

RESUMO EXECUTIVO

Clusterização e elaboração de personas das MPE no Brasil

Sistema SEBRAE

Brasília - DF, 23 de Dezembro de 2022





Todos os direitos reservados.

A reprodução não autorizada desta publicação, no todo ou em parte, constitui violação aos direitos autorais (Lei nº 9.610).

Serviço Brasileiro de Apoio às Micro e Pequenas Empresas – SEBRAE

Unidade de Gestão Estratégica e Inteligência

SGAS 605 – Conjunto A – Asa Sul – Brasília/DF – CEP 70200-904

Tel.: 55 61 3348-7180

Site: <https://www.sebrae.com.br/>

CONSELHO DELIBERATIVO NACIONAL

Presidente

José Roberto Tadros

DIRETORIA EXECUTIVA

Diretor-Presidente

Carlos do Carmo Andrade Melles

Diretor Técnico

Bruno Quick Lourenço de Lima

Diretor de Administração e Finanças

Eduardo Diogo

Gerente da Unidade de Gestão Estratégica e Inteligência

Adriane Ricieri Brito

Gerente Adjunto da Unidade de Gestão Estratégica e Inteligência

Fausto Ricardo Keske Cassemiro

Coordenador do Núcleo de Pesquisa e Gestão do Conhecimento

Kennyston Costa Lago

Equipe Técnica

Tomaz Back Carrijo

Felipe Marcel Neves

Existem inúmeras vantagens em se dividir uma população em grupos menores, sendo que uma delas é facilitar a identificação e consequente caracterização dos subconjuntos gerados. A diminuição da variabilidade nos grupos facilita criar estratégias de atendimento diferenciadas para cada tipo de pessoa, podendo ser definidas metas, objetivos e ações mais específicos. Uma das formas de segmentar dados é através da utilização de modelos de aprendizado de máquina não-supervisionados, processo comumente chamado de clusterização, que consiste em treinar um modelo matemático para encontrar padrões em um conjunto de dados.

Com objetivo de fomentar o atendimento do Sebrae e aprimorar ações de comunicação focados no microempreendedor, o presente estudo apresenta o resultado da modelagem estatística para clusterização do perfil das MPE. A Pesquisa Perfil do MPE é bianual e está em sua 3ª edição. Seu público-alvo é composto por empresas de pequeno porte e microempresas localizados em todas as unidades federativas do Brasil que exercem atividades em todos os grandes setores do IBGE. É uma pesquisa quantitativa realizada por telefone (CATI), dividida em 32 perguntas fechadas e 4 abertas, com seleção aleatória dos entrevistados a partir de base de dados fornecida pelo Sebrae (período de coleta foi entre 10 de fevereiro a 5 de maio de 2022). Foram realizadas um total de 6.911 entrevistas distribuídas em cotas por unidade da federação, setor de atividade e ano de abertura (até 2019 e após 2020).

Em virtude de todas as variáveis utilizadas serem categóricas, somado ao fato de ser uma pesquisa amostral, foi aplicado o método *k-modes* ponderado com pesos amostrais como algoritmo de agrupamento. Essa metodologia está implementada no software R versão 4.2.0, na função `wKModes`. Uma vez que os perfis das MPE são mais homogêneos do que os observados para MEI, os resultados de sua clusterização foram obtidos com o algoritmo considerando a população de ME (5.841.103; 84%) e EPP (929.441; 16%) separadamente.

Os resultados apresentam a combinação de ambas as clusterizações. As EPP foram divididas em dois *clusters* - EPP_A1 (30%) e EPP_B1 (70%) - e as ME em 4 *clusters* - ME_A1 (19%), ME_B1 (37%), ME_C1 (18%), ME_C2 (19%). As descrições dos grupos e *clusters* estão a seguir:

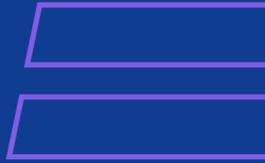
- **Grupo de EPP:** Grupo de *clusters* presentes em Empresas de Pequeno Porte;
 - Cluster A1:** Homens negros, com ensino médio completo, atuando no setor de comércio.
 - Cluster B1:** Homens brancos, com ensino superior completo, atuando no setor de serviços.
- **Grupo de ME:** Grupo de *clusters* presentes em Microempresas.
 - Cluster A1:** Mulheres brancas, com ensino superior completo, no setor de serviços.
 - Cluster B1:** Homens brancos, com ensino superior completo, e ex-carteira assinada.
 - Cluster C1:** Homens negros, com ensino médio completo, e empresa com mais de 10 anos de vida, atuando no setor de comércio.
 - Cluster C2:** Homens negros, com ensino superior completo, trabalhando em casa.

A Figura 1 abaixo apresenta as proporções das variáveis mais relevantes selecionadas pelo modelo para efetuar essa classificação usando escala de cor.

Variável/Pergunta	Subcategoria	EPP_A1	EPP_B1	ME_A1	ME_B1	ME_C1	ME_C2
Setor	Comércio	63,26	26,96	12,22	49,8	59,24	12,35
	Construção / Indústria	19,54	17,33	6,74	15,22	12,79	23,02
	Serviços	17,2	55,7	81,04	34,98	27,97	64,64
P9_Local	Em casa	6,77	17,47	23,77	9,69	9,12	53,07
	Em um estabelecimento comercial	84,99	73,25	61,29	78,54	82,83	32,91
P10_Tempo_de_vida	Entre 01 e 02 anos	10,1	12,89	16,13	11,21	10,6	22,82
	Entre 03 e 05 anos	16	13,9	21,75	15,81	14,36	38,04
	Entre 06 e 10 anos	13,5	17,08	18,69	29,96	22,21	26,22
	Mais de 10 anos	55,28	51,6	38,4	38,52	48,22	8,91
P11_Motivo_para_empresender	Conhecimento/experiência na área em que atua	37,19	58,51	57,87	19,91	61,5	65,92
P12_Principal_ocup_anterior	Empregado(a) com carteira	30,73	61,24	25,33	70,45	27,76	74,74
P19_Outra_fonte_renda	Não possui nenhuma outra fonte de renda	74,63	58,23	80,13	59,8	81,69	79,42
P20_Principal_motivo_formalizar	Possibilidade de emitir Nota Fiscal	9,89	11,51	8,69	11,65	6,57	28,88
	Ter uma empresa formal	42,85	51,9	43,8	35,38	37,32	15,06
P25_Pagamento_impostos	Pag_impostos_Não se aplica	73,9	74,45	68,61	75,16	71,01	71,78
	Pag_impostos_Sim	16,64	16,1	20,05	15,42	16,42	17,61
P27_Escolaridade	Fundamental Completo	16,13	12,26	13,6	12,88	12,09	7,57
	Medio Completo	61,75	18,36	21,82	36,42	67,91	34,66
	Superior Completo	18,02	67,11	62,73	47,12	13,64	56,08
P28_Renda_familiar_mensal	1 a 3 SM	9,59	4,89	9,53	14,23	19	10,63
	3 a 5 SM	17,04	9,91	34,85	22,56	12,91	12,75
	5 a 10 SM	37,34	20,93	20,93	31,8	21,13	49,04
	Mais de 10 SM	16,81	46,14	19,49	14,42	30,02	18,39
P30_Idade	25 a 29 anos	11,31	3,35	6,89	7,13	6,05	9,54
	30 a 39 anos	13,16	21,34	50,7	29,37	19,75	56,31
	40 a 49 anos	33,53	19,47	12,77	39,37	24,71	22,9
	50 a 64 anos	36,07	39,09	22,44	15,91	40,33	8,16
P31_Cor	Branco(a)	44,74	83,15	78,13	67,16	30,45	34,52
	Negro(a) (Preto + Pardo)	47,45	12,01	16,06	28,93	60,96	62,23
P32_Sexo	Feminino	18,73	22,86	73,83	21,01	20,7	15,83
	Masculino	81,27	77,14	26,17	78,99	79,3	84,17

Figura 1 – **Proporção das principais variáveis em relação a cada Cluster.** As cores da tabela indicam variáveis com maior (tom verde) e menor (tom vermelho) proporção.

Fonte: Elaborado pelos autores.



SEBRAE

50+50

