

Desafios da Gestão Municipal 2018

Análise por cluster

Os 100 maiores municípios brasileiros (aqui chamados de 100+) concentram 39% da população do país, 50% do PIB e 54% dos empregos formais. O estudo Desafios da Gestão Municipal (DGM), produzido pela consultoria Macroplan, especializada em cenários futuros e gestão pública, analisa o grupo dos 100+ em um conjunto de **15 indicadores** de **4 áreas de resultado** - Educação, Saúde, Segurança e Saneamento e Sustentabilidade. Esses indicadores formam o indicador sintético que varia entre 0 e 1 seguindo metodologia semelhante à do Índice de Desenvolvimento Humano (IDH) do Programa das Nações Unidas para o Desenvolvimento (PNUD). A escolha de indicadores finalísticos para compor o índice sintético permite avaliar a entrega de resultados com maior impacto na qualidade de vida e na competitividade do município.

A amostra dos 100+ não é homogênea. O grupo dos 100+ apresenta heterogeneidades internas que podem afetar a capacidade do setor público entregar resultados e a possibilidade de avanço dos indicadores finalísticos. Esta edição traz, portanto, uma análise por *cluster* que permite a comparação do município com o grupo de municípios mais similares em termos de complexidade da gestão e disponibilidade de recursos. Cada território tem suas especificidades que devem ser consideradas, mas a comparação com municípios pode se mostrar adequada para identificar desafios e potenciais avanços. O desafio é criar grupos semelhantes para comparação.

A análise por cluster, ou análise por agrupamentos, é amplamente utilizada para classificar elementos (pessoas, países, cidades) em grupos de acordo com semelhanças nas variáveis. Com isso, pretende-se constituir grupos mais parecidos em termos de desafios para a gestão pública e permitir melhor comparabilidade.

O método de clustering escolhido para este trabalho foi o K-Means¹, um dos mais consolidados e utilizados para este tipo de análise. O método é baseado na minimização de distâncias entre cada ponto e o centroide (ponto médio entre todas as observações de cada grupo) de seu cluster. A partir dos centroides iniciais, se definem os grupos com base na distância dos pontos para estes centroides. Após o agrupamento inicial, novos centroides são calculados de acordo com as médias de cada um dos pontos de seus respectivos grupos, sendo feito novo agrupamento de acordo com as distâncias para estes novos centroides. O processo se repete até que os centroides não mudem mais de uma iteração para a seguinte.

A utilização deste método permite que os grupos sejam formados com os municípios mais próximos em relação a uma série de variáveis selecionadas. No caso, as variáveis representam a caracterização do município em relação à sua (i) complexidade e (ii) disponibilidade de recursos.

¹ Yadav, Jyoti; Sharma, Monika. A Review of K-mean Algorithm. International Journal of Engineering Trends and Technology (IJETT) – Volume 4 Nº 7- Julho 2013.

Para criação dos clusters, foram utilizados os 150 maiores municípios, de acordo com a população residente estimada pelo IBGE em 2017. A amostra original, de 100 municípios, foi expandida para que os menores desses 100 não ficassem restritos a comparações com municípios maiores.

As variáveis escolhidas para clusterização e suas respectivas fontes foram as seguintes: i) variáveis que caracterizem a complexidade do municípios para a gestão: população residente estimada (IBGE, 2017), densidade populacional (População/Área, IBGE, 2017) e percentual de pessoas com 25 anos mais com ensino médio completo (Censo 2010), e ii) variáveis que mensuram a disponibilidade de recursos: renda domiciliar per capita (Censo 2010), Índice de Gini (Censo 2010), receita municipal per capita (Siconfi, 2016) e geração de receita própria (Meu Município, 2016²).

As variáveis categorizadas como Complexidade são: População, Densidade e Percentual de Adultos com Ensino Médio. Todas as outras entram na lista das variáveis de disponibilidade de Recursos. A escolha se justifica na tentativa de explicar o tamanho do desafio de cada município. Aos municípios com maior complexidade e menos disponibilidade de recursos são atribuídos maiores desafios de gestão.

O método k-means é sensível aos pontos iniciais escolhidos e essa escolha inicial depende da ordem em que os municípios estão organizados no bando de dados. Como há uma sensibilidade do algoritmo à ordem, foi criado um indicador sintético com base nas sete variáveis selecionadas para o ordenamento das linhas. Os municípios foram ordenados dos maiores índices (menos complexos e com mais recursos) para os menores.

A construção do indicador sintético segue a metodologia do IDGM³. É feita primeiramente a padronização de cada uma das variáveis com base nos seus máximos e mínimos observados. Cada variável assumiu um sinal de forma a caracterizar os municípios mais complexos aos menos complexos e com menos recursos aos com mais recursos. As variáveis com sinais invertidos são: população, densidade e Gini. Portanto, quanto menos complexo e mais recursos, maior o índice sintético do município.

A quantidade de clusters é definida pelo método “Elbow”, que leva em conta a soma dos quadrados das distâncias entre os pontos e os centroides de seus clusters. Quando a soma de quadrados passa a cair pouco com a adição de mais um cluster, é um indicativo de que a quantidade está próxima da ideal.

O indicador de qualidade do cluster – a soma dos desvios quadrados entre os grupos sobre a soma dos desvios quadrados da observação da média geral – que caracteriza o melhor ajustamento do modelo de clusterização, foi de 0,745.

Os 100 municípios foram divididos em 8 grupos. Ao organizar os municípios dessa forma, foi possível avaliar a performance de cada um comparando com um grupo de municípios similares, além de identificar um potencial de melhora de indicadores mais factível com a realidade de cada local. O resultado desta técnica de clusterização está disponível na próxima página.

O estudo completo e a análise por cluster está disponível em: www.desafiosdosmunicipios.com.br

Clusters

Cluster A

#	Município	UF
1	Maringá	PR
2	Piracicaba	SP
3	São José do Rio Preto	SP
4	Jundiá	SP
5	São José dos Campos	SP
6	Sorocaba	SP
7	Ribeirão Preto	SP
8	Londrina	PR
9	São Bernardo do Campo	SP
10	Blumenau	SC
11	Taubaté	SP
12	Joinville	SC
13	Mogi das Cruzes	SP
14	Caxias do Sul	RS
15	Palmas	TO
16	Bauru	SP
17	Campo Grande	MS
18	Praia Grande	SP
19	Juiz de Fora	MG
20	Santa Maria	RS
21	Guarujá	SP
22	Cuiabá	MT

Cluster B

#	Município	UF
1	Sumaré	SP
2	Ribeirão das Neves	MG
3	Itaquaquecetuba	SP
4	Caucaia	CE
5	Gravataí	RS
6	Cariacica	ES
7	Paulista	PE
8	Aparecida de Goiânia	GO
9	Várzea Grande	MT
10	São Gonçalo	RJ
11	Nova Iguaçu	RJ
12	Duque de Caxias	RJ
13	Belford Roxo	RJ

Cluster C

#	Município	UF
1	Limeira	SP
2	Franca	SP
3	Cascavel	PR
4	Uberlândia	MG
5	Uberaba	MG
6	Ponta Grossa	PR
7	Suzano	SP
8	Betim	MG
9	Petrópolis	RJ
10	Contagem	MG
11	Guarulhos	SP
12	São José dos Pinhais	PR
13	Governador Valadares	MG
14	São Vicente	SP
15	Anápolis	GO
16	Serra	ES
17	Pelotas	RS
18	Canoas	RS
19	Camaçari	BA
20	Porto Velho	RO

Cluster D

#	Município	UF
1	Mauá	SP
2	Taboão da Serra	SP
3	Diadema	SP
4	Osasco	SP
5	Carapicuíba	SP
6	Olinda	PE
7	São João de Meriti	RJ

Cluster E

#	Município	UF
1	Campinas	SP
2	Santos	SP
3	Curitiba	PR
4	Vitória	ES
5	Florianópolis	SC
6	Belo Horizonte	MG
7	Santo André	SP
8	Niterói	RJ
9	Porto Alegre	RS

Cluster F

#	Município	UF
1	Montes Claros	MG
2	Campina Grande	PB
3	Boa Vista	RR
4	Petrolina	PE
5	Feira de Santana	BA
6	Vitória da Conquista	BA
7	Teresina	PI
8	Campos dos Goytacazes	RJ
9	Mossoró	RN
10	Rio Branco	AC
11	Caruaru	PE
12	Santarém	PA
13	Jaboatão dos Guararapes	PE
14	Macapá	AP
15	Ananindeua	PA

Cluster G

#	Município	UF
1	Goiânia	GO
2	Vila Velha	ES
3	Fortaleza	CE
4	João Pessoa	PB
5	Salvador	BA
6	Recife	PE
7	Aracaju	SE
8	Natal	RN
9	Manaus	AM
10	São Luís	MA
11	Maceió	AL
12	Belém	PA

Cluster H

#	Município	UF
1	São Paulo	SP
2	Rio de Janeiro	RJ

Para mais informações sobre a biblioteca na linguagem R que foi utilizada, acesse:

<https://cran.r-project.org/web/packages/ClusterR/ClusterR.pdf>.