



Entrada de multinacional no mercado brasileiro

Cassiano Gross Schuler

IBOPE
inteligência

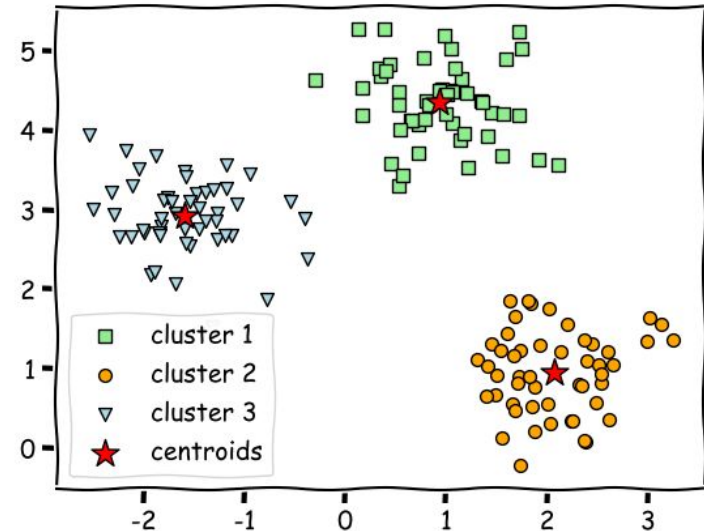
Agrupando Municípios

O agrupamento neste projeto é realizado a partir de um algoritmo de aprendizagem de máquina chamado K-Means. O projeto foi realizado em linguagem Python um Jupyter Notebook postado na plataforma Github. O link encontra-se no último slide, Referências.

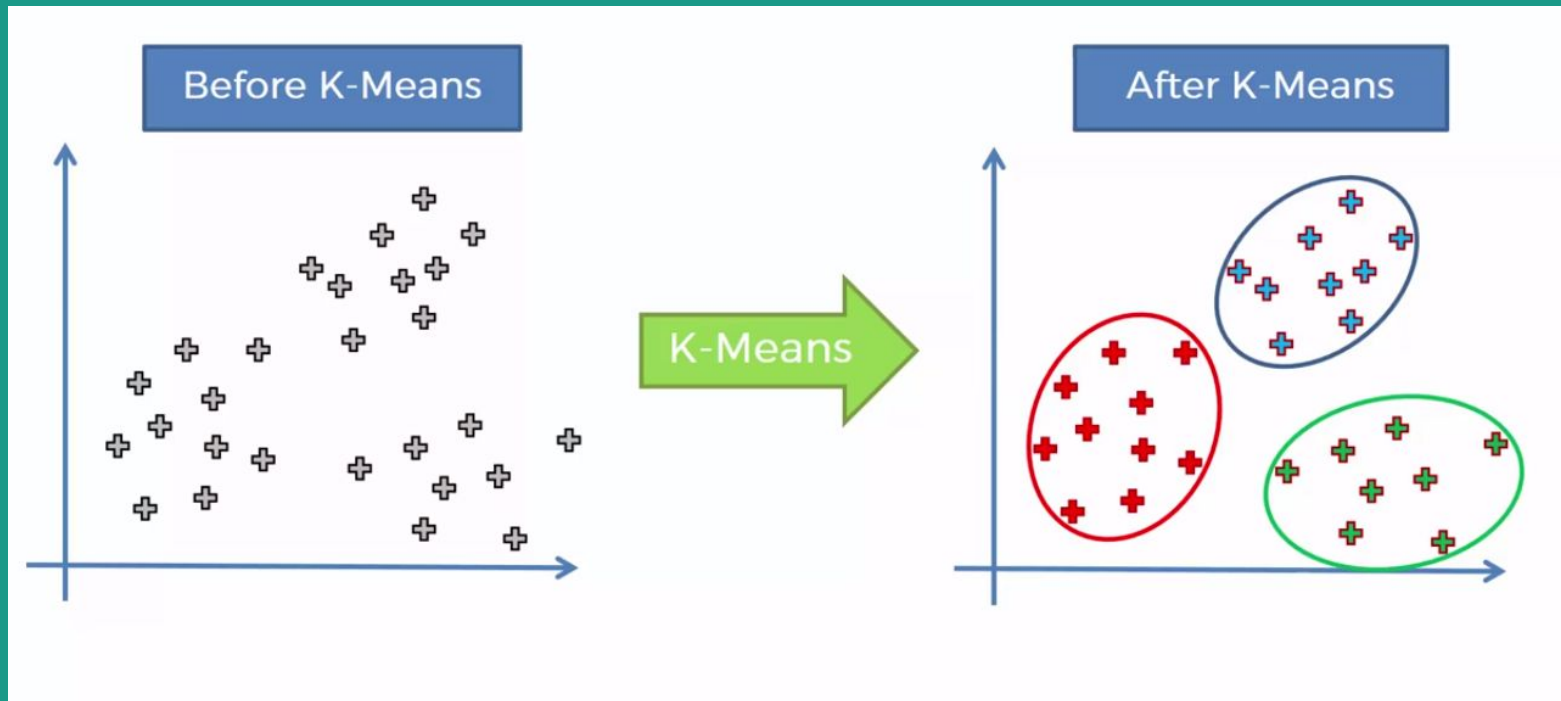


Algoritmo: K-Means

Este algoritmo faz uma busca por proximidade nos dados. Pontos centrais para cada grupo de dados são ajustados e atualizados algumas vezes até encontrarem uma melhor posição entre esses dados. Esses pontos são chamados centróides e agregam os dados em seu grupo por proximidade em um cluster.



Algoritmo: K-Means

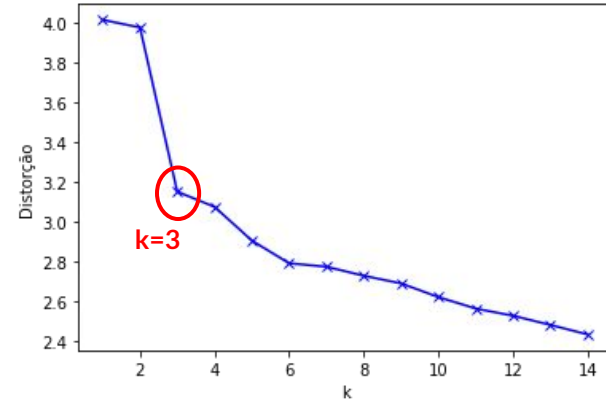


Número de grupos

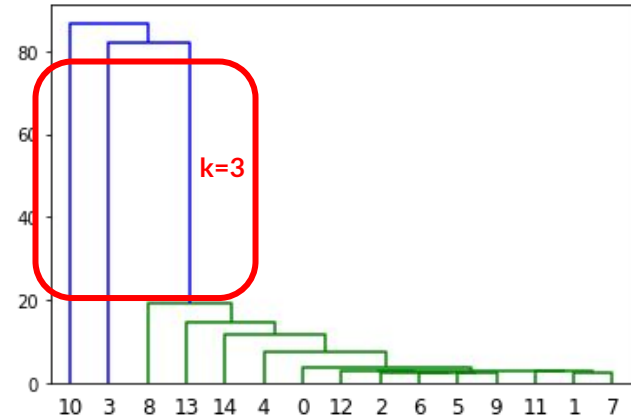
Apesar de todos os indicadores sugerirem preferencialmente 3 grupos (k), essa utilização gerou um grupo igual ao grupo A, que veremos a seguir e mais dois grupos grandes.

Para uma melhor classificação neste projeto foram utilizados 7 grupos, o que ainda é algo consistente com os indicadores.

Método do cotovelo



Dendrograma





Grupo A

Cidades: 2

Rio de Janeiro (RJ)
São Paulo (SP)

Alta

População, crescimento populacional, renda, anos de estudo.

Média

Área.

Baixa

População rural, distância à capital, analfabetismo, taxa de fecundidade.



Grupo B

Cidades: 11

Recife (PE)
Goiânia (GO)
Salvador (BA)
Fortaleza (CE)
Brasília (DF)

...

Alta

Crescimento populacional, população urbana, população, renda.

Média

Área.

Baixa

População rural, distância à capital, analfabetismo, taxa de fecundidade.



Grupo C

Cidades: 102

Camaragibe (PE)

Pelotas (RS)

São José dos Campos (SP)

Contagem (MG)

Nova Iguaçu (RJ)

...

Alta

Densidade demográfica, crescimento populacional, média de anos de estudo, renda.

Média

População rural, área, intensidade de pobreza.

Baixa

População rural, analfabetismo, distância a capital, taxa de fecundidade, mortalidade infantil.



Grupo D

Cidades: 1487

Bom Sucesso (MG)
Espírito Santo do Pinhal (SP)
Lages (SC)
Três Fronteiras (SP)
Anchieta (ES)

...

Alta

Proporção urbana, anos de estudo, renda, IDH, alfabetização, esperança de vida.

Média

População.

Baixa

Proporção rural, analfabetismo, pobreza, mortalidade infantil, taxa de fecundidade.



Grupo E

Cidades: 1421

Riachuelo (RN)

Umburatiba (MG)

Campo Redondo (RN)

Rafael Godeiro (RN)

Chapada do Norte (MG)

...

Alta

Pobreza, analfabetismo, mortalidade infantil.

Média

Taxa de crescimento, população rural, distância à capital.

Baixa

Renda per Capita, IDH, anos de estudo e esperança de vida ao nascer.



Grupo F

Cidades: 1205

Malhada de Pedras (BA)
Jussara (BA)
Aurora do Pará (PA)
Belém do Brejo do Cruz (PB)
Ruy Barbosa (RN)

...

Alta

Analfabetismo, mortalidade infantil, pobreza, taxa de fecundidade.

Média

Taxa de crescimento, população rural, distância à capital.

Baixa

IDH, alfabetização, anos de estudo, renda per Capita, e esperança de vida ao nascer.



Grupo G

Cidades: 1279

Arandu (SP)

Congonhinhas (PR)

Vargem Bonita (SC)

Serra da Saudade (MG)

Marapoama (SP)

...

Alta

Esperança de vida ao nascer, alfabetização, IDH, proporção rural da população.

Média

Distância à capital, população em geral.

Baixa

Indigência, pobreza, analfabetismo, mortalidade infantil .

São Paulo



Portas de entrada no Brasil



Rio de Janeiro



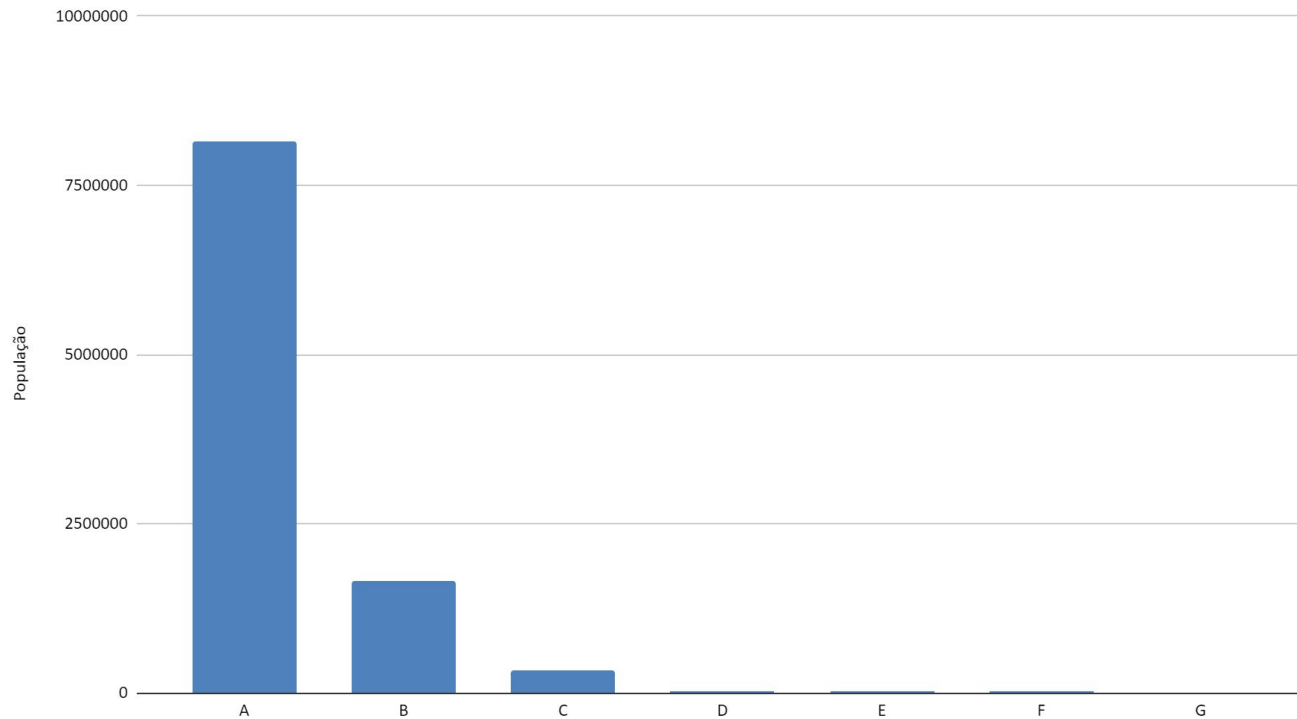
Fatores importantes a se considerar

- Maior concentração populacional gera mais visibilidade, mais mercado.
- Maior renda per capita gera mais compradores em potencial.
- Menor distância até a capital gera um menor custo com logística e distribuição.
- Taxa de aumento populacional é um investimento futuro, estará em um local com uma maior população daqui a alguns anos.
- Uma maior média de anos de estudo serve para contratação de melhores profissionais locais.

Dados consideráveis: Médias por cada grupo

GRUPOS	Distância à capital (km)	Renda per Capita, 2000	Média de anos de estudo das pessoas de 25 anos ou mais, 2000	População total, 2000	Taxa de crescimento populacional em 10 anos	Quantidade de cidades no grupo
A	0.00	603.34	8.12	8146078.00	580934.50	2
B	1.28	451.05	7.77	1645196.73	268160.36	11
C	68.19	348.58	6.72	343293.29	65491.19	102
D	256.63	265.53	5.34	32184.36	5256.31	1487
E	255.54	115.11	3.46	17995.54	1827.17	1421
F	271.69	69.70	2.44	14295.31	1147.44	1205
G	246.56	200.59	4.42	7613.98	302.13	1279

População média total, 2000



População total média municipal de acordo com cada grupo

Porta de entrada: Grupo A

As cidades de São Paulo e Rio de Janeiro são as melhores opções para um primeiro contato com o país, pois são capitais com grande população e renda, sem contar que ainda há uma futura previsão de crescimento populacional.





Próxima etapa de expansão: Grupo B

Este grupo é composto majoritariamente por capitais, 11 cidades que seguem a mesma linha do Grupo A, porém não são cidades tão populosas e ricas.

572	
584	Belém (PA)
742	Belo Horizonte (MG)
1487	Brasília (DF)
1807	Curitiba (PR)
1908	Fortaleza (CE)
2010	Goiânia (GO)
2846	Guarulhos (SP)
3896	Manaus (AM)
4059	Porto Alegre (RS)
4279	Recife (PE)
	Salvador (BA)



Etapas de expansão seguintes:

- A ordem alfabética dos grupos não foi dada por acaso, esta é a ordem de expansão que a empresa deve seguir.
- Cada grupo, em sequência possui um número de cidades que vai aumentando conforme a empresa vai tendo condições de se expandir mais e mais.

GRUPOS	Quantidade de cidades no grupo
A	2
B	11
C	102
D	1487
E	1421
F	1205
G	1279

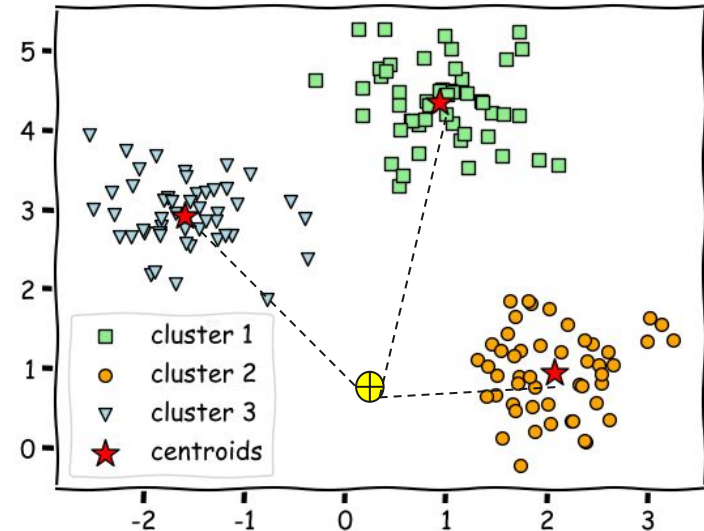
Modelo de classificação para a probabilidade


O este modelo está criado no notebook, mas mostrarei um pouco dos passos realizados

Ponto Amarelo

Baseado na mesma imagem mostrada anteriormente temos uma distância de um ponto de interesse até os centróides dos clusters. Quanto mais próximo do centróide maior a chance de pertencimento do ponto a esse grupo.

Então essas probabilidades estão relacionadas com as distâncias: mais próximo é mais provável.





```
[520] # Previsão pelo modelo
      amostra = dados.sample()
      amostra_drop = func_prepara_para_padronizar(amostra)
      amostra_padronizada = scaler.transform(amostra_drop); amostra_padronizada
```

Neste exemplo estou gerando o dado a partir de uma amostra aleatória, porém eu poderia estar entrando com um dado novo.

Nesta etapa a amostra está sendo tratada para ficar padronizada conforme os dados que foram tratados durante o programa.

‘amostra_padronizada’ vai gerar ter todos os dados de ‘amostra’ só que padronizados pela função `scaler.transform()`.

```
def probabilidade_de_proximidade(amostra_padronizada):
    distancias = modelo.transform(amostra_padronizada)
    distancias2 = 1/distancias
    soma = distancias2.sum()
    proporcoes = distancias2 / soma

    grupo = 0
    for item in proporcoes[0]:
        print(f'grupo {dicionario_nivel_cidade[grupo]} ou {grupo}: {round(item*100,2)}%')
        grupo += 1
```

```
x = probabilidade_de_proximidade(amostra_padronizada); x
```

```
grupo D ou 0: 37.03%
grupo E ou 1: 18.46%
grupo B ou 2: 2.4%
grupo A ou 3: 0.51%
grupo G ou 4: 21.95%
grupo C ou 5: 8.52%
grupo F ou 6: 11.13%
```

Então é criada uma função que ao receber a amostra padronizada devolve as probabilidades da amostra de estar em cada um dos grupos.

A variável distâncias recebe as posições do ponto 'amarelo' em relação a cada um dos grupos. Estas distâncias então são tratadas para que se tornem as probabilidades de pertencimento a determinado grupo.

Referência e Projeto:



https://github.com/Zorug/Projetos_Futuros/blob/master/Ibope_Inteligencia/Principal_ibope_project_more_clusters.ipynb

Contato:



<https://www.linkedin.com/in/cassiano-gross/>



cassiano_gross@hotmail.com