívios) e barreiras à liderança.

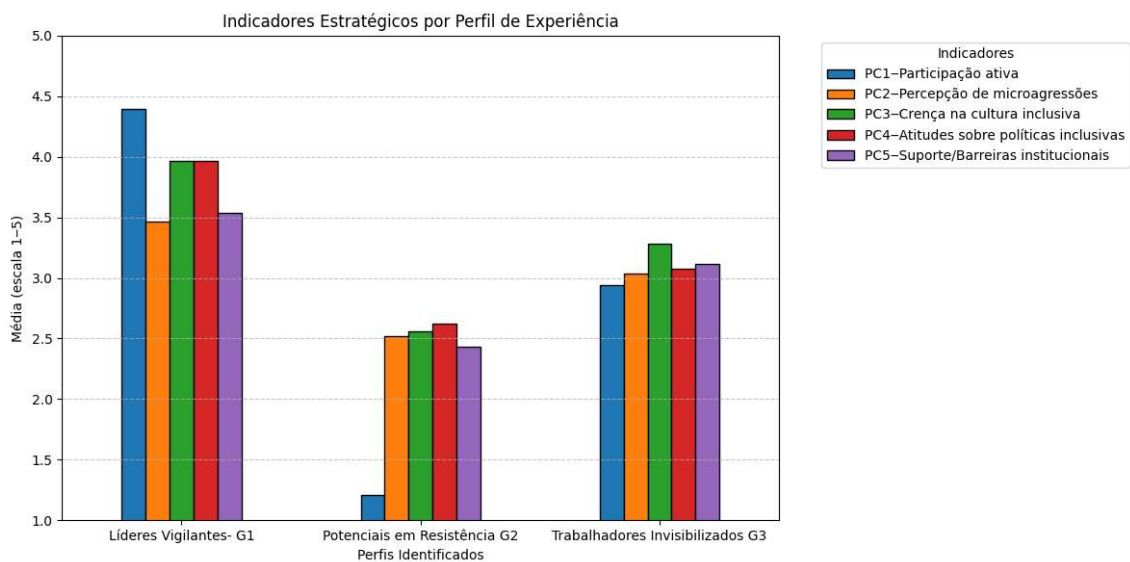
**Variáveis:** percebo\_facil\_branco\_lider, facil\_branças\_integrar, influencia\_raramente\_convívio (todas com valência negativa).

O **Grupo 1** apresenta média mais elevada ( $\approx 3,5$ ), reconhecendo privilégios raciais e sinalizando atenção crítica a barreiras; o Grupo 2 situa-se abaixo ( $\approx 2,4$ ), o que pode refletir distanciamento do debate ou normalização; o Grupo 3 é intermédio ( $\approx 3,1$ ). Em termos comportamentais, exclusão relacional (menos convites informais) fragiliza laços de troca com lideranças (**LMX**) e reduz oportunidades (Graen & Uhl-Bien, 1995), mesmo quando o discurso organizacional (CP3) é positivo.

<b>Indicador (CP)</b>	<b>% Var. Explicada</b>	<b>G1 (Média)</b>	<b>G2 (Média)</b>	<b>G3 (Média)</b>
CP1	38.8	4.40	1.21	2.94
CP2	8.7	3.47	2.52	3.03
CP3	7.4	3.96	2.56	3.28
CP4	4.7	3.96	2.62	3.08
CP5	4.3	3.54	2.43	3.12

**Tabela 1:** Média aritmética das respostas por CP x Grupos (escala *Likert* 1 a 5)

A Figura 5 evidencia, de forma sintética, o posicionamento médio de cada perfil nas cinco dimensões estratégicas. Observa-se que os Líderes Vigilantes (Grupo 1) apresentam valores consistentemente elevados em todas as dimensões, destacando-se na Participação Ativa ( $\approx 4,4$ ). Em contraste, os Potenciais em Resistência (Grupo 2) registam o desempenho mais baixo, sobretudo na mesma dimensão contrastando com o Grupo 1, Participação Ativa ( $\approx 1,2$ ). Os Trabalhadores Invisibilizados (Grupo 3) ocupam uma posição intermédia, com níveis moderados em todas as dimensões (entre 2,9 e 3,3), revelando integração parcial, mas ainda frágil. A comparação visual sublinha, portanto, a clivagem clara entre um grupo altamente incluído e dois grupos que carecem de políticas mais específicas de reconhecimento e apoio, sendo o Grupo 2 particularmente vulnerável. Essa matriz ajuda a compreender como os grupos diferem não apenas em suas posições objetivas, mas também nas variáveis relacionadas à crenças, emoções e percepções de justiça, o que é central para orientar políticas organizacionais inclusivas.



**Figura 5:** Indicadores estratégicos por perfil de experiência

### 3.4 Recomendações metodológicas para futuros inquéritos

Um aspeto relevante para iniciativas orientada por dados encontra-se indissociavelmente ligada à qualidade do instrumento de recolha. Nesta conformidade, delineiam-se, infra, um conjunto de orientações estratégicas suscetíveis de incrementar a validade científica e a utilidade prática do projeto.

#### 3.4.1 Avaliação da consistência interna e validação da estrutura latente

A fiabilidade interna de cada dimensão deve ser estimada com o  $\alpha$  de Cronbach, (Cronbach, 1951) e, preferencialmente, com o  $\omega$  de McDonald (total e/ou hierárquico), por oferecer uma aproximação mais adequada sob modelos congenericamente realistas (McDonald, 1999). Dado o carácter ordinal das escalas *Likert*, recomenda-se o uso de correlações policóricas e a apresentação de  $\alpha/\omega$  ordinais (Gadermann, Guhn, & Zumbo, 2012). Quando a estrutura fatorial não estiver consolidada, procede-se a Análise Fatorial Exploratória (AFE) e, subsequentemente, a Análise Fatorial Confirmatória (AFC), utilizando estimadores apropriados a dados ordinais, como o WLSMV, e relatando de forma transparente os critérios de qualidade de ajustamento (Flora & Curran, 2004; Brown, 2015; Worthington & Whittaker, 2006). Por fim, atendendo às comparações entre género  $\times$  etnia  $\times$  nível, é desejável testar invariância de medida pelo menos aos níveis configural, métrico e escalar, assegurando a comparabilidade entre subgrupos antes de interpretar diferenças (Meredith, 1993; Putnick & Bornstein, 2016).

### **3.4.2 Modularização do inquérito e normalização das escalas**

A transformação do inquérito em módulos independentes aumenta a flexibilidade da sua aplicação, reduz a fadiga do respondente e facilita a subsequente análise segmentada dos dados. A uniformização das escalas, preferindo a adoção de escalas *Likert* de seis ou sete pontos promove a coerência métrica, simplifica procedimentos de pontuação e viabiliza comparações dentro do próprio estudo e entre diferentes estudos. A padronização da arquitetura de dados constitui, outrossim, uma boa prática consagrada para reforçar a qualidade da investigação, fomentar a transparência e assegurar a reprodutibilidade dos resultados (Defossez et al., 2020).

Note-se, todavia, que os denominados dados Q, por assentarem em autorrelato, estão sujeitos a enviesamentos de desejabilidade social ou de (auto)perceção; impõe-se, pois, uma triangulação metodológica com outras fontes de evidência comportamental sempre que tal se revele exequível.

### **3.5 Redefinição de Produtos Oferecidos e Posicionamento**

Com vista a reforçar a proposta de valor e a consolidação da identidade institucional, a redefinição dos produtos oferecidos articula-se em três eixos complementares: (i) reconfiguração da oferta formativa, (ii) desenvolvimento de ferramentas analíticas orientados para a geração transversal de valor e (iii) estratégias de posicionamento e liderança intelectual. As propostas que se seguem convertem a evidência empírica deste estudo em soluções aplicadas. Partem dos mecanismos identificados ao longo dos resultados: inclusão processual e voz (CP1), pressão subjetiva/desconforto (CP2), discurso de inclusão vs. práticas de partilha (CP3), atitudes a políticas e liderança (CP4) e barreiras percebidas/redes informais (CP5). E ainda dos três segmentos obtidos por *K-means*. Cada iniciativa alinha: (i) o mecanismo comportamental observado, (ii) a alavanca de intervenção (formação, produto analítico, comunicação científica) e (iii) os indicadores de sucesso. O objetivo é oferecer soluções consistentes com a evidência e úteis para decisão. A seguir, apresentam-se as principais iniciativas previstas em cada um dos eixos:

#### **3.5.1 Redefinição da oferta formativa: formações orientadas por evidências**

Para garantir uma formação que produza bons resultados para as organizações, é proposta a realização de diagnósticos e identificação de comportamentos observáveis: quem é convidado para o quê (voz e participação), como circula a informação (acesso), que patrocínio é acionado (LMX) e como se mitigam microagressões e hipervigilância (Sue et al., 2007; Edmondson, 1999). Assim, os conteúdos formativos derivam diretamente dos gaps revelados por CP1

(participação), CP2 (microagressões/hipervigilância), CP4 (atitudes/políticas) e CP5 (barreiras estruturais e exclusão relacional):

- **Oficina “Radar Inclusivo”:** Curso de curta duração direcionado a líderes e RH no qual realizam leituras de indicadores (convites para reuniões, convites para autodesenvolvimento, partilha efetiva), casos-tipo extraídos dos três perfis (Líderes Vigilantes; Potenciais em Resistência; Trabalhadores Invisibilizados) e micro-rotinas para ampliar voz e acesso (Edmondson, 1999; Shore et al., 2011)
- **Curso “Data-Driven DEI Strategy” (formato *blended*, 16 h):** Aborda a integração de métricas DEI em sistemas *Objectives and Key Results* (OKR), com foco em patrocínio estruturado (LMX aplicado a oportunidades e feedback) e em sinais culturais que alinham discurso e prática (Graen & Uhl-Bien, 1995; Schein, 2010/2017). A lógica é: diagnóstico → microintervenções → revisão de processos, com avaliação periódica dos efeitos.

### 3.5.2 Produtos analíticos e leitura integrada (medir, comparar, melhorar)

A gestão de DEI exige medidas estáveis, comparáveis no tempo e entre grupos, e um quadro de leitura que una dimensões culturais e processuais. Para tal, a CLOO organiza os resultados num painel de indicadores e numa métrica compósita, preservando a natureza proprietária do modelo:

- **Desenvolvimento do Índice CLOO-DEI:** O Índice é posicionado como *score* compósito de maturidade (0–100) alinhado aos mecanismos acima (inclusão processual, cultura/justiça, experiência de microagressões, suporte da liderança e atitudes). O relatório entrega leitura por pilar e por segmento (género × etnia × nível), orientando metas e *benchmarking* setorial. (*Detalhes proprietários omitidos por confidencialidade.*) Para tal, a identificação de agrupamentos homogêneos será precedida por uma mediante Análise Arquetípica (AA), técnica estatística que visa identificar perfis extremos, os denominados arquétipos, situados nos limites do espaço amostral (isto é, no invólucro convexo dos dados). Ao contrário de abordagens centradas em soluções médias, como a ACP, a AA procura decompor cada observação como uma combinação convexa destes “tipos puros”, possibilitando uma caracterização mais expressiva e intuitiva das diferentes manifestações de inclusão ou exclusão (Cutler & Breiman, 1994). Os coeficientes obtidos pela AA, que expressam o grau em que cada entidade se aproxima de cada arquétipo, servirão subsequentemente como base para técnicas de agrupamento, as quais permitirão identificar grupos homogêneos de

organizações ou colaboradores, em função das suas combinações particulares destes perfis extremos.

Esta abordagem apoia-se no trabalho seminal de Cutler e Breiman (1994), que introduziram formalmente a Análise Arquetípica como um método de decomposição de dados multivariados, destacando a sua utilidade para captar padrões “puros” ou representações extremas no conjunto de dados.

### 3.5.3 Posicionamento e referência estratégica e científica

A consolidação de uma prática orientada por evidência implica também produção pública de conhecimento aplicado, o que reforça a confiança de clientes e parceiros, atrai talento e alimenta ecossistemas de aprendizagem:

- **White Paper “Data Science para DEI”:** Publicação de carácter técnico-científico destinada a documentar a metodologia subjacente ao Índice CLOO-DEI e assim posicionar o projeto como referência na aplicação de ciência de dados à gestão de DEI.
- **Dados & Cultura de Inclusão:** Indicadores que guiam decisões promove diálogo com empresas-cliente, combinando leituras de padrões recorrentes (p. ex., desfasamento entre discurso de inclusão e partilha efetiva) com soluções fundamentadas em LMX, segurança psicológica e gestão de microagressões (Graen & Uhl-Bien, 1995; Edmondson, 1999; Sue et al., 2007). Esta frente posiciona a CLOO como referência técnica, reforçando a legitimidade das recomendações e a transferência do que funciona. Em síntese, as recomendações ora apresentadas visam reforçar a base metodológica dos programas, potenciar a criação transversal de valor nas organizações e afirmar uma liderança intelectual ancorada em evidências empíricas.

### 3.6 Limitações e caminhos para investigações futuras

A presente análise comporta um conjunto de limitações que importa reconhecer e explicitar, tanto do ponto de vista da estrutura dos dados como das opções metodológicas adoptadas. A sua identificação permite, por um lado, uma interpretação mais rigorosa dos resultados obtidos e, por outro, a delimitação de caminhos para investigações subsequentes com maior rigor metodológico.

#### 3.6.1 Estrutura e qualidade dos dados

A base de dados utilizada apresenta limitações estruturais que se revelam particularmente relevantes no contexto da ciência de dados comportamentais. A ausência de padronização na

nomenclatura das variáveis, a escassa documentação associada à codificação das respostas e a variabilidade nos níveis de granularidade (i.e., mistura de dados ao nível do participante e ao nível do ensaio) comprometem a integridade e reprodutibilidade da análise. Tal como apontado por Defossez et al. (2020), a inexistência de normas consolidadas para organização e partilha de dados comportamentais representa um entrave ao progresso científico.

### **3.6.2 Redução da dimensionalidade e segmentação**

A utilização de técnicas de redução da dimensionalidade (e.g., Análise de Componentes Principais) e de segmentação (e.g., *k-means*) introduz elementos de subjetividade, nomeadamente quanto à seleção do número de componentes ou *clusters* a reter. Estas técnicas assumem pressupostos fortes de homogeneidade e são sensíveis a *outliers*, podendo comprometer a estabilidade dos perfis identificados. A investigação futura poderá se beneficiar da exploração de modelos baseados em distribuições mistas ou abordagens hierárquicas mais flexíveis (James et al., 2023).

### **3.6.3 Representatividade e generalização**

A amostra em estudo, conquanto numerosa, resulta de um processo de seleção não probabilístico, pelo que os resultados não são generalizáveis para o conjunto da população ativa. Acresce que determinadas variáveis contextuais de relevo, como informação longitudinal, indicadores de desempenho organizacional ou de dinâmicas grupais, não se encontram presentes, restringindo a interpretação multivariada dos padrões de perceção identificados.

### **3.6.4 Avaliação psicométrica do instrumento**

A avaliação da fiabilidade e da validade confirmatória do instrumento, nomeadamente por meio do coeficiente *alfa* de Cronbach e de Análise Fatorial Confirmatória (AFC), não foi realizada no âmbito deste estudo em virtude de limitações temporais, sendo estas etapas reservadas para investigações futuras. Esta limitação compromete a validação empírica da estrutura latente do instrumento, impedindo a confirmação dos constructos subjacentes (Mor-Barak, 2014).

Face às limitações identificadas, propõem-se as seguintes direções para futuras investigações:

- Adoção de normas de estruturação e documentação de dados comportamentais, conforme o modelo *Behavarse* (Defossez et al., 2020), que possibilitem maior rigor, reprodutibilidade e integração de dados.

- Exploração de metodologias orientadas para a interpretabilidade dos resultados, com ênfase em ferramentas de visualização e explicação de modelos, conforme discutido por Iodice D'Enza et al. (2024).
- Integração de dados longitudinais e contextuais, aptos a captar a dinâmica das percepções e sua relação com trajetórias organizacionais.
- Articulação com abordagens qualitativas ou mistas, com vista a compreender os sentidos atribuídos pelos participantes às suas experiências de inclusão, reconhecimento e pertença.

Estas linhas orientadoras permitirão aprofundar a compreensão empírica e conceptual das dinâmicas de inclusão em contextos organizacionais diversos, contribuindo para o desenvolvimento de práticas mais informadas, justas e eficazes.

## Conclusão e Investigação Futura

O presente trabalho examinou, em regime retrospectivo (*post-hoc*), o inquérito de diversidade e inclusão aplicado pela CLOO *Behavioral Insights Unit* em 2022 a 1500 respondentes. A aplicação sequencial de procedimentos de limpeza e enriquecimento dos dados, da Análise de Componentes Principais, de técnicas de segmentação *K-means* tornou possível detetar estruturas latentes que não haviam sido reveladas pelas análises descritivas originais, oferecendo-lhes um novo enquadramento interpretativo e conferindo-lhes utilidade prática no apoio à tomada de decisões.

A investigação demonstrou a existência de cinco componentes principais, responsáveis por 63,8% da variância total, que corporizam dimensões fundamentais da experiência organizacional, designadamente a participação ativa, a perceção de microagressões, as crenças de igualdade, as atitudes perante políticas inclusivas e o suporte institucional. A segmentação subsequente em três *clusters* homogéneos, validada por métricas internas e por análise de variância, evidenciou perfis diferenciados de risco de exclusão e permitiu delinear intervenções de diversidade, equidade e inclusão (DEI) ajustadas às necessidades específicas de cada grupo.

Para a literatura, o estudo comprova que a integração entre enquadramentos conceptuais de inclusão-exclusão (Mor-Barak, 2014) e algoritmos de *machine learning* enriquece o diálogo entre ciências comportamentais e ciência de dados. Para a CLOO, os resultados facultam bases empíricas para o refinamento dos produtos formativos oferecidos às empresas-cliente, fornecem indicadores de monitorização imediata — como o grau de convite ao autodesenvolvimento — e reforçam o posicionamento comercial da empresa enquanto consultora orientada por evidências.

Não obstante a relevância das recolhas, subsistem limitações que enquadram a sua interpretação. O valor moderado do *silhouette* (0,298) evidencia fronteiras difusas entre grupos, o possível viés de autosseleção dos respondentes aconselha prudência na generalização externa, e a utilização de ACP sobre escalas de *Likert* impõe reservas quanto à intervalaridade subjacente. Estas restrições sugerem que futuras investigações deverão recorrer a desenhos longitudinais, testar algoritmos de segmentação alternativos, ponderar a aplicação de correlações policóricas e adotar estratégias de ponderação pós-estratificação.

Levando-se em conta as ressalvas e adequações apresentadas no parágrafo anterior, foi possível identificar oportunidades de investigação futura, como por exemplo: Identificação de Preditores Críticos de Progressão para Cargos de Liderança, Desenvolvimento de um estudo longitudinal,

recorrendo a modelação estatística multivariada (v.g., regressão logística multinível, árvores de decisão e algoritmos de *machine learning*), com o propósito de isolar variáveis individuais, contextuais e institucionais que predizem a probabilidade de ascensão a funções de chefia. O objetivo consiste em construir um quadro preditivo que sustente políticas de gestão de talento baseadas em evidência empírica.

O caminho que se abre convida, ainda, à aplicação de métodos de análise de redes para mapear influências interpessoais, à exploração de técnicas de processamento de linguagem natural que revelem narrativas subtis de exclusão e à avaliação multicritério da robustez dos algoritmos de segmentação utilizados. A periodicidade da recolha de dados, combinada com triangulação qualitativa e com a adoção sistemática de padrões internacionais, potenciará a construção de um observatório contínuo de práticas de inclusão.

Em síntese, foi demonstrado que a aplicação de técnicas avançadas de ciência de dados a conjuntos de dados comportamentais permite diagnosticar fenómenos complexos de diversidade e inclusão e convertê-los em recomendações operacionais capazes de amplificar o impacto social e económico das organizações. A CLOO, ao internalizar este processo, afirma-se como referência na transformação *data-driven* de políticas de DEI e posiciona-se, a médio prazo, como laboratório de excelência em ciência de dados comportamentais aplicada à igualdade e justiça organizacional.

## Referências Bibliográficas

- Asfahani, A. M. (2024). *Discover Sustainability Fusing talent horizons: the transformative role of data integration in modern talent management*. 5, 25. <https://doi.org/10.1007/s43621-024-00212-7>
- Borgatti, S. P., & Halgin, D. S. (2011). On network theory. *Organization Science*, 22(5), 1168–1181. <https://doi.org/10.1287/orsc.1100.0641>
- Bro, R., & Smilde, A. K. (2014). Principal component analysis. *Analytical Methods*, 6(9), 2812–2831. <https://doi.org/10.1039/C3AY41907J>
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). Guilford Press.
- Cattell, R. B. (1950). Personality: A systematic theoretical and factual study, 1st ed. In *Personality: A systematic theoretical and factual study, 1st ed.* McGraw-Hill. <https://doi.org/10.1037/10773-000>
- Cox, T. H., & Blake, S. (1991). Managing Cultural Diversity: Implications for Organizational Competitiveness. *The Executive*, 5(3), 45–56. <http://www.jstor.org/stable/4165021>
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334.
- Cutler, A., & Breiman, L. (1994). Archetypal analysis. *Taylor & Francis A Cutler, L Breiman Technometrics*, 1994•Taylor & Francis, 36(4), 338–347. <https://doi.org/10.1080/00401706.1994.10485840>
- Defossez, A., Ansarinia, M., Clocher, B., Schmück, E., Schrater, P., & Cardoso-Leite, P. (2020). *The structure of behavioral data*. <https://doi.org/https://doi.org/10.48550/arXiv.2012.12583>
- Dulebohn, J. H., Bommer, W. H., Liden, R. C., Brouer, R. L., & Ferris, G. R. (2012). A meta-analysis of antecedents and consequences of leader–member exchange. *Journal of Management*, 38(6), 1715–1759.

- Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, *44*(2), 350–383.
- Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster analysis* (5th ed.). Wiley.
- Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of estimation methods for CFA with ordinal data. *Psychological Methods*, *9*(4), 466–491.
- Gadermann, A. M., Guhn, M., & Zumbo, B. D. (2012). Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Practical Assessment, Research & Evaluation*, *17*, Article 3. <https://doi.org/10.7275/n560-j767>
- Gerstner, C. R., & Day, D. V. (1997). Meta-analytic review of leader–member exchange theory: Correlates and construct issues. *Journal of Applied Psychology*, *82*(6), 827–844.
- Giordani, P., Henk, ·, & Kiers, A. L. (2023). Weighted least squares for archetypal analysis with missing data. *Behaviormetrika*, *51*, 441–475. <https://doi.org/10.1007/s41237-023-00220-3>
- Graen, G. B., & Uhl-Bien, M. (1995). Relationship-based approach to leadership: LMX theory. *The Leadership Quarterly*, *6*(2), 219–247.
- Greenberg, J. (1990). Organizational justice: Yesterday, today, and tomorrow. *Journal of Management*, *16*(2), 399–432.
- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2009). *ANÁLISE MULTIVARIADA DE DADOS* (Bookman, Ed.).
- Iodice D’Enza, A., Markos, A., & Kurihara, K. (2024). Issues in behavioral data science. *Behaviormetrika*, *51*(1). <https://doi.org/10.1007/s41237-023-00222-1>
- Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, *31*(8), 651–666. <https://doi.org/10.1016/j.patrec.2009.09.011>
- James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). *An Introduction to Statistical Learning*. <https://doi.org/10.1007/978-3-031-38747-0>
- Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2065), 20150202. <https://doi.org/10.1098/rsta.2015.0202>

- Kaiser, H. F. (1960). The Application of Electronic Computers to Factor Analysis. *Educational and Psychological Measurement*, 20(1), 141–151. <https://doi.org/10.1177/001316446002000116/ASSET/A9E82F80-3621-4CD1-9787-DDA676302364/ASSETS/001316446002000116.FP.PNG>
- Ketemaw, A., Ayenew, Z., & Zewde, S. (2024). *Managing diversity and multi-cultural collaboration: A systematic review*. *SAGE Open*, 14(4), 1–10. <https://doi.org/10.1177/21582440241273933>
- Kaufman, L., & Rousseeuw, P. J. (2005). *Finding groups in data: An introduction to cluster analysis*. Wiley.
- Kelleher, J. D., & Tierney, B. (2018). *Data science*. MIT Press.
- Kuhn, M., & Johnson, K. (2013). Applied predictive modeling. *Applied Predictive Modeling*, 1–600. <https://doi.org/10.1007/978-1-4614-6849-3/COVER>
- Liden, R. C., & Maslyn, J. M. (1998). Multidimensionality of leader–member exchange (LMX): Development of a measure and test of theory. *Organizational Behavior and Human Decision Processes*, 73(3), 267–293.
- Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2), 129–137. <https://doi.org/10.1109/TIT.1982.1056489>
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In L. M. Le Cam & J. Neyman (Eds.), *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, pp. 281–297). University of California Press.
- Matos, J., Cruz, C., Mattar, L. D., & Mauro, C. (2021). *Pessoas e práticas: Como as ciências comportamentais podem promover a diversidade e a inclusão nas organizações*. CLOO – Behavioral Insights Unit; Instituto Mattos Filho.
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Lawrence Erlbaum.
- Meredith, W. (1993). Measurement invariance, factorial invariance, and factorial comparison. *Psychometrika*, 58(4), 525–543.

- Mor-Barak, M. E. (2014). *Managing Diversity: Toward A Globally Inclusive Workplace*. Michalle. In *SAGE Publications* (fourth edition). SAGE.
- Palazzo, L., Iannario, M., & Palumbo, F. (2024). Integrated assessment of financial knowledge through a latent profile analysis. *Behaviormetrika*, *51*(1), 319–339. <https://doi.org/10.1007/S41237-023-00217-Y/FIGURES/6>
- Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting. *Child Development Perspectives*, *10*(4), 299–305.
- P.S., Varsha. (2023). How can we manage biases in artificial intelligence systems – A systematic literature review. *International Journal of Information Management Data Insights*, *3*(1), 100165. <https://doi.org/https://doi.org/10.1016/j.ijime.2023.100165>
- Saxena, S., Deogaonkar, A., Pais, R., & Pais, R. (2023). *Workplace productivity through employee sentiment analysis using machine learning*. *International Journal of Professional Business Review*, *8*(4), e01216. <https://doi.org/10.26668/businessreview/2023.v8i4.1216>
- Schein, E. H. (2010/2017). *Organizational culture and leadership* (4th ed.). Jossey-Bass.
- Shearer C. (2000). The CRISP-DM model: the new blueprint for data mining”, *Journal of Data Warehousing* *5* (2000) ; 5:13—22. *J Data Warehousing* , *5*(4).
- Shore, L. M., Randel, A. E., Chung, B. G., Dean, M. A., Ehrhart, K. H., & Singh, G. (2011). Inclusion and diversity in work groups: A review and model for future research. *Journal of Management*, *37*(4), 1262–1289. <https://doi.org/10.1177/0149206310385943;JOURNAL:JOURNAL:JOMA;WGROU:STRING:PUBLICATION>
- Sue, D. W., Capodilupo, C. M., Torino, G. C., et al. (2007). Racial microaggressions in everyday life. *American Psychologist*, *62*(4), 271–286.
- van Dijk, H., van Engen, M., & Paauwe, J. (2012). Reframing the Business Case for Diversity: A Values and Virtues Perspective. *Journal of Business Ethics*, *111*(1), 73–84. <https://doi.org/10.1007/S10551-012-1434-Z>

Worthington, R. L., & Whittaker, T. A. (2006). Scale Development Research: A Content Analysis and Recommendations for Best Practices. *The Counseling Psychologist*, 34(6), 806–838. <https://doi.org/10.1177/0011000006288127>

Xu, R., & Wunsch, D. (2005). Survey of clustering algorithms. *IEEE Transactions on Neural Networks*, 16(3), 645–678. <https://doi.org/10.1109/TNN.2005.845141>

## Apêndice I - Descrição das questões originais e dimensões

Variável	Pergunta original
<b>Dimensão Fatores de inclusão na organização</b>	
influencia_meu_grupo	1. Tenho influência nas decisões tomadas pelo meu grupo de trabalho no que diz respeito à organização de tarefas.
influencia_colegas_partilham_info	2. Os meus colegas partilham comigo informações relacionadas ao trabalho.
influencia_envolvido_atividades	3. Frequentemente sou envolvido e convidado para participar ativamente em atividades relacionadas com o trabalho do meu grupo de trabalho.
influencia_sentimento	4. Sinto que posso influenciar decisões que afetam a minha organização.
influencia_ultimo_ter_conhecimento	5. Geralmente sou um dos últimos a ter conhecimento de mudanças importantes na organização.
influencia_convidado_reunioes	6. Geralmente sou convidado para reuniões importantes na minha organização.
influencia_supervisor_opinioao	7. O meu supervisor regularmente pede a minha opinião antes de tomar decisões importantes.
influencia_supervisor_nao_partilha	8. O meu supervisor não partilha informação comigo.
influencia_convidado_autodesenvolvimento	9. Sou convidado pela minha liderança para participar de reuniões sobre meu desenvolvimento
influencia_convidado_contribuir_reunioes	10. Sou frequentemente convidado a contribuir com a minha opinião em reuniões com pessoas com posição hierárquica superior à minha supervisão imediata.
influencia_comunicacao_superior	11. Recebo frequentemente comunicações de pessoas com posição hierárquica superior ao meu supervisor (por exemplo, e-mails, memorandos).
influencia_convidado_superior	12. Sou frequentemente convidado a participar em reuniões com pessoas com posição hierárquica superior ao meu supervisor.
influencia_convidado_colegas	13. Sou frequentemente convidado pelos meus colegas para contribuir no planeamento de atividades sociais não relacionadas com o meu trabalho.
influencia_informado_atividades	14. Sou sempre informado de atividades sociais informais e eventos sociais da organização.

influencia_raramente_convivio	15. Raramente sou convidado a juntar-me aos meus colegas quando vão almoçar juntos ou conviver após o trabalho.
<b>Fatores da organização:</b>	
percebo_facil_branco_lider	Percebo que é mais fácil para uma pessoa branca exercer liderança do que uma pessoa negra.
facil_exercer_lideranca	É fácil exercer liderança na minha organização
todas_bemvindas	Na minha organização todas as pessoas são bem-vindas
igualdade_negras_branças	Na minha organização as pessoas negras são tratadas de igual modo que as pessoas brancas.
facil_branças_integrar	É mais fácil para uma pessoa branca integrar-se na cultura da organização.
nao_necessita_medidas_inclusivas	Não são necessárias medidas adicionais para tornar a organização inclusiva.
lideranças_sensíveis_inclusão_racial	As lideranças são sensíveis às questões de inclusão racial.
<b>Bloco de avaliação de cenários de microagressões</b>	
acredito_submetido_desconfortáveis	Acredito que sou submetido a situações desconfortáveis no cotidiano
acredito_cauteloso	Acredito que preciso ser sempre cauteloso em todas as situações
acredito_fazer_mais	Acredito que sempre tenho que fazer mais que os meus colegas de trabalho para ter reconhecimento
sempre_maximo_reconhecido	Sempre busquei fazer o máximo das atividades para ser reconhecido
lideranças_auxiliaram_desenvolvimento	Lideranças auxiliaram meu desenvolvimento
lideranças_desenvolveram_oportunidades	Lideranças desenvolveram oportunidades nas organizações
existiam_lideranças_negras	Nas organizações por onde passei, existiam lideranças negras. (S/N)
suporte_lideranças_negras	Tive o suporte de lideranças negras (apenas se S na resposta anterior)
suporte_lideranças_branças	Tive o suporte de lideranças brancas

**Apêndice II - Tabela ANOVA questões mais determinantes para a formação dos grupos**

Variável	Estatística F	p-valor
influencia_informado_atividades	1378.581	0.000
influencia_convidado_autodesenvolvimento	1273.700	0.000
influencia_convidado_reunioes	1268.821	0.000
influencia_convidado_colegas	1254.668	0.000
influencia_convidado_superior	1245.767	0.000
influencia_convidado_contribuir_reunioes	1237.534	0.000
influencia_envolvido_atividades	1227.645	0.000
influencia_comunicacao_superior	1136.713	0.000
influencia_supervisor_opinioao	1106.821	0.000
influencia_sentimento	878.432	0.000
influencia_colegas_partilham_info	864.248	0.000
lideranças_desenvolveram_oportunidades	428.435	0.000
lideranças_auxiliaram_desenvolvimento	355.662	0.000
facil_exercer_lideranca	299.291	0.000
influencia_raramente_convivio	281.326	0.000
suporte_lideranças_branças	237.144	0.000
influencia_ultimo_ter_conhecimento	189.229	0.000
todas_bemvindas	183.166	0.000
sempre_maximo_reconhecido	174.693	0.000
influencia_supervisor_nao_partilha	173.683	0.000
igualdade_negras_branças	165.538	0.000
lideranças_sensíveis_inclusão_racial	155.741	0.000
nao_necessita_medidas_inclusivas	111.510	0.000
acredito_fazer_mais	85.514	0.000

acredito_cauteloso	55.873	0.000
acredito_submetido_desconfortaveis	50.769	0.000
influencia_meu_grupo	40.563	0.000
facil_branças_integrar	32.331	0.000
percebo_facil_branco_lider	26.250	0.000

### Apêndice III Distribuição por gênero × etnia nas quatro dimensões

	Subgrupo (gênero × etnia)	Inclusão Organizacional	Cultura Inclusiva	Microagressões	Apoio Lideranças	Síntese interpretativa
<b>1 – “Líderes Vigilantes”</b>	<b>Homens (brancos)</b>	4,00	4,43	4,25	4,25	Elevada integração formal e reconhecimento; forte consciencialização crítica.
	<b>Mulheres (brancas)</b>	3,98	4,33	3,73	4,11	Níveis igualmente altos, mas reportam microagressões ligeiramente menores que os homens.
	<b>Homens (negros)</b>	4,08	4,03	3,66	4,06	Inclusão elevada; percebem menos apoio simbólico que brancos, embora próximo.
	<b>Mulheres (negras)</b>	4,20	3,61	3,46	3,71	Maior percepção de microagressões; inclusão formal ainda alta mas reconhecimento simbólico inferior.
	<b>Mulheres (outras etnias)</b>	4,10	4,03	3,83	4,09	Experiência positiva, em linha com média do grupo; microagressões mais salientadas.
	<b>2 – “Potenciais em Resistência”</b>	<b>Mulheres (brancas)</b>	1,33	1,94	1,69	1,69
	<b>Homens (brancos)</b>	1,45	2,71	2,29	2,15	Inclusão baixa; microagressões moderadas; déficit de apoio institucional (≈2,2).
	<b>Mulheres (negros)</b>	1,38	2,79	2,68	2,42	Vulnerabilidade acrescida: menor inclusão — microagressões elevadas, escasso apoio.

	<b>Subgrupo (género × etnia)</b>	<b>Inclusão Organizacio nal</b>	<b>Cultura Inclusiva</b>	<b>Microa gressões</b>	<b>Apoio Lideranças</b>	<b>Síntese interpretativa</b>
	<b>Homens (negros)</b>	1,50	2,13	1,89	1,55	Pior cenário absoluto de apoio ( $\leq 1,6$ ); sensação de invisibilidade institucional.
	<b>Mulheres (outras etnias)</b>	1,37	2,88	2,70	2,21	Experiência de exclusão estrutural; microagressões evidentes; apoio incipiente.
<b>3</b>	<b>Mulheres (brancas)</b>	2,99	3,03	3,01	2,71	Experiência mediana; percepções homogêneas às de outros subgrupos.
	<b>Homens (brancos)</b>	3,00	3,25	3,13	3,03	Ligações institucionais razoáveis, mas sem reconhecimento pleno.
	<b>Mulheres (negros)</b>	2,90	3,17	3,09	2,97	Níveis muito próximos dos homens negros; microagressões ligeiramente superiores (+0,15).
	<b>Homens (negros)</b>	3,10	3,03	2,89	2,64	Inclusão moderada; menor apoio comparativamente às mulheres brancas.
	<b>Homens (outras etnias)</b>	2,94	3,31	3,10	3,03	Perfil intermédio; indicador de cultura inclusiva ligeiramente superior à média do grupo.