

# Cluster de saúde - versão 8 de abril

## Análise de Estratégias de Agrupamento - 08/04

### Tentativas Realizadas

#### 1. Uso de RobustScaler

- Resultado: **Não funcionou**

#### 2. Cálculo de peso com base no tamanho inverso dos clusters

- Resultado: **Não funcionou**

#### 3. Uso do algoritmo DBSCAN

- Resultado: **Forneceu uma visão mais especial para este caso específico**

### Observações sobre o DBSCAN

- **Tratamento de outliers:**
  - DBSCAN classifica pontos como outliers (cluster -1).
  - Esses pontos normalmente são **excluídos do Silhouette Score**, pois não pertencem a nenhum cluster propriamente dito.
- **Clusters de formato arbitrário:**
  - DBSCAN é ótimo para **formas não-esféricas**.
  - Porém, o Silhouette Score tende a favorecer **clusters compactos e bem separados**, o que pode **subestimar a qualidade** dos agrupamentos com formas complexas.
- **Densidade variável:**
  - DBSCAN se baseia em densidade.
  - O Silhouette Score pode **não capturar bem a qualidade** de agrupamentos com densidades diferentes.

### Métricas Alternativas para Avaliação

Além do Silhouette Score, outras métricas úteis para o DBSCAN:

- **Índice de Davies-Bouldin:** Avalia a separação entre clusters.
- **Validação baseada em densidade (DBCv):** Density-Based Clustering Validation.
- **Avaliação visual:** Usando técnicas de redução de dimensionalidade como PCA ou t-SNE.

### Considerações sobre os Parâmetros do DBSCAN

- A escolha dos parâmetros `eps` e `min_samples` tem **grande impacto** na qualidade do agrupamento.
  - Recomenda-se realizar uma **busca de parâmetros** e avaliar os resultados com diferentes métricas e avaliação visual.
- 

## Conclusão da Análise por Tentativa e Erro

- Para esses dados com **proximidade muito alta**, o **melhor agrupamento até o momento** é o atual, que:
  - **Generaliza a maioria das cidades**, principalmente devido ao grande número de zeros (mesmo com taxas transformadas, como vacinação).
  - Utiliza uma **estratégia de pré-processamento** com:
    - `IQR`, `standard_score()`, `z_score`
    - PCA antes do agrupamento
    - Remoção de outliers
    - Implementação artificial de distribuição

## Observação sobre Mortalidade Materna

- A variável de **mortalidade materna**, por possuir **poucos dados**, pode estar influenciando negativamente nos clusters.
- A estratégia será **testar manter essa variável fora da análise**.