

# UTILIZAÇÃO DE FERRAMENTA DE CLUSTERIZAÇÃO PARA ALINHAR A ESTRATÉGIA DE UMA EMPRESA ÀS NECESSIDADES DE GESTÃO DE SEUS TALENTOS

Bruna Villa Todeschini – UFRGS – bruna.todeschini@ufrgs.br

Cláudia Medianeira Cruz Rodrigues – UFRGS – claudia.rodrigues@producao.ufrgs.br

Michel José Anzanello – UFRGS – anzanello@producao.ufrgs.br

## Resumo

A busca por melhores práticas em gestão de pessoas é um tema que se torna cada vez mais presente em empresas de diversos segmentos, e visa, entre outros, entender o comportamento dos colaboradores para direcionar com mais assertividade os investimentos em capital humano. No entanto, pessoas com características diferentes podem receber estratégias corporativas de maneiras diversas. Este artigo descreve um estudo de clusterização por k-means realizado com um grupo de funcionários de uma empresa brasileira. As variáveis utilizadas referem-se a características individuais e demográficas dos indivíduos dessa população. Os resultados apontaram para a divisão da população em dois grupos que se diferenciam em relação a salários, idade, tempo de empresa e número de horas extras trabalhadas. Para cada um dos grupos, foram propostas estratégias de posicionamento da empresa em relação a seus colaboradores.

**Palavras-chave:** Comportamento Organizacional, People Analytics, Clusterização.

## 1. Introdução

Cada dia mais as empresas têm se preocupado em orientar suas ações para o bem-estar dos colaboradores. Neste sentido, a Qualidade de Vida no Trabalho vem sendo pesquisada desde a década de 70 e seus estudos vêm sendo aplicados em muitas áreas de atuação. Rossi *et al.* (2004) ressaltam que tomadores de decisão devem escolher como alocar recursos escassos para colocá-los em utilização otimizada, tendo em vista tanto a questão da qualificação técnica quanto os aspectos comportamentais, referindo-se, principalmente, às atitudes em ambiente de trabalho.

Entre os estudos da última década, o conceito de *People Analytics* (também conhecido por *Talent Analytics e Human Resources Analytics*), inicialmente utilizado em grandes empresas de tecnologia, foi suscitado com o objetivo de transformar os dados e as métricas de Recursos Humanos em conhecimento rigoroso e relevante para as empresas (BOUDREAU;

RAMSTAD, 2004). Esse conhecimento deve ser conduzido de maneira que o negócio tome decisões embasadas em informações relevantes acerca do potencial humano. Por isso, uma estrutura lógica deve garantir que as análises estejam centradas nas questões certas para a descoberta de dados e resultados que sejam acionáveis (LEVENSON, 2010).

Davenport *et al.* (2010) sugerem a existência de seis ramos em *People Analytics*: (i) o chamado Fatos do Capital Humano, trabalhando os principais indicadores de desempenho individual e de nível empresarial; (ii) Análises de Recursos Humanos, que busca juntar dados de diferentes naturezas para obter *insights* sobre departamentos ou funções específicas; (iii) Análise de Investimentos em Capital Humano, que busca identificar que tipos de ações geram maior impacto no desempenho do negócio; (iv) Previsão de Força de Trabalho, que analisa rotatividade, planos de sucessão e outros dados, a fim de identificar excessos ou escassez de mão de obra com antecedência; (v) Modelo de Valor para Talentos, que usa análises para calcular o que os funcionários mais valorizam e, assim, criar modelos que aumentem as taxas de retenção; e (vi) a Cadeia de Suprimentos de Talentos, que ajuda as empresas a tomarem decisões em tempo real sobre suas pessoas em relação a demandas como a otimização dos horários de trabalho, de acordo com a previsão de demanda em vendas em uma área comercial, por exemplo.

Entre os ramos sugeridos por Davenport *et al.* (2010), a Análises de Recursos Humanos e Análise de Investimento em Capital Humano são estudos complementares à medida que o primeiro gera bases de dados e as hipóteses iniciais para a aplicação do segundo. Davenport *et al.* (2010) ainda comentam que, em suas pesquisas, nenhuma organização que obteve sucesso na aplicação de *People Analytics* fez uso de apenas um dos métodos.

Boudreau e Ramstad (2004) enfatizam que a dificuldade atual reside em criar uma estrutura de organização e um sistema de Gestão de Pessoas que envolvam habilidades além daquelas comuns ao tradicional modelo de prestação de serviços de Recursos Humanos onde elas possam ter o maior efeito. Nesse sentido, é necessário buscar informações além dos domínios de gestão de pessoas, a fim de entender a lógica de funcionamento de cada organização para propor soluções à gestão de talentos.

Dessa forma, o objetivo principal deste estudo é identificar padrões de características em grupos de colaboradores de uma empresa, a fim de propor estratégias por parte da empresa adequadas às necessidades de cada um desses grupos, proporcionando condições para

melhoria de desempenho e de qualidade de vida no trabalho. A proposta é identificar esses grupos através de um conjunto de dados de colaboradores organizados por variáveis que representam características individuais e demográficas (DUBRIN, 2003) para gerar *clusters* de indivíduos através de análise de agrupamentos (HAIR *et al.*, 2010). As características comuns dos indivíduos inseridos nesses grupos são avaliadas para identificar padrões de comportamento e gerar iniciativas de melhoria em desenvolvimento humano.

É necessário ressaltar a importância de estudos como o presente, visto que cada vez mais as empresas tentam manter bons colaboradores proporcionando-lhes bem-estar, benefícios e investimentos em educação. A identificação de *clusters* entre os funcionários por meio de uma técnica estatística pode gerar o embasamento necessário para a adequação de planos de desenvolvimento, alinhando as expectativas entre empresa e colaborador.

Este artigo está organizado da seguinte maneira: após a contextualização do tema e da definição do objetivo do artigo na Introdução, segue a seção 2 que apresenta a revisão da literatura sobre o assunto abordado, tanto no que se refere à problemática de gestão de pessoas, quanto no que se refere à técnica de clusterização. Os procedimentos metodológicos são apresentados na seção 3, seguidos da descrição dos resultados do estudo na seção 4. Na seção 5, é apresentada a conclusão do artigo com uma discussão sobre os resultados obtidos.

## **2. Referencial Teórico**

Esta seção apresenta uma revisão bibliográfica sobre os tópicos tratados na pesquisa, que estão divididos em três subseções: a primeira apresenta aspectos do comportamento organizacional, a segunda comenta a importância das características avaliadas em cada indivíduo e a terceira apresenta fundamentos das ferramentas de clusterização.

### **2.1. Comportamento Organizacional**

O conceito de comportamento organizacional, que vem sendo estudado desde a década de 40, é apresentado por Wagner e Hollenbeck (2010) como um campo de estudo que busca compreender, explicar, prever e mudar o comportamento humano no contexto organizacional. Essa linha da ciência é dividida em três dimensões: comportamentos micro, meso e macro-organizacionais (WAGNER; HOLLENBECK, 2010). A dimensão micro estuda o comportamento individual dos trabalhadores, enquanto a dimensão meso tem seu foco no entendimento das relações entre pessoas trabalhando em times e grupos. Já o estudo do macrocomportamento organizacional estuda os comportamentos das organizações como um

todo. Robbins *et al.* (2003) acrescentam uma quarta dimensão, o ambiente externo que interage com a organização.

Esta pesquisa está focada na dimensão do microcomportamento organizacional, pois irá trabalhar com variáveis individuais de funcionários de uma empresa a fim de identificar características similares em conjuntos a serem formados analiticamente, gerando base para a tomada de decisão e sugerindo planos de ação em relação aos grupos para criar condições de se ter um melhor desempenho geral da empresa. Cascio e Boudreau (2010) enfatizam que as análises em Recursos Humanos são importantes à medida que tornam melhores as decisões vitais sobre talentos e sobre como eles são organizados. Nesse sentido, Chi *et al.* (2009) mostram que clientes tendem a ter uma melhor experiência com empresas que possuem funcionários com níveis elevados de satisfação e engajamento, enquanto que Koys (2001) relata que o comportamento de maneira cidadã por parte dos funcionários influencia a rentabilidade das organizações. Além dessas características, outras também são relevantes ao estudo do nível individual de comportamento organizacional, apresentadas na próxima seção.

## **2.2. Características dos funcionários**

DuBrin (2003) afirma que compreender diferenças entre empregados pode levar a melhorias relevantes na produtividade individual e organizacional. Assim, ele apresenta as principais fontes de diferenças individuais e demográficas entre colaboradores. Diferenças individuais dizem respeito à maneira como as pessoas reagem a uma mesma situação com base em características pessoais; já as diferenças demográficas trabalham com as diversidades em fatores históricos relacionados à força de trabalho que ajudam a moldar atitudes e comportamentos dos trabalhadores (DUBRIN, 2003).

Como principais diferenças individuais, DuBrin (2003) apresenta: (i) produtividade; (ii) habilidade e talento; (iii) propensão a alcançar resultados de alta qualidade; (iv) maneira como as pessoas querem ser empoderadas e envolvidas; (v) estilo de liderança; (vi) necessidade de contato com os outros; (vii) comprometimento e lealdade à empresa; e (viii) autoestima. Em relação às diferenças demográficas, ele apresenta: (i) sexo; (ii) idade e experiência; (iii) etnia; e (iv) estado de deficiência.

Dessa maneira, os conceitos das diferenças levantadas podem ser utilizados na análise dos grupos a serem formados. Campos (1992) refere que aumentar a produtividade significa produzir mais e melhores saídas em um processo com cada vez menos insumos. Hunter *et al.* (1990) afirmam que, quanto maior o nível de complexidade de uma tarefa executada por

trabalhadores, maior a variabilidade em relação à produtividade dessa tarefa. Nesse sentido, em relação às diferenças de habilidade, Robbins *et al.* (2003) colocam que essa característica influencia diretamente o nível de desempenho e a satisfação de um funcionário de acordo com a adaptação entre habilidade e trabalho exercido.

Em relação às divergências entre trabalhadores sobre a propensão que cada um possui em alcançar resultados de alta qualidade, DuBrin (2003) enfatiza que algumas pessoas tendem com naturalidade a buscar fortemente a alta qualidade porque são conscienciosas, têm uma boa capacidade para serem precisas e se orgulham de seus trabalhos. Da mesma forma, trabalhadores menos conscienciosos, menos precisos e que se orgulham pouco de seu trabalho tendem a ter maiores dificuldades para atingir resultados com o nível de qualidade dos primeiros.

Sobre o modo como as pessoas querem ser empoderadas e envolvidas, Horochovski e Meirelles (2007) mencionam que empoderamento é um conceito similar à autonomia, pois trata da capacidade de indivíduos e grupos decidirem sobre as questões que lhes dizem respeito e escolher entre cursos de ação alternativos. No entender de DuBrin (2003), é necessário incentivar os trabalhadores a sugerirem melhorias e dar mais autonomia para que tomem decisões.

DuBrin (2003) ainda refere que as pessoas diferem no estilo de liderança que preferem e de que necessitam. Muitos indivíduos buscam a maior liberdade possível dentro de um cargo e sabem lidar com isso, enquanto outras pessoas preferem ter seus gestores sempre por perto. Do mesmo modo, as pessoas também necessitam de supervisão em níveis diferentes: geralmente, trabalhadores menos competentes, menos experientes e menos motivados precisam de uma supervisão mais constante do que outros funcionários.

Sobre a necessidade de contato com os outros, Soto (2002) define o conceito de sociabilidade como o desejo de compartilhar atividades e receber dos demais a atenção e estimulação que é parte da interação social. Ser sociável supõe valorizar intrinsecamente o processo de interagir com outras pessoas.

Em relação ao comprometimento com a empresa, um estudo de Klehe *et al.* (2011) mostra que, entre uma série de fatores testados, apenas a exploração do trabalho confirma uma redução da lealdade dos trabalhadores à empresa. DuBrin (2003), porém, afirma que existem trabalhadores mais comprometidos com seus empregadores, estando atentos em produzir bens e serviços com alta qualidade e tendo uma característica de pontualidade,

enquanto outros não se sentem comprometidos, não sentindo culpa quando geram refugos ou faltam ao trabalho por motivos fúteis. DuBrin (2003) ainda refere que a autoestima influencia a produtividade e a capacidade de aceitar responsabilidades adicionais, uma vez que leva as pessoas a acreditarem que conseguem carregar consigo desafios básicos da vida e que podem ser dignos de felicidade.

Sobre as diferenças demográficas, são citados sexo, idade, experiência, etnia e estado de deficiência. Em relação às diferenças de sexo, experimentos realizados por Powell (2011) concluíram que as mulheres têm tendência a ter um estilo mais interpessoal do que os homens ao lidar com colegas de trabalho. Além disso, tendem a ser mais democráticas e, quando em posição de liderança, tendem a ser menos autocráticas em todas as circunstâncias avaliadas. Mesmo assim, o estudo mostra que a maneira de realizar uma mesma tarefa não difere de homens para mulheres.

Mahlberg *et al.* (2013) testaram hipóteses sobre a relação de envelhecimento e produtividade. O estudo não apontou indícios de que envelhecimento leva necessariamente a um declínio da produtividade no trabalho. DuBrin (2003) ressalta que avaliar idade e experiência tem se tornado relevante, tendo em vista que a taxa de inserção de pessoas na população economicamente ativa em grande parte dos países ocidentais não corresponde à taxa de aposentadoria dos mais velhos.

No que se trata de etnias, é apresentado o conceito de racioetnicidade, que é o termo corrente utilizado para referir uma série de diferenças étnicas e raciais. DuBrin (2003) fala que, no desempenho do trabalho e no comportamento das pessoas, essas diferenças são mais atribuíveis à cultura do que à experiência étnica ou racial em si.

Sobre diferenças em relação a deficiências, a World Health Organization (1993) conceitua deficiência como perda ou anomalia de estrutura ou função psicológica, fisiológica ou anatômica temporária ou permanente. É importante atentar para esse aspecto, visto que a deficiência, entre elas a mental, está ganhando atenção como fonte de diversidade no local de trabalho (DUBRIN, 2003).

Com as diferenças individuais e demográficas entre colaboradores apresentadas, é possível utilizar variáveis que as representem para realizar a análise de agrupamentos. Na próxima seção, são apresentados conceitos sobre análise multivariada de formação de *clusters*.

### **2.3. Ferramentas de Formação de *Clusters***

Segundo Hair *et al.* (2010), a análise de agrupamentos (*cluster analysis*) é um grupo de técnicas multivariadas cuja finalidade principal é agrupar objetos com base nas suas características. Seu critério essencial é o de tentar maximizar as diferenças entre os grupos em confronto com a variação dos mesmos (CORRAR *et al.*, 2009).

Os métodos de análise de agrupamentos são divididos em dois tipos: hierárquicos e não hierárquicos. Hair *et al.* (2005) afirmam que os procedimentos hierárquicos são procedimentos de agrupamentos apoiados em fases e que envolvem uma combinação ou divisão dos objetos em agrupamentos. Já os procedimentos não hierárquicos atribuem objetos a um grupo uma vez especificado o número de grupos a ser formado (CORRAR *et al.*, 2009).

Os métodos não hierárquicos levam vantagem sobre os hierárquicos quando se trabalha com conjuntos grandes de dados, visto que não requerem o cálculo de matrizes de semelhança entre todas as observações, apenas o cálculo da semelhança entre cada observação e o centroide do *cluster*. Além disso, os métodos não hierárquicos são menos suscetíveis aos *outliers*, à medida de distância usada, e à inclusão de variáveis irrelevantes ou inapropriadas (HAIR *et al.*, 2010).

Os algoritmos não hierárquicos do tipo *k-means* geram um número especificado pelo usuário de porções (grupos) de dados e, em seguida, de forma iterativa, alocam observações aos clusters até que algum critério numérico seja cumprido (HAIR *et al.*, 2010). Cada uma dessas observações contém valores para cada variável utilizada na caracterização da amostra a ser clusterizada. Para rodar o algoritmo, são usadas medidas de distâncias entre as observações e entre os grupos formados, dentre as quais se destaca distância Euclidiana (distância de uma linha reta traçada entre dois pontos).

Para avaliar a qualidade dos agrupamentos formados através das ferramentas de clusterização, Rousseeuw (1987) apresenta o conceito de silhuetas para interpretar e validar resultados de análises de *clusters*. Dessa forma, o *Silhouette Index* (SI), apresentado na equação (1), avalia quanto um objeto é similar aos demais objetos em seu grupo se comparado aos objetos alocados no grupo mais próximo a ele (STROIEKE *et al.*, 2013). Quanto mais elevado o valor do índice, melhor a alocação do objeto em seu *cluster* de destino.

$$SI(i) = \frac{b(i) - a(i)}{\max\{b(i), a(i)\}}, \quad (1)$$

onde  $a(i)$  é a distância média da observação  $i$  em relação às demais observações do grupo em que foi alocado e  $b(i)$  é a distância média da observação  $i$  em relação às observações do grupo

vizinho mais próximo. Assim, o valor de  $SI(i)$  varia de -1 a +1, sendo que valores de SI próximos a +1 denotam observações adequadamente inseridas no cluster de destino.

A análise de agrupamentos por *k-means* tem sido utilizada no reconhecimento de padrões de comportamento humano. Pimentel *et al.* (2003) fizeram uso da técnica para identificar grupos de aprendizes em ensino presencial e possibilitar a utilização de um sistema tutor inteligente, enquanto que Horst *et al.* (2012) utilizaram *k-means* para avaliar a satisfação de funcionários em um hospital. Stroeke *et al.* (2013) utilizaram o algoritmo para gerar grupos homogêneos de trabalhadores em curvas de aprendizado; os resultados foram aferidos através do *Silhouette Index*.

### **3. Procedimentos Metodológicos**

Esta etapa apresenta a descrição do cenário em que foi aplicado o estudo e a caracterização dos métodos de pesquisa e de trabalho.

#### **3.1 Descrição do Cenário**

O presente estudo foi realizado em um grupo de empresas brasileiras que se caracteriza por ser de grande porte segundo critério do BNDES (2013), com empresas em quatro estados das regiões sudeste e sul. O grupo conta com empresas em diversos ramos, tendo destaque comunicações e tecnologia da informação.

Por ser um grupo de grande porte, conta hoje com uma estrutura de mais de 5000 colaboradores. Dessa forma, o estudo está focado na área de Recursos Humanos, que hoje centraliza os processos relacionados a pessoas de todo o grupo. A área lidera três macroprocessos: (i) encontrar pessoas; (ii) gerir pessoas; e (iii) desenvolver pessoas. A fim de contribuir com esses macroprocessos, um método de formação de agrupamentos foi utilizado para buscar a identificação de padrões no comportamento e nas características de colaboradores da empresa. Com os grupos identificados e analisados, foi feita uma proposição de ações a serem realizadas visando encontrar, gerir e desenvolver pessoas com mais eficiência e eficácia.

#### **3.2 Caracterização do Método de Pesquisa**

O estudo está classificado segundo quatro critérios: natureza, abordagem, objetivos e procedimentos. Em relação à natureza, é uma pesquisa aplicada, uma vez que investiga hipóteses sugeridas por modelos sobre comportamento organizacional em um caso real. Sobre a abordagem, segue em duas linhas: uma que busca caracterizar comportamentos de pessoas,



tendo um caráter qualitativo, e outra que busca comprovar esses comportamentos por meio de análises quantitativas. Tem, portanto, uma abordagem quantiquantitativa, comum entre pesquisas sociais aplicadas (GOMES; ARAÚJO, 2005). No que se refere aos objetivos, caracteriza-se como uma pesquisa exploratória, uma vez que o uso de uma técnica de análise multivariada de dados atrelada ao comportamento de indivíduos numa organização a fim de identificar padrões e propor melhorias na gestão de pessoas é um tema pouco estudado. Em relação aos procedimentos, é um estudo de caso, já que analisa, utiliza dados e coleta informações de uma situação individual.

### **3.3 Caracterização do Método de Trabalho**

O estudo foi realizado em cinco etapas: (i) coleta de dados; (ii) tratamento dos dados; (iii) geração e avaliação de qualidade dos agrupamentos; (iv) interpretação dos *clusters* formados; (v) proposição de estratégias de posicionamento da empresa em relação aos grupos formados.

A primeira etapa consiste em escolher e coletar dados que representam as diferenças individuais e demográficas propostas por DuBrin (2003). Essas informações foram coletadas de bases de dados de sistemas utilizados pela área de Recursos Humanos da empresa em estudo e são necessárias para representar as características da dimensão micro-organizacional, que trata dos colaboradores individualmente, na formação do comportamento organizacional (WAGNER; HOLLENBECK, 2010). A coleta foi feita com dados de colaboradores de todas as empresas do grupo para todos os níveis hierárquicos disponíveis e com dados de avaliação de desempenho válidos. As bases de dados foram organizadas em matrizes, onde cada coluna representa uma variável de interesse e cada linha representa um colaborador.

A segunda etapa consistiu em realizar o tratamento dos dados. Essa etapa teve início na filtragem de linhas da base geral em que todos os campos estavam preenchidos, facilitando a análise posterior. A partir dessa base de dados, foram testadas duas técnicas para tratamento dos dados: padronização e normalização dos dados. Ambas objetivaram minimizar problemas na análise de variáveis com diferentes magnitudes, onde uma escala maior pode ofuscar a influência de escalas menores na medição de distâncias realizadas pela sistemática de clusterização. Na padronização dos dados, são escolhidos os valores máximo e mínimo do conjunto de variáveis; valores máximos recebem o escore 1 e os mínimos recebem 0. Os valores intermediários recebem escores ponderados entre esses dois valores. Já na

normalização, calcularam-se média e desvio padrão para dar origem ao escore  $Z$ , que substituiu o dado de cada indivíduo para cada uma das variáveis analisadas.

A terceira etapa apoia-se em dois procedimentos (tanto para o conjunto de dados normalizados quanto para os dados padronizados): geração de *clusters* e avaliação da qualidade desses agrupamentos. A partir da ferramenta de agrupamento *k-means* (HAIR *et al.*, 2010), foram realizadas rodadas de clusterização no *software* R (pacote *cluster*) para vários valores de  $k$  ( $2 \leq k \leq 11$ ), que representa o número de grupos formados. Avaliou-se, para cada valor de  $k$ , a qualidade do arranjo gerado pelo *Silhouette Index* (SI) (ROUSSEEUW, 1987). O arranjo que apresentou o maior valor de SI foi escolhido. Essa rotina foi usada com as bases de dados completas e testando-se a retirada de variáveis do banco, a fim de verificar eventuais incrementos no valor de SI que justificassem (juntamente com a análise do contexto da empresa) a retirada de determinadas variáveis.

A quarta etapa consistiu na interpretação dos resultados da clusterização. Ela foi realizada identificando as características preponderantes internas de cada grupo, explicitando os fatores de maior relevância e similaridade entre os dados. Esta etapa tem característica empírica de mineração de dados, buscando entender os padrões de comportamento dos indivíduos inseridos nos grupos formados.

Na última etapa, propuseram-se estratégias de posicionamento da empresa em relação aos grupos identificados na quarta etapa. Essas estratégias foram fundamentadas no diagnóstico gerado na etapa anterior, buscando melhorias em relação a bem-estar no trabalho, alinhamento à cultura da empresa, treinamentos e plano de remuneração, entre outros. O conjunto de ações foi avaliado em uma priorização por dois aspectos: necessidade de retenção de talentos (funcionários com alto desempenho e alto potencial de desenvolvimento) e viabilidade da ação.

#### 4. Resultados

O banco de dados utilizado no estudo foi composto de 2152 observações (colaboradores) e sete variáveis que representam características individuais e demográficas propostas por DuBrin (2003). A descrição das variáveis segue na Tabela 1.

**Tabela 1:** Apresentação de variáveis a serem utilizadas na análise de agrupamentos

Variável (código)	Significado	Diferenças individuais e demográficas (DUBRIN, 2003) relacionadas	Unidade de Medida
----------------------	-------------	---	-------------------------

Satisfação (satisf)	Representa a satisfação geral do colaborador em relação à empresa.	Diferenças individuais de necessidade contato com os outros, comprometimento e lealdade à empresa e autoestima	-
Horas Extras (hextras)	Representa o número de horas extras realizadas no mês utilizado para a análise (proporcional aos minutos realizados).	Diferenças individuais de produtividade e habilidade e talento	horas
Idade (idade)	Representa a idade do colaborador proporcional ao número de dias de vida no último dia do mês utilizado para a análise.	Diferenças demográficas de idade e experiência	anos
Nota da Avaliação de Desempenho (mhct)	Representa a nota dada ao colaborador por seu gestor e um comitê de Recursos Humanos sobre a avaliação de desempenho do colaborador no trimestre anterior ao utilizado para a análise.	Diferenças individuais de produtividade e propensão a alcançar resultados de alta qualidade	-
Remuneração Total (salário)	Representa a soma de remuneração fixa e variável do colaborador no mês utilizado para a análise.	Diferenças individuais de produtividade, comprometimento e lealdade à empresa, habilidade e talento, diferenças demográficas de idade e experiência	reais
Compa-ratio (comp_out)	É a comparação do salário do colaborador com a mediana de mercado, ou seja, comparação do salário da pessoa com o valor praticado por outras empresas para a mesma função exercida.	Diferenças individuais de habilidade e talento e diferenças demográficas de experiência	percentual
Tempo de Empresa (tempo_empr)	Representa o período de tempo iniciado na data de admissão do colaborador na empresa e finalizado na data de demissão (para colaborador demitido na data da análise) ou no último dia do mês utilizado para a análise (para colaborador contratado na data da análise) proporcional ao número de dias do período.	Diferenças demográficas experiência e diferenças individuais de comprometimento e lealdade à empresa	anos

Na sequência, foi verificada a correlação entre as variáveis utilizadas com o intuito de avaliar a necessidade de retirar alguma variável que tivesse um comportamento semelhante a outra no banco (alta correlação). Os resultados seguem na Tabela 2.

**Tabela 2:** Medidas de correlação entre as variáveis utilizadas no banco de dados

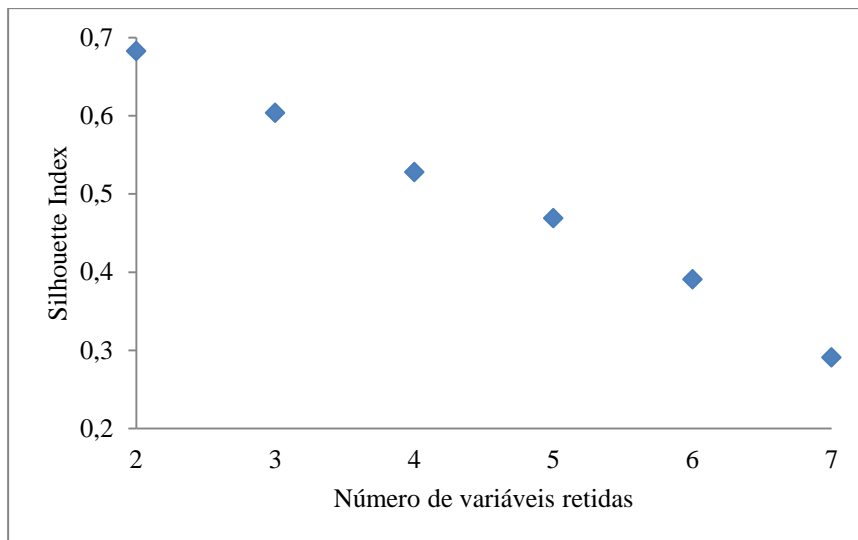
Correlação	Nota da Avaliação de Desempenho	Remuneração Total	Compa-ratio	Idade	Satisfação	Tempo de Empresa	Horas Extras
Nota da Avaliação de Desempenho	1,0000						
Remuneração Total	0,1763	1,0000					
Compa-ratio	0,0759	0,1805	1,0000				
Idade	-0,0245	0,2156	0,4204	1,0000			

<b>Satisfação</b>	0,0883	0,0806	0,0717	0,1341	1,0000		
<b>Tempo de Empresa</b>	0,0165	0,1177	0,2658	0,6919	0,0638	1,0000	
<b>Horas Extras</b>	-0,0226	-0,0911	-0,1056	-0,0649	-0,0816	-0,1055	1,0000

A Tabela 2 mostra que não há correlações próximas a +1 ou -1, o que demonstra que o banco apoia-se em variáveis que apresentam características diversas entre si (pouco correlacionadas). No caso das variáveis Tempo de Empresa e Idade, nota-se uma correlação maior que as demais; optou-se, porém, pela manutenção de ambas por tratarem de características distintas no comportamento micro-organizacional (especificamente, características individuais e demográficas). Ainda com base nos valores de correlação apresentados pelas variáveis, entende-se como dispensável o uso de ferramentas com o propósito de transformar variáveis para diminuir a correlação entre elas, como a técnica de Análise dos Componentes Principais, entre outras.

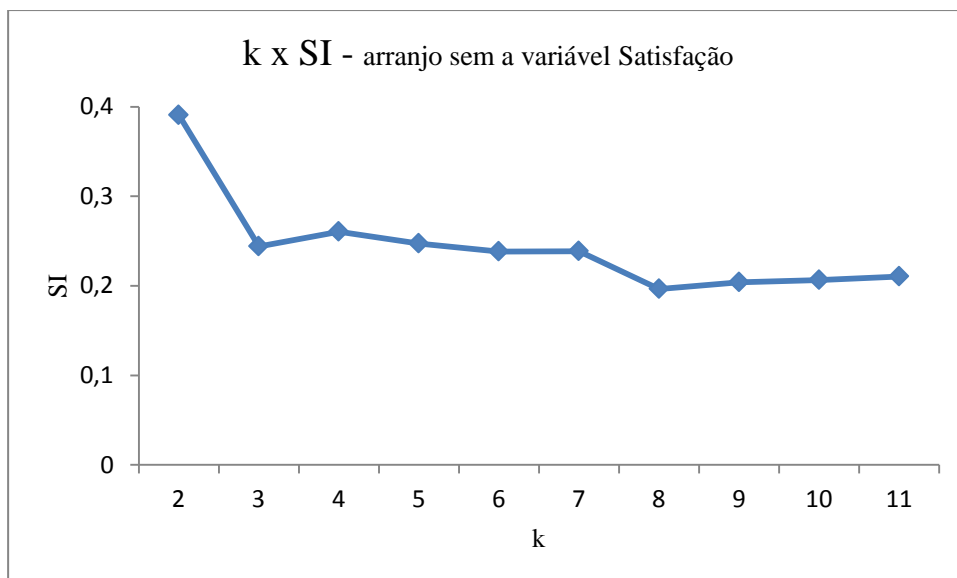
O tratamento da base de dados anterior à clusterização foi realizado de duas maneiras: padronização e normalização dos dados. Os resultados das análises de SI mostraram que os dados padronizados tiveram desempenho superior aos normalizados, tanto no caso em que foram utilizadas todas as variáveis, quanto nos casos em que foram feitas tentativas de retirada de variáveis. Essas tentativas foram realizadas em fases: inicialmente, com as sete variáveis no modelo, foram feitos testes com a omissão de todas as variáveis, uma a uma. O modelo que gerou o maior SI foi o escolhido e o procedimento foi repetido até restar apenas duas variáveis no modelo. A Figura 1 mostra os valores de SI para os dados padronizados conforme o número de variáveis remanescentes no banco.

Considerando o incremento no SI e o contexto das variáveis na empresa, foi feita a retirada de uma variável: Satisfação. Essa variável, além de ter sido a responsável pelo maior incremento no SI ao ser retirada, apresenta aspectos negativos por ser menos objetiva do que as outras e por ter necessidade de melhorias na pesquisa de satisfação já identificadas pela empresa.

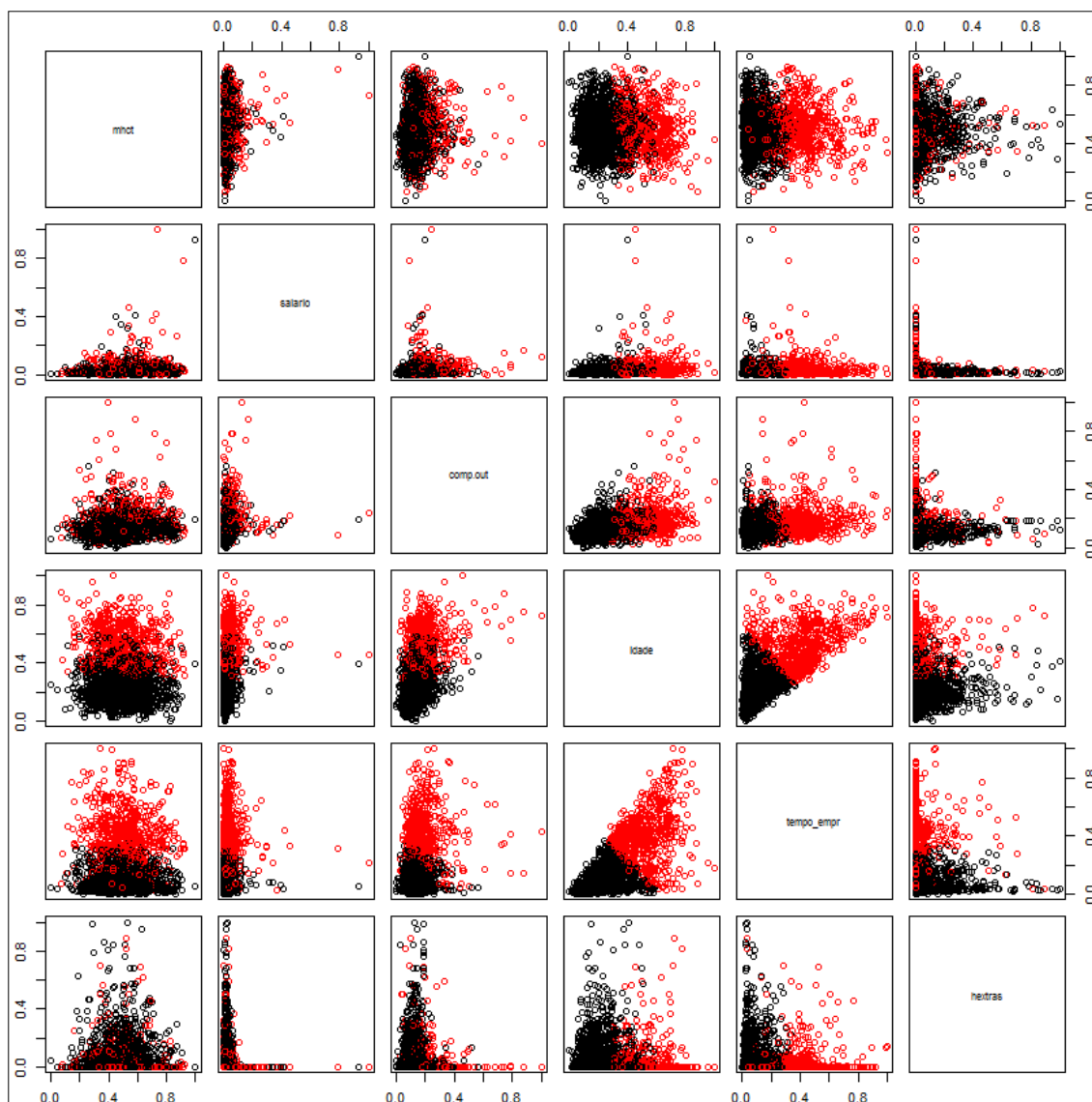


**Figura 1:** Valores de Silhouette Index para o número de variáveis retidas na clusterização

De tal forma, permaneceram no banco as seguintes variáveis: nota de avaliação de desempenho, idade, compa-ratio, tempo de empresa, remuneração total e horas extras. Dessa maneira, a figura 2 mostra os valores de SI para cada valor de k no arranjo em que essas variáveis permaneceram. Assim, se chegou a um arranjo de dois grupos ( $k=0,3907$ ), representados na Figura 3 pelas cores vermelha e preta, que mostra a dispersão dos grupos para os pares de variáveis utilizadas.



**Figura 2:** Valores de SI para cada número de grupos formados na clusterização sem a variável Satisfação



**Figura 3:** Dispersões dos agrupamentos gerados pela clusterização

Com as análises realizadas, foi possível perceber que um dos grupos (representado pela cor vermelha na Figura 3) é formado por colaboradores com maior idade, que realizam menos horas extras e têm mais tempo de empresa. Além disso, percebe-se uma tendência a apresentar maiores salários do que o outro grupo.

Dadas essas características, espera-se um posicionamento da empresa em relação aos funcionários avaliados. Para o grupo de cor vermelha, por ser um grupo com características mais seniores, sugerem-se algumas decisões gerenciais:

- (i) Promover uma política de previdência privada mais atraente do que para os funcionários do outro grupo. Assim, essas pessoas poderiam investir mais em suas carreiras dentro da empresa e a organização teria mais segurança em oferecer a essas pessoas vagas consideradas críticas no seu modelo de gestão;

- (ii) Gerar uma política de remuneração variável de maior impacto, dando mais peso na remuneração total para a remuneração variável e menos para a remuneração fixa do que nos dias atuais. Assim, existe mais um motivo para promover o engajamento desse público nas questões financeiras da empresa; e
- (iii) Realizar um acompanhamento de carreira personalizado para aqueles com maiores notas de desempenho, visando conferir mais segurança acerca dos caminhos que essas pessoas pretendem seguir em suas carreiras.

Por sua vez, o grupo apresentado em preto na Figura 3 tem um perfil mais jovem, que realiza mais horas extras, menos tempo de empresa e salários menores do que o grupo vermelho. Em relação a esse grupo, sugerem-se as seguintes medidas:

- (i) Possibilidade de realização de um programa de *mentoring* para os indivíduos do grupo. A decisão em realizar ou não viria do próprio colaborador e os mentores seriam colegas de empresa com mais experiência nos assuntos a serem focados no programa;
- (ii) Investimento mais forte do que é feito hoje em programas de estágio e *trainee*; e
- (iii) Capacitações mais recorrentes nos assuntos de interesse desse grupo de funcionários. Isso ajudaria a compensar o fato de estarem recebendo salários proporcionalmente menores do que o outro grupo.

A implementação de ações com o objetivo de alinhar a estratégia da empresa em relação a esses grupos de funcionários requer a definição de um plano de ação envolvendo a equipe de recursos humanos, gestores, funcionários e apoiadores externos. Assim, a partir dessas definições, poderão ser iniciadas as ações relativas às propostas realizadas nesse estudo.

## 5. Conclusões

Esse estudo buscou identificar padrões de características comuns a grupos de colaboradores de uma empresa. A proposta objetivou, com os padrões identificados, auxiliar na tomada de decisão sobre estratégias de gestão de pessoas mais adequadas às necessidades de cada grupo, bem como gerar subsídios para proporcionar um aumento na qualidade de vida no trabalho e melhores desempenhos individuais e de equipe.

O método proposto apoiou-se na clusterização não hierárquica por *k-means* e no *Silhouette Index* como índice de qualidade dos arranjos formados. De posse de uma base de

dados previamente definida, foram inicialmente realizados testes para definir o formato mais adequado dos dados de entrada (escala normalizada ou padronizada) com vistas a favorecer o processo de formação dos grupos. Definidos os agrupamentos, passou-se então à interpretação dos membros integrantes dos clusters e proposição de estratégias da empresa em relação aos grupos formados.

O método proposto foi aplicado em um banco de dados oriundo de um grupo de empresas associadas a comunicações e tecnologia da informação. Os testes de retirada de variáveis mostraram que os dados padronizados tiveram desempenho superior aos normalizados. Essa comparação de escalas contribuiu para a análise dos arranjos em duas frentes: na elevação da qualidade da clusterização mediada através do *Silhouette Index*, e na análise visual dos arranjos formados. Durante os testes de retirada de variáveis realizados, percebeu-se que a variável Satisfação não colaborava expressivamente na identificação de grupos de colaboradores. Esse fato, somado ao conhecimento prévio do contexto da empresa, determinou a retirada dessa variável da base de dados. A clusterização resultou na geração de dois grupos de colaboradores. Um deles é composto por colaboradores com maiores idade, tempo de empresa e salário e que realizam menos horas extras em relação ao outro grupo. Para cada grupo, foi gerada uma série de estratégias alinhadas com o perfil identificado.

Em trabalhos futuros, sugere-se utilizar variáveis categóricas que representam características micro-organizacionais. Para tanto, torna-se necessário utilizar uma ferramenta de clusterização ou medida de distância que se ajuste ao uso de variáveis contínuas e categóricas num mesmo arranjo.

## **Referências**

BOUDREAU, John; RAMSTAD, Peter. **Talentship and Human Resource Measurement and Analysis: From ROI to Strategic Organizational Change**. CEO Publication, 2004.

CAMPOS Vicente F. **Controle de Qualidade Total**. Fundação Christiano Ottoni, 1992.

CASCIO Wayne, BOUDREAU John. **Investimento em pessoas: como medir o impacto financeiro das iniciativas em recursos humanos**. Bookman, 2010.

CHI Christina G, GURSOY Dogan. **Employee satisfaction, customer satisfaction, and financial performance: an empirical examination**. International Journal of Hospitality Management, vol 28, pgs 245–253, 2009.



CORRAR Luiz J, PAULO Edilson, DIAS FILHO José M. **Análise multivariada: para os cursos de administração, ciências contábeis e economia.** Atlas, 2009.

DAVENPORT, Thomas; HARRIS, Jeanne; SHAPIRO, Jeremy. **Competing on Talent Analytics.** Harvard Business Review, 2010.

DUBRIN Andrew J. **Fundamentos do Comportamento Organizacional.** Thomson, 2003.

EVERITT Brian, HOTHORN Torsten. **An Introduction to Applied Multivariate Analysis with R.** Springer Science+Business Media, 2011.

GOMES Fabrício P, ARAÚJO Richard M. **Pesquisa Quantiquantitativa em Administração: uma visão holística do objeto em estudo.** Seminários em Administração FEA/USP, 2005.

HAIR Joseph F, ANDERSON Rolph E, TATHAM Ronald L, BLACK William C. **Análise Multivariada de Dados - 5ª edição.** Bookman, 2005.

HAIR Joseph F, BLACK William C, BABIN Barry J, ANDERSON Rolph E. **Multivariate Data Analysis - 7th edition.** Pearson, 2010.

HOROCHOVSKI Rodrigo R, MEIRELLES Giselle. **Problematizando o conceito de empoderamento.** Anais do II Seminário Nacional de Movimentos Sociais, Participação e Democracia, 2007.

HORST Fabio, MARTINS JUNIOR Celso, SOUZA Luzia V de. **Técnicas de Data-Mining e reconhecimento de padrões aplicada na predição do nível de satisfação dos colaboradores de um hospital na cidade de Guarapuava/PR.** Anais do Congresso de Matemática Aplicada e Computacional, 2012.

[http://www.bndes.gov.br/SiteBNDES/bndes/bndes\\_pt/Institucional/Apoio\\_Financeiro/porte.html](http://www.bndes.gov.br/SiteBNDES/bndes/bndes_pt/Institucional/Apoio_Financeiro/porte.html). Acesso em: 05 nov. 2013.

HUNTER John E, SCHMIT Frank L, JUDIESCH Michael K. **Individual differences in output variability as a function of job complexity.** Journal of Applied Psychology, vol 75, pgs 28-42, 1990.

KLEHE Ute-Christine, ZIKIC Jelena, VIANEN Annelies E M V, PATER Irene E D P. **Career adaptability, turnover and loyalty during organizational downsizing.** Journal of Vocational Behavior, vol 79, pgs 217-229, 2011.

KOYS Daniel. **The effects of employee satisfaction, organizational citizenship behavior and turnover on organizational effectiveness: a unit-level, longitudinal study.** Personnel Psychology, vol 54, pgs 101-114, 2001.

LEVENSON, Alec. **Using Targeted Analytics to Improve Talent Decisions.** CEO Publication, 2010.

MAHLBERG Bernhard, FREUND Inga, CUARESMA Jesús C, PRSKAWETZ Alexia. **Ageing, productivity and wages in Austria.** Labour Economics, vol 22, pgs 5-15, 2013.

PIMENTEL Edson P, FRANÇA Vilma F, OMAR Nizam. **A identificação de grupos de aprendizes no ensino presencial utilizando técnicas de clusterização.** Anais do XIV Simpósio Brasileiro de Informática na Educação, 2003.

POWELL, Gary N. **The Gender and Leadership Wars.** Organizational Dynamics, vol 40, pgs 1-9, 2011.

ROBBINS Stephen P, ODENDAAL Aletta, ROODT Gert. **Organisational Behaviour: Global and Southern African Perspectives.** Pearson, 2003.

ROSSI, Peter Henry; LIPSEY, Mark W; FREEMAN Howard E. **Evaluation: a systematic approach.** Seventh edition, Sage Publications, 2004.

ROUSSEEUW Peter J. **Silhouettes: a graphical aid to the interpretation and validation of cluster analysis.** Journal of Computational and Applied Mathematics, vol 20, pgs 53-65, 1987.

SOTO Eduardo. **Comportamento organizacional: o impacto das emoções.** Thomson, 2002.

STROIEKE Renato E, FOGLIATTO Flavio S, ANZANELLO Michel J. **Análise de conglomerados em curvas de aprendizado para formação de agrupamentos homogêneos de trabalhadores.** Produção vol 23, pgs 537-547, 2013.

WAGNER John A, HOLLENBECK John R. **Organizational Behavior: Securing Competitive Advantage.** Routledge, 2010.

WORLD HEALTH ORGANIZATION. **International classification of impairments, disabilities, and handicaps: a manual of classification relating to the consequences of disease.** 1993.