

Case Plussoft Data Analyst

Daniel Volponi





CENÁRIO

Uma grande multinacional varejista do ramo de supermercados deseja ingressar no mercado brasileiro e o Data Analyst está responsável pelo estudo demográfico que servirá de base para a estratégia de entrada dessa empresa no Brasil.

OBJETIVOS

Realizar as seguintes tarefas a partir das informações demográficas dos municípios brasileiros:

- Classificar os municípios brasileiros com base nas informações disponíveis e caracterizá-los em grupos;
- Elaborar uma forma de classificar um novo município entre os grupos já criados na etapa anterior;
- Indicar quais grupos de municípios deveriam ser a porta de entrada para a empresa no país e porque.





FERRAMENTAS UTILIZADAS



Todos os notebooks e análises realizadas podem ser vistas no [repositório no Github](#).

BASES UTILIZADAS

1. **Base de Informações Demográficas dos Municípios Brasileiros:** base disponibilizada pela Plusoft com informações demográficas do IBGE de todos os municípios do Brasil.
2. Para compor a análise foi utilizado a **Base do Cadastro Nacional da Pessoa Jurídica:** dados da Receita Federal, com informações do registro de CNPJ por município.

PREMISSAS

- Para fins de estudo foram utilizadas somente linhas com pelo menos 100.000 habitantes, totalizando 223 municípios.

Variáveis Utilizadas

1. **Densidade demográfica, 2000:** distribuição espacial dos habitantes por quilômetro quadrado e os principais sistemas de transporte.
2. **Renda per Capita:** A renda per capita é obtida mediante a divisão da Renda Nacional (isto é, Produto Nacional Bruto menos os gastos de depreciação do capital e os impostos diretos) pelo número de habitantes do país.
3. **Intensidade da Pobreza:** percentual das pessoas consideradas como indigentes, ou seja, aqueles cuja renda familiar per capita é inferior ao valor necessário para atender tão-somente às necessidades básicas de alimentação.
4. **Índice de Desenvolvimento Humano Municipal:** uma medida resumida do progresso a longo prazo em três dimensões básicas do desenvolvimento humano: renda, educação e saúde.
5. **Variação População 1991 a 2000:** variação da população entre 1991 e 2000.
6. **Percentual População Urbana:** População Urbana sobre a população total.
7. **População por Supermercado:** quantidade total de CNPJs do CNAE 4711301 e 471102, referente aos supermercados disponibilizados pela Receita Federal do Brasil, divididos pela população do município.

Clusterização com K-means

O que é K-Means?

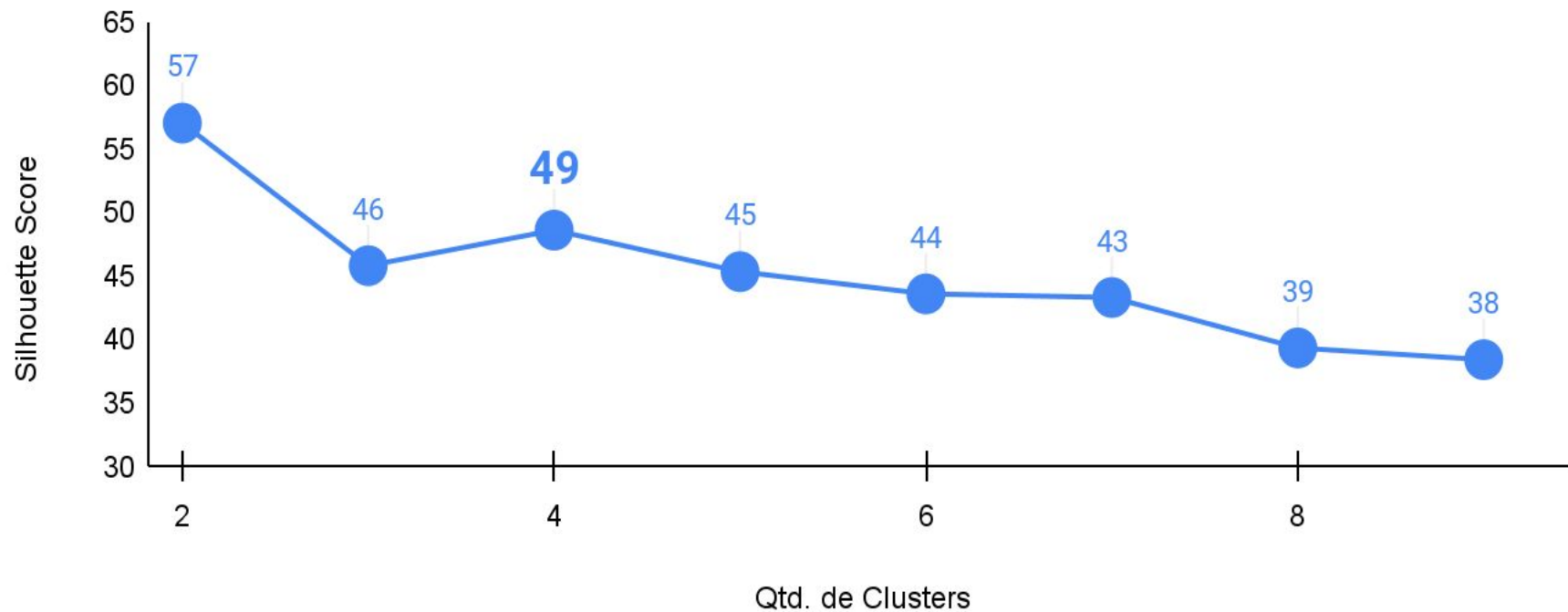
K-Means é um algoritmo de clusterização (ou agrupamento) disponível na biblioteca Scikit-Learn. É um algoritmo de aprendizado não supervisionado (ou seja, que não precisa de inputs de confirmação externos) que avalia e clusteriza os dados de acordo com suas características. [\[Fonte\]](#)

Coeficiente de Silhouette

O Coeficiente Silhouette é calculado usando a distância média intra-cluster (a) e a distância média mais próxima do cluster (b) para cada amostra.



Definindo a Quantidade Ideal de Clusters



Clusters Encontrados



Populoso

Cidades: 73
Densidade Demográfica: 3.330
Intensidade da Pobreza: 46
IDH: 79
Variação População: 26%
População Urbana: 98%
População/Supermercado: 502

Exemplos:
Rio de Janeiro (RJ)
São Paulo (SP).



Metropolitano

Cidades: 51
Densidade Demográfica: 621
Intensidade da Pobreza: 44
IDH: 78
Variação População: 34%
População Urbana: 94%
População/Supermercado: 576

Exemplos:
São José dos Campos (SP)
Joinville (SC)



Competitivo

Cidades: 70
Densidade Demográfica: 131
Intensidade da Pobreza: 40
IDH: 80
Variação População: 25%
População Urbana: 92%
População/Supermercado: 321

Exemplos:
Brasília (DF)
Manaus (AM)

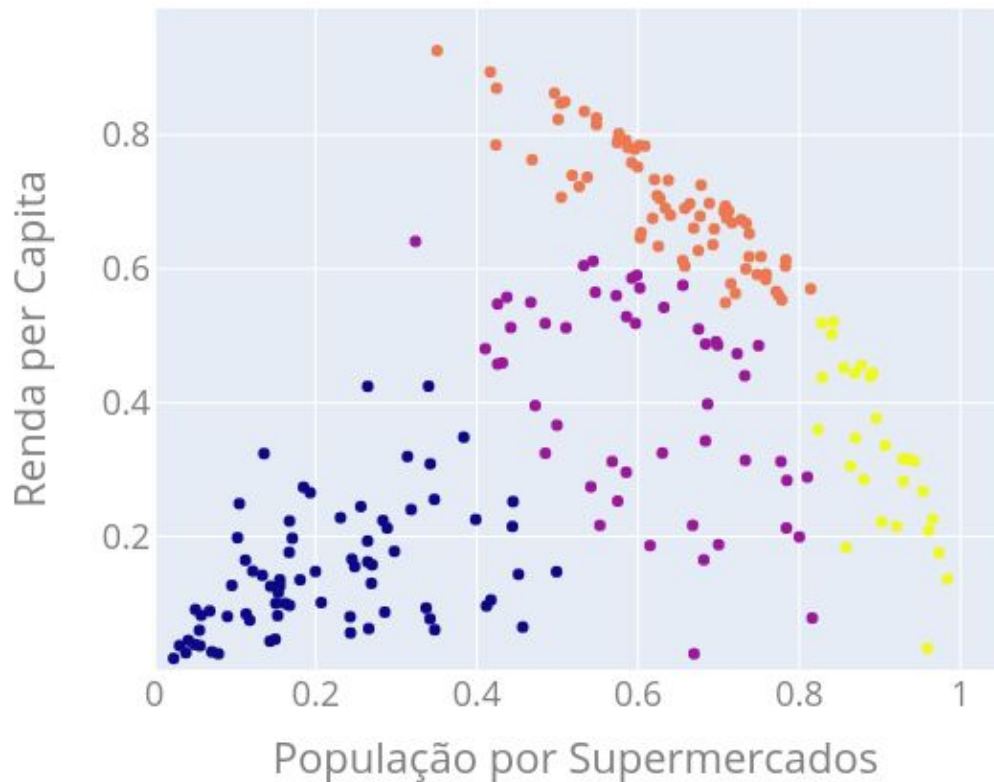


Oportuno

Cidades: 29
Densidade Demográfica: 179
Intensidade da Pobreza: 46
IDH: 72
Variação População: 22%
População Urbana: 85%
População/Supermercado: 690

Exemplos:
Santarém (PA)
Caruaru (PE).

Renda Per Capita x População por Supermercados



Populoso



Metropolitano

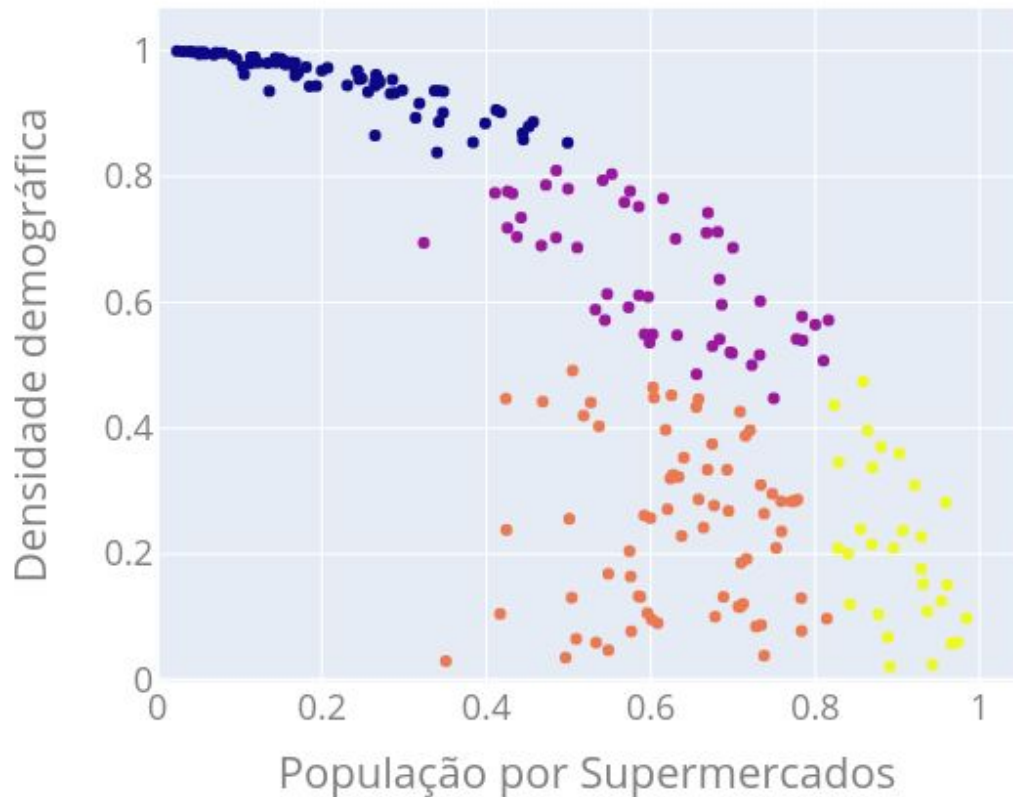


Competitivo



Oportuno

Densidade Demográfica x População por Supermercados



Populoso



Metropolitano



Competitivo



Oportuno

Sugestão de Investimento: Cluster Oportuno

Abertura 10 Maiores Cidades



Município	População	Renda per Capita	IDH	População por Supermercados
Santarém (PA)	262,538	139.9	0.746	421
Caruaru (PE)	253,634	209.76	0.713	593
Caucaia (CE)	250,479	129.23	0.721	525
Ilhéus (BA)	222,127	170.22	0.703	459
Petrolina (PE)	218,538	201.23	0.747	387
Mossoró (RN)	213,841	179.59	0.735	528
Marília (SP)	197,342	421.22	0.821	705
Sete Lagoas (MG)	184,871	281.65	0.791	644
Divinópolis (MG)	183,962	327.64	0.831	619
Juazeiro (BA)	174,567	175.15	0.683	354

Oportuno

Cidades: 29

Densidade Demográfica: 179

Intensidade da Pobreza: 46

IDH: 72

Variação População: 22%

População Urbana: 85%

População/Supermercado: 690

Exemplos:

Santarém (PA)

Caruaru (PE).



CONCLUSÕES

- Considerando as cidades com pelo menos 100 mil habitantes, as cidades do cluster “Oportuno” são as melhores alternativas para a empresa ingressar no Brasil, uma vez que apresentam menor índice de competitividade (População por Supermercado).
- Não é indicado o investimento nas cidades do cluster “Competitivo”. Estas já possuem alto volume de supermercados por habitantes.
- O cluster “Populoso” pode ser considerado uma alternativa conservadora de investimento, visto que possui as maiores cidades do país.