



# Atividade 05

Agrupamento em grafos

Abril, 2025



# Equipe



**Enzo Borges Rocha Lima**

Membro



**Luis Felipe do Nascimento Moura**

Membro



# Objetivo



- 
- Aplicar os algoritmos para agrupar dados :
    - K-Means
    - DBSCAN
  - Estimar os parâmetros mais adequados
  - Comparar e discutir o melhor agrupamento.
- 



# Algoritmos

## K-Means

- Algoritmo de clusterização que divide os dados em k grupos.
- Escolhe k centróides (pontos centrais), atribuindo cada ponto ao centróide mais próximo e, depois, recalculando os centróides com base nos pontos atribuídos.
- Repete o processo até estabilizar ou alcançar o limiar estabelecido.

## DBSCAN

- Algoritmo de agrupamento baseado em densidade.
- Define um raio (eps) e um número mínimo de pontos (min\_samples).
- Um ponto é considerado parte de um cluster se houver pontos suficientes dentro do raio.
- Pontos em regiões com baixa densidade são considerados ruído (outliers).

# Implementação

---

- Foram utilizados diferentes conjuntos de parâmetros(colunas), são eles:
  - **Combinações básicas**
    - ['fixed acidity', 'volatile acidity', 'alcohol'],
    - ['fixed acidity', 'pH', 'alcohol'],
    - ['citric acid', 'residual sugar', 'alcohol'],
    - ['density', 'alcohol', 'chlorides'],
  - **Combinações relacionadas à acidez**
    - ['fixed acidity', 'volatile acidity', 'citric acid', 'pH', 'alcohol'],
    - ['fixed acidity', 'citric acid', 'pH'],
    - ['volatile acidity', 'citric acid', 'sulphates'],
  - **Combinações relacionadas à doçura e corpo**
    - ['residual sugar', 'alcohol', 'density'],
    - ['residual sugar', 'alcohol', 'chlorides'],
    - ['residual sugar', 'free sulfur dioxide', 'total sulfur dioxide'],

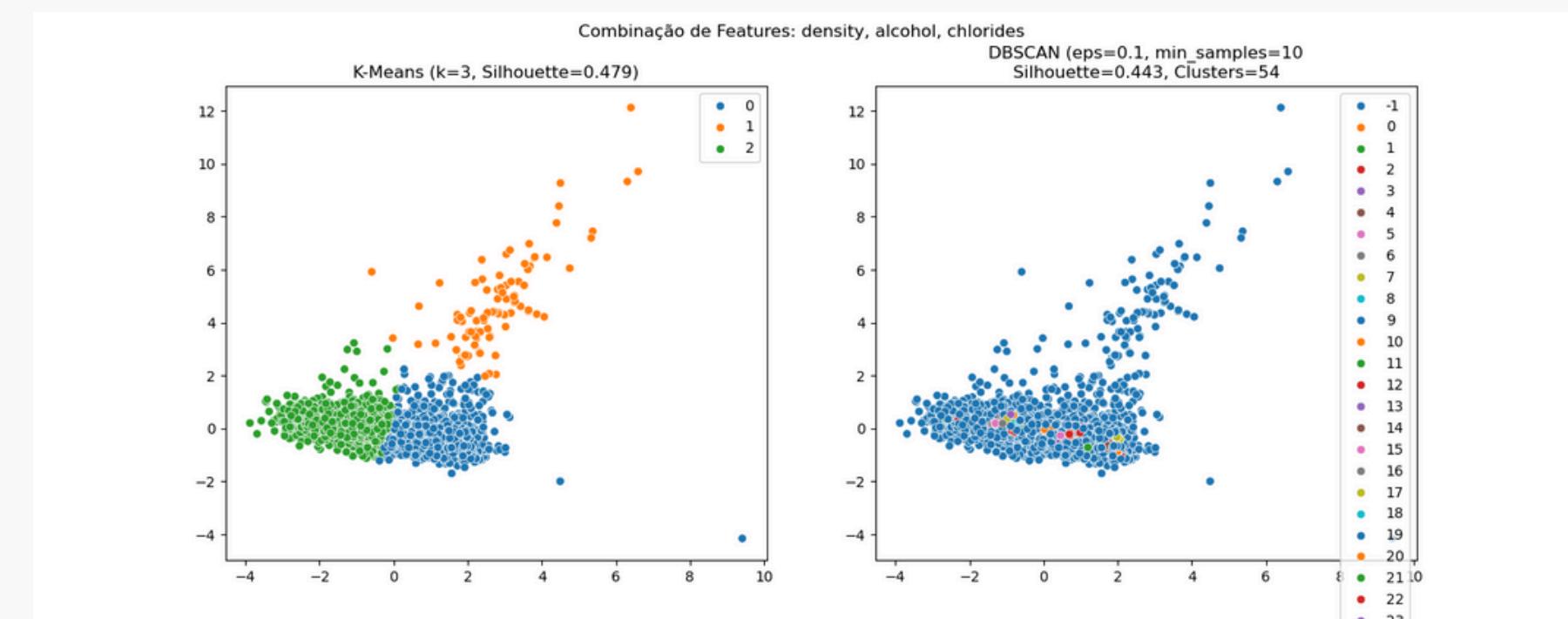
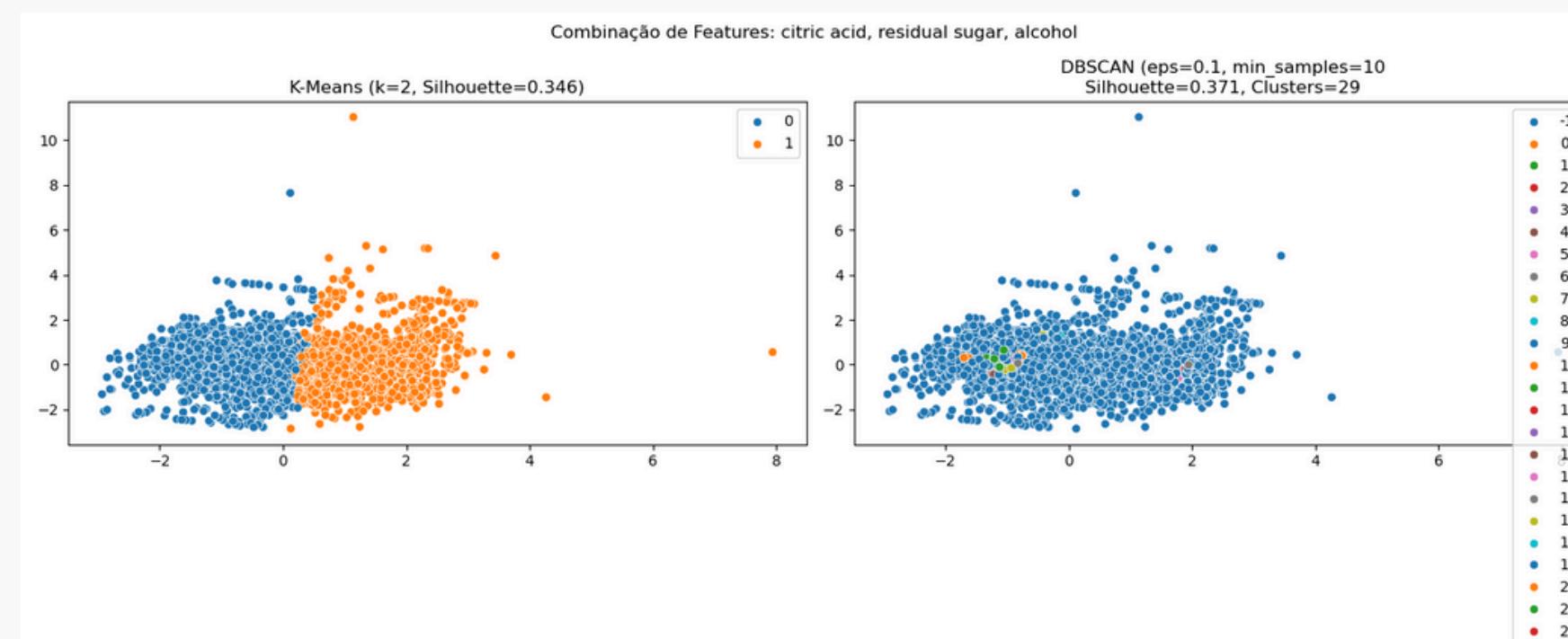
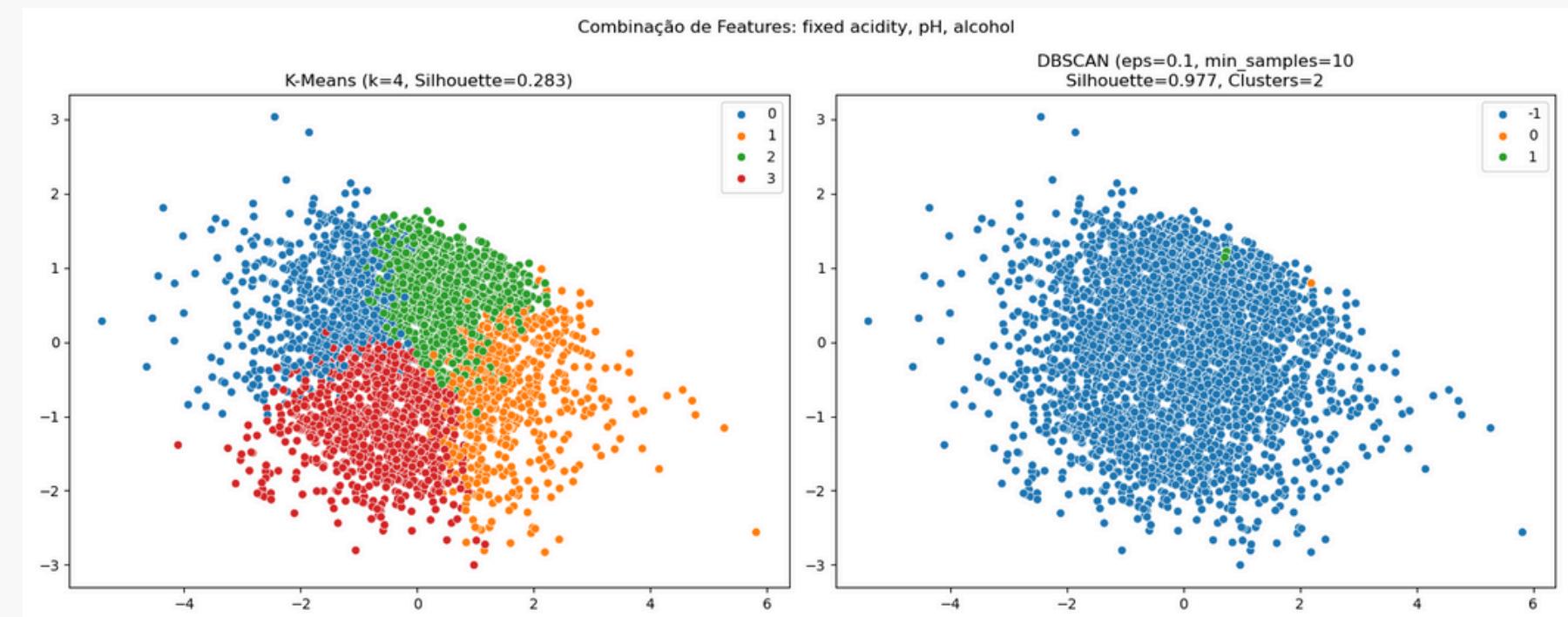
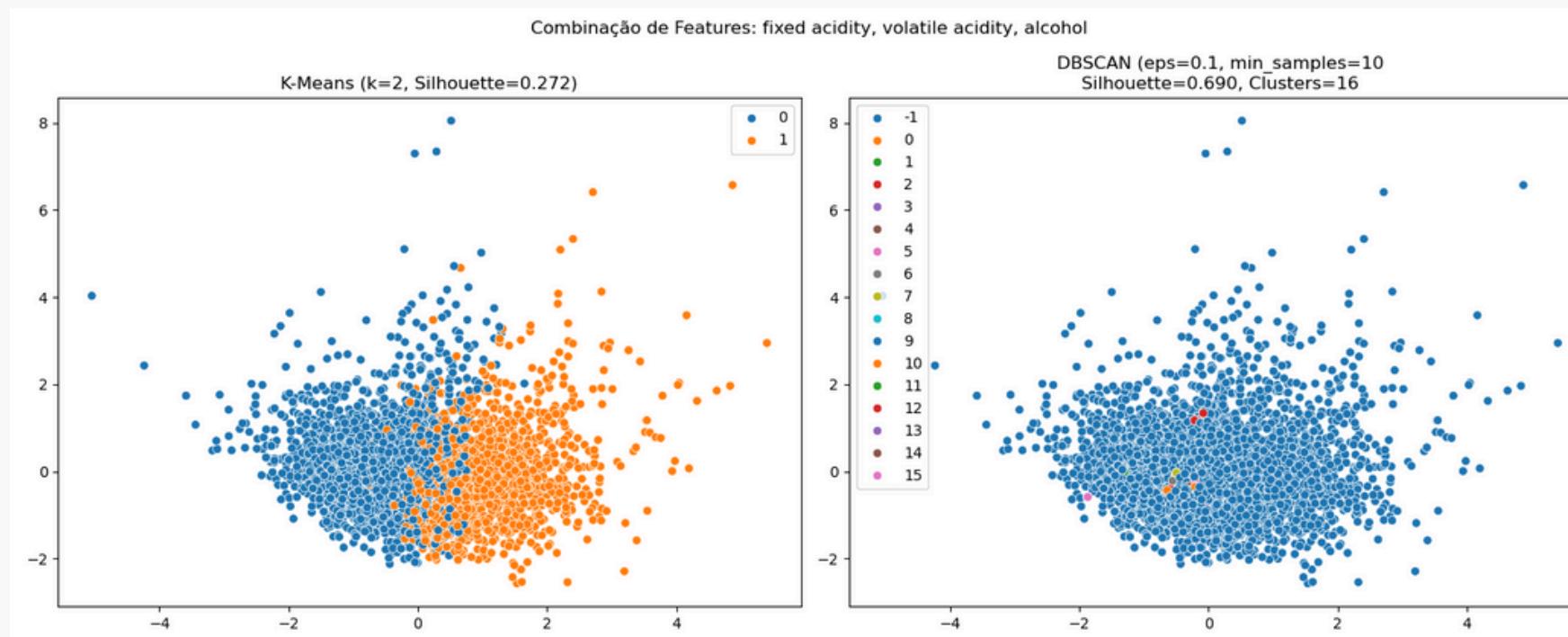
# Implementação

---

- **Combinações relacionadas a compostos sulfurados**
  - ['free sulfur dioxide', 'total sulfur dioxide', 'sulphates'],
  - ['chlorides', 'sulphates', 'alcohol'],
  - ['free sulfur dioxide', 'total sulfur dioxide', 'pH'],
- **Combinações mais complexas**
  - ['fixed acidity', 'volatile acidity', 'citric acid', 'residual sugar', 'chlorides', 'alcohol', 'sulphates'],
  - ['fixed acidity', 'pH', 'alcohol', 'density', 'sulphates'],
  - ['volatile acidity', 'citric acid', 'residual sugar', 'chlorides', 'total sulfur dioxide'],
- **Combinações específicas para características distintas**
  - ['alcohol', 'density', 'pH', 'sulphates'], # Corpo e estrutura
  - ['fixed acidity', 'volatile acidity', 'alcohol', 'sulphates'], # Acidez e preservação
  - ['residual sugar', 'alcohol', 'chlorides', 'total sulfur dioxide'], # Doçura e conservantes

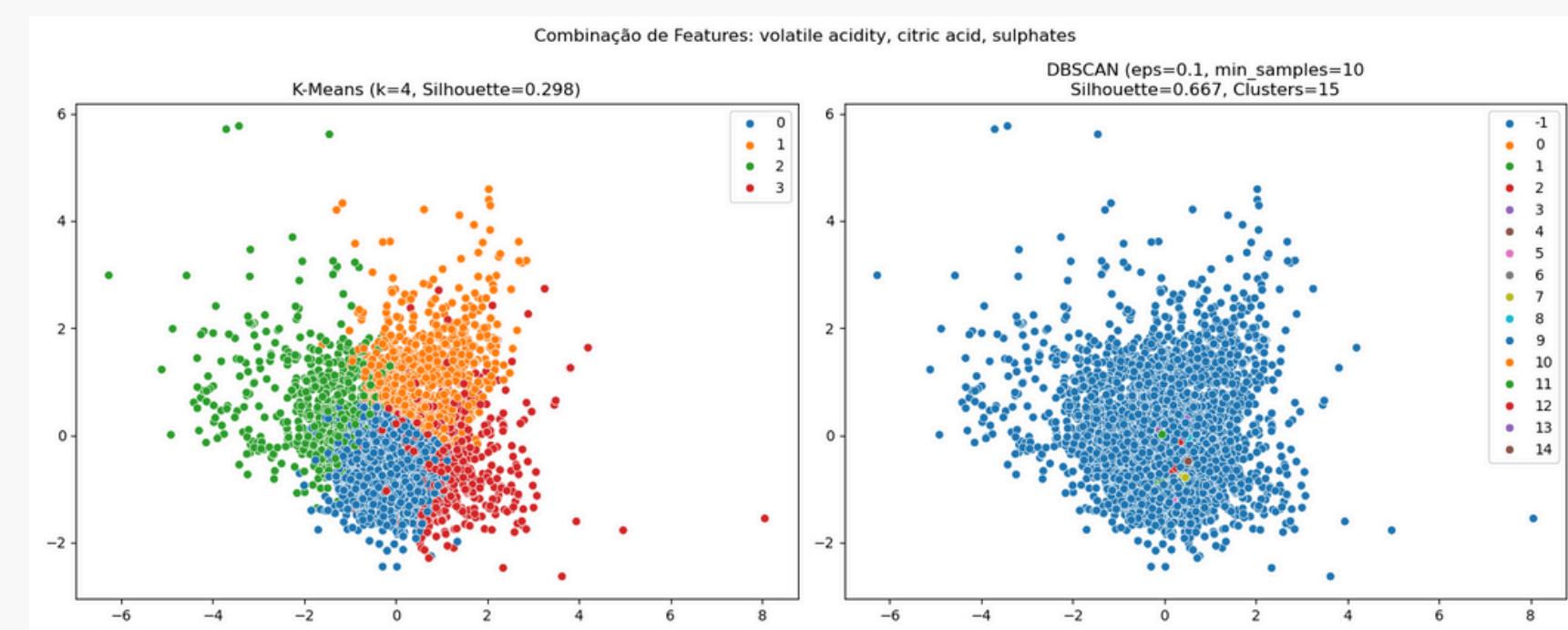
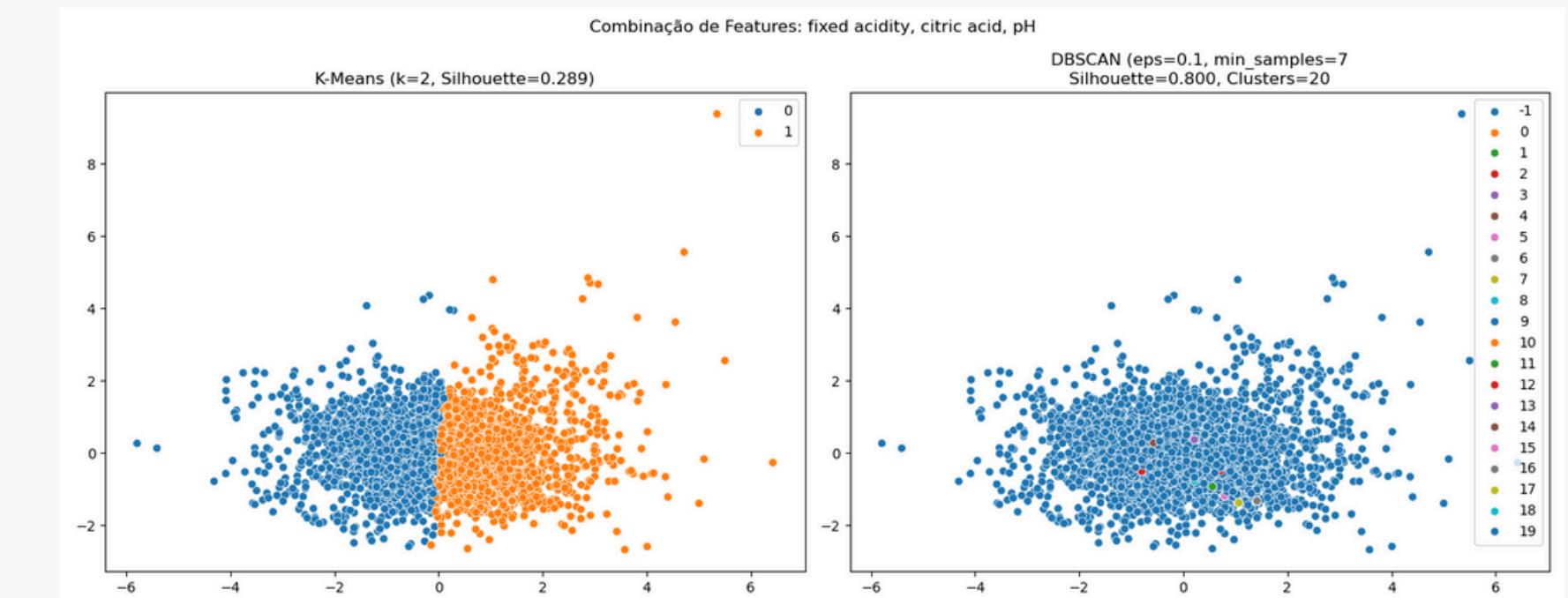
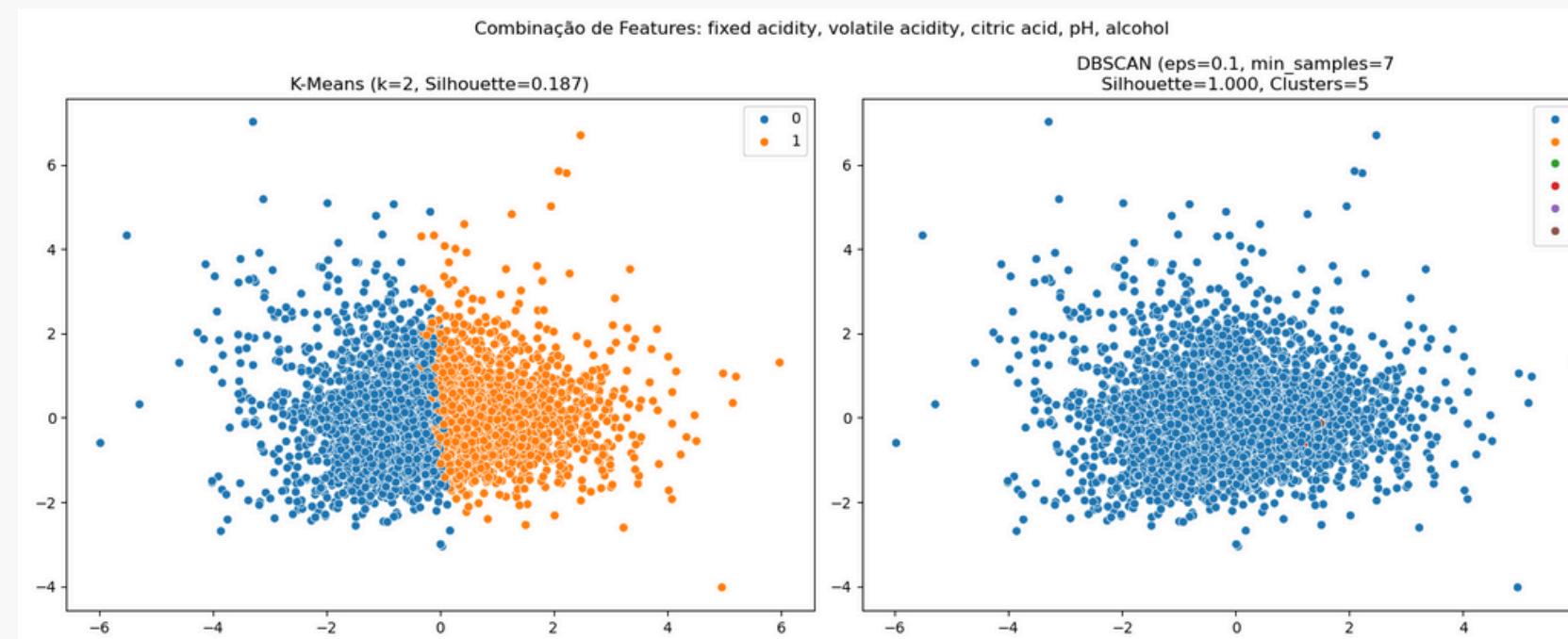
# Resultados

## Combinações básicas



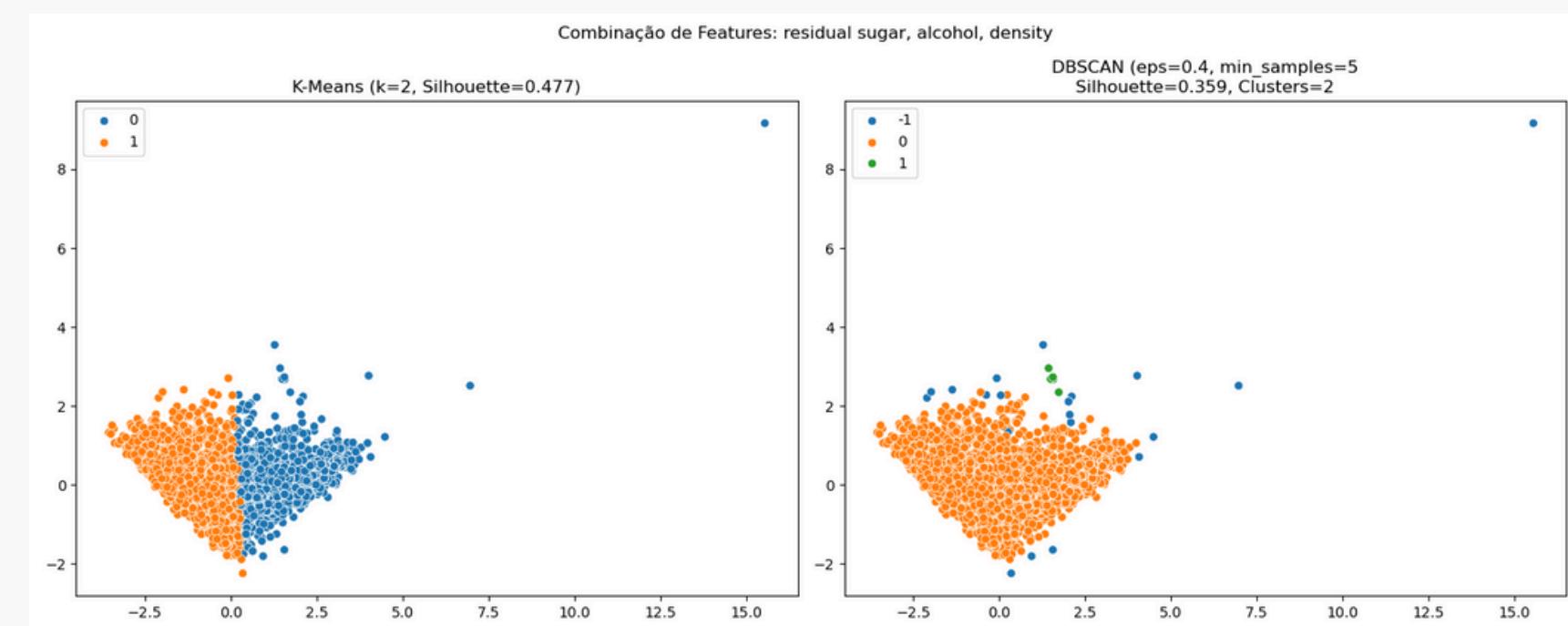
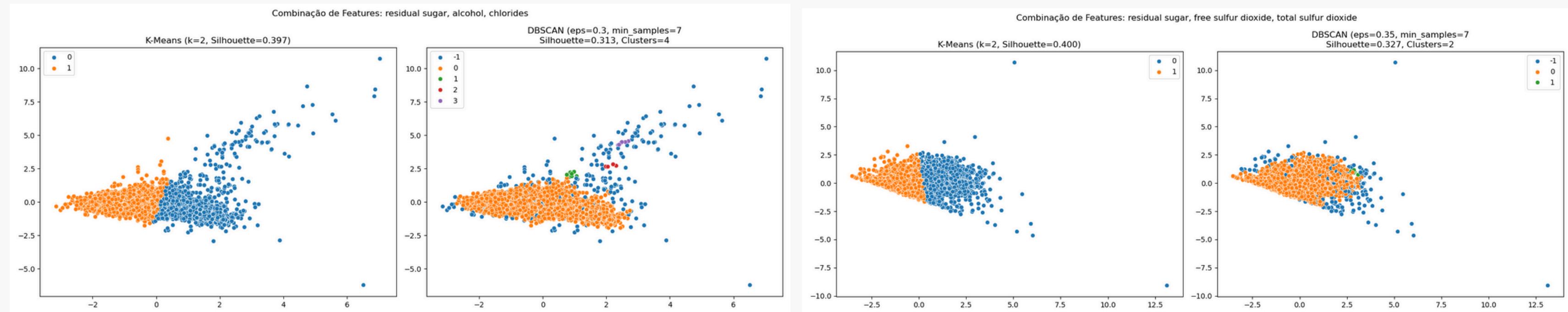
# Resultados

## Combinações relacionadas à acidez



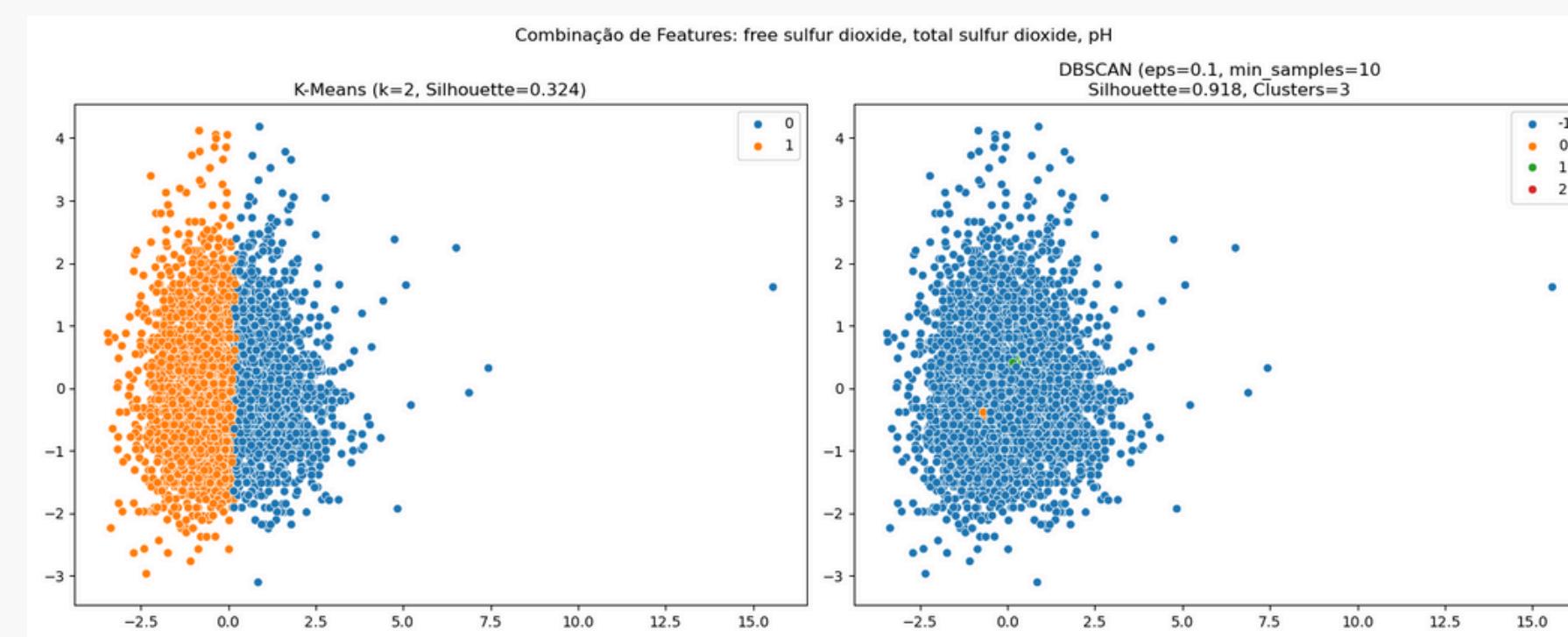
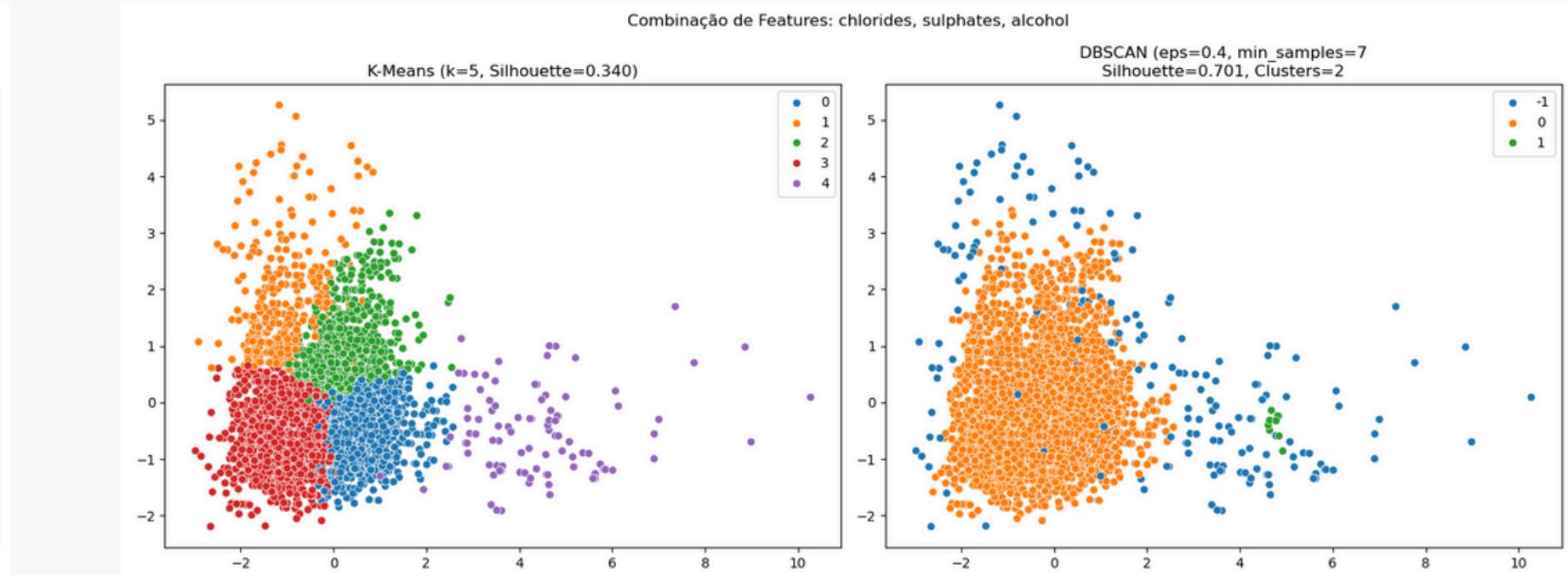
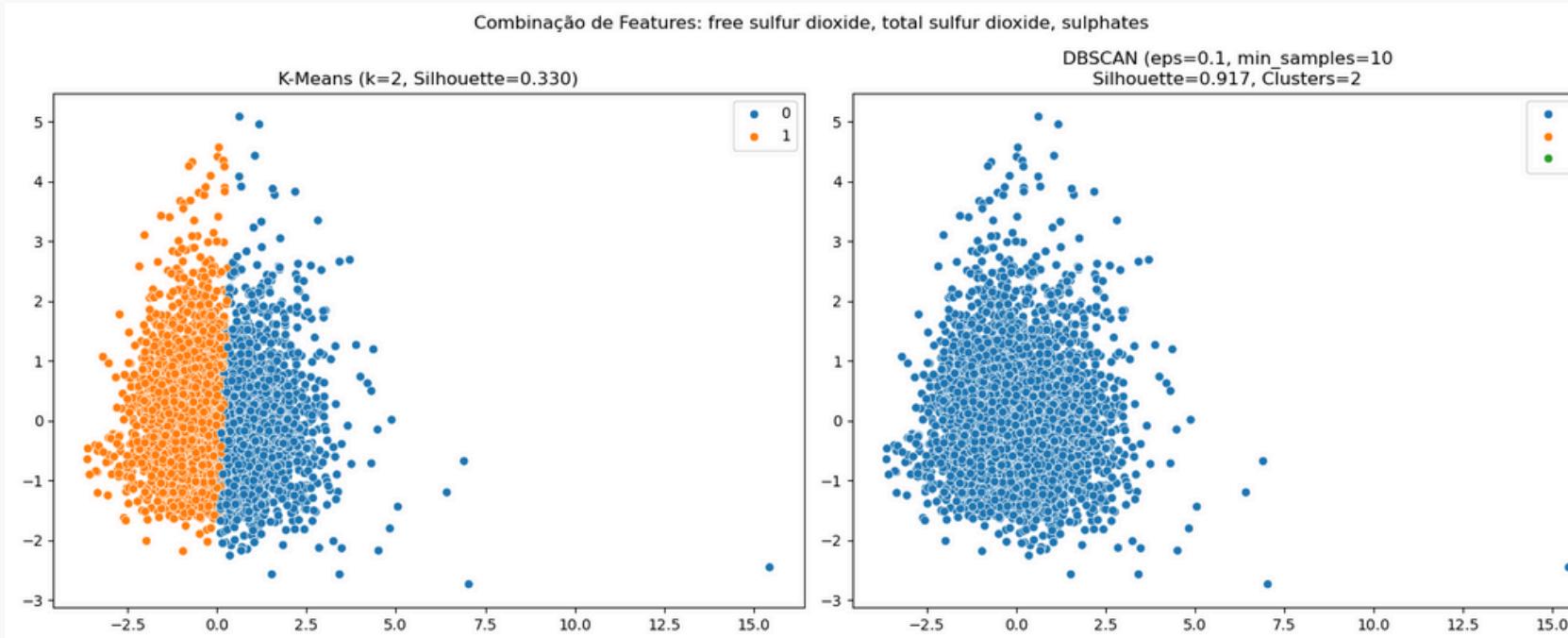
# Resultados

Combinações relacionadas à docura e corpo



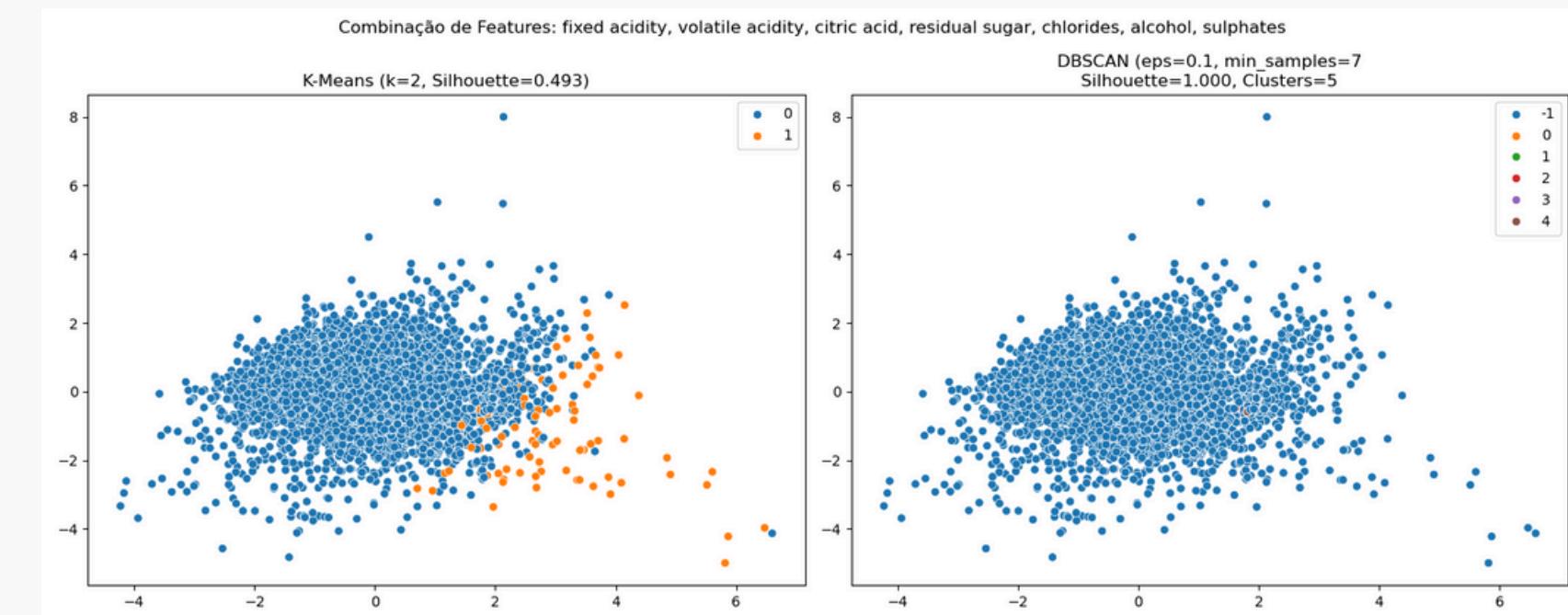
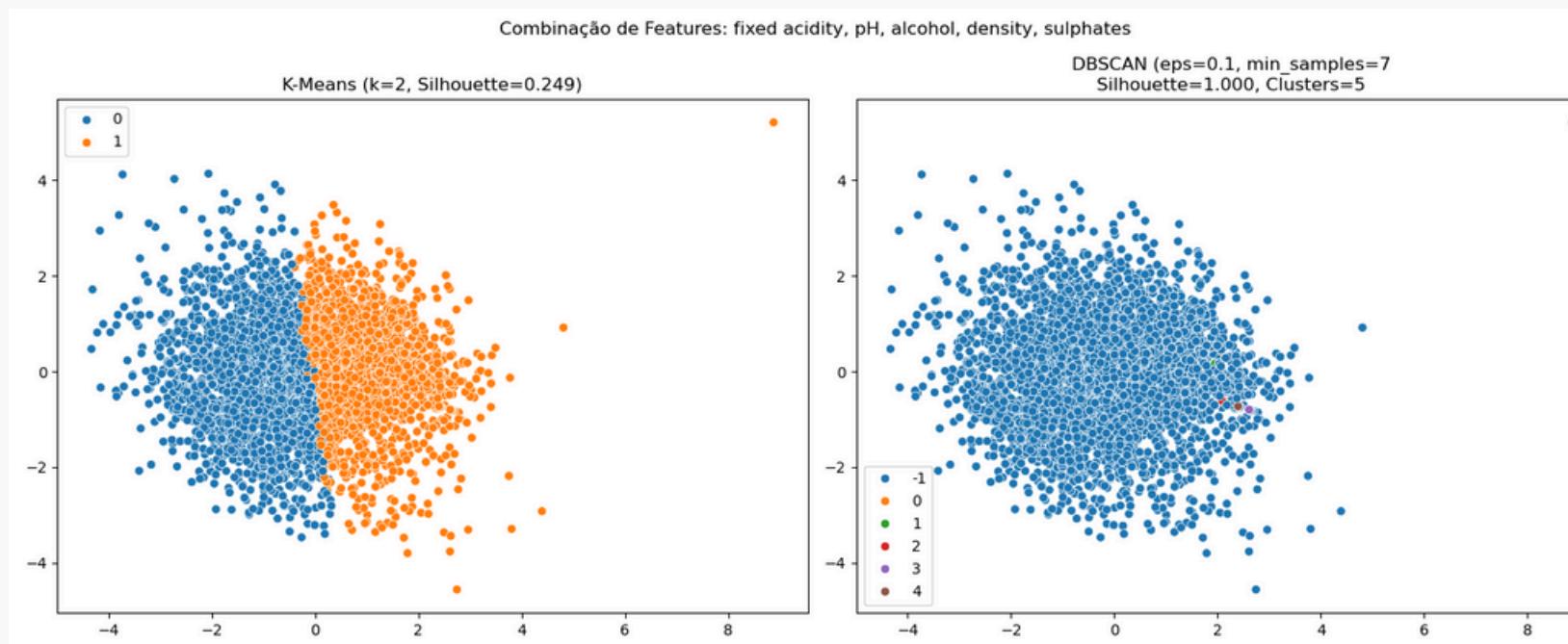
# Resultados

Combinações relacionadas a compostos sulfurados



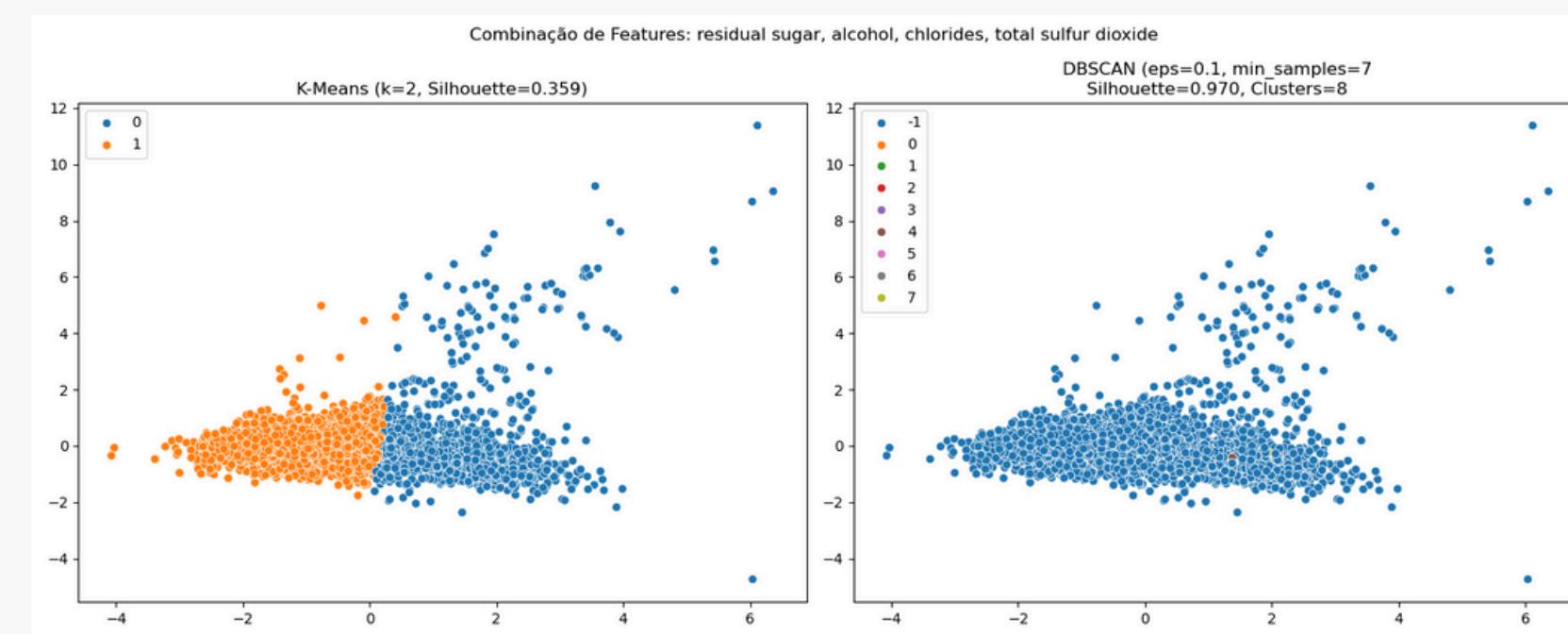
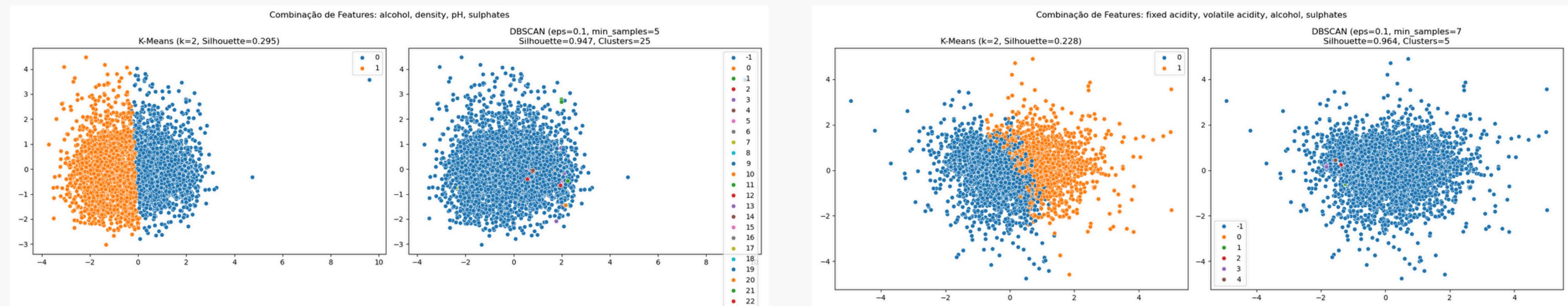
# Resultados

## Combinações mais complexas



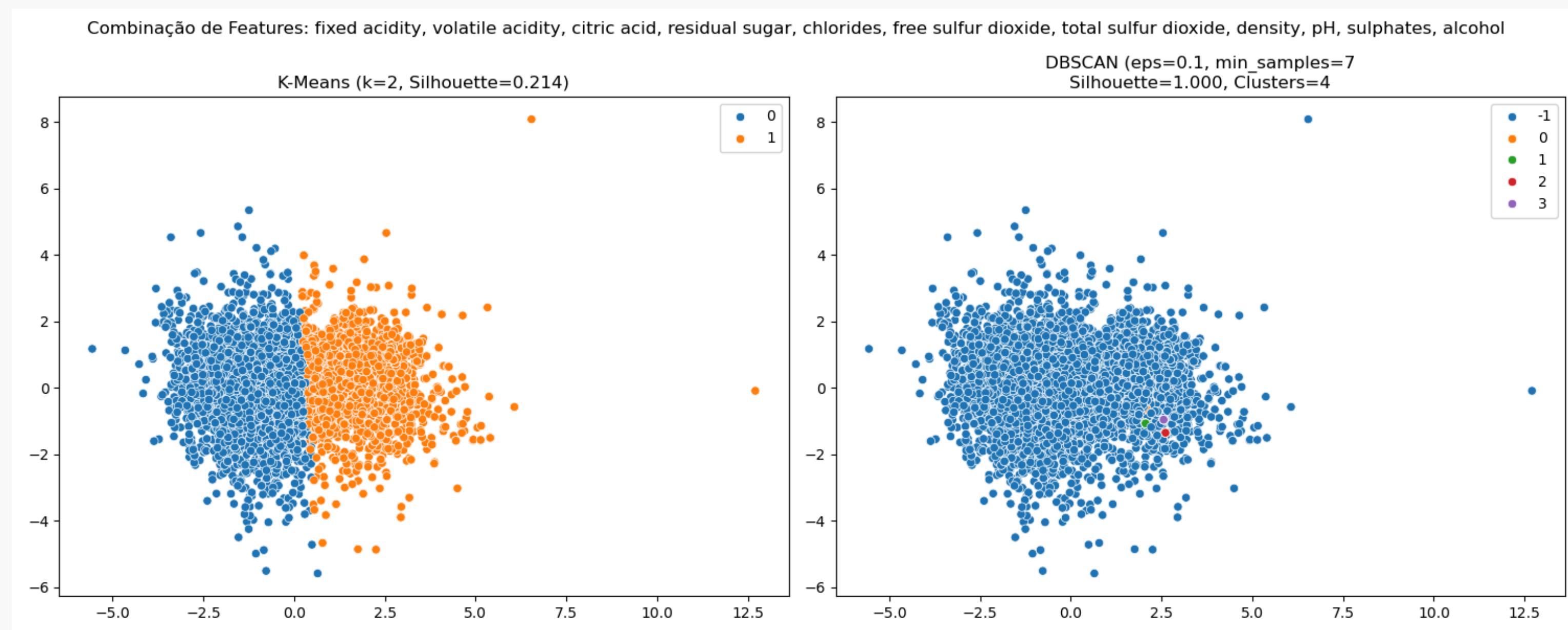
# Resultados

Combinações específicas para características distintas



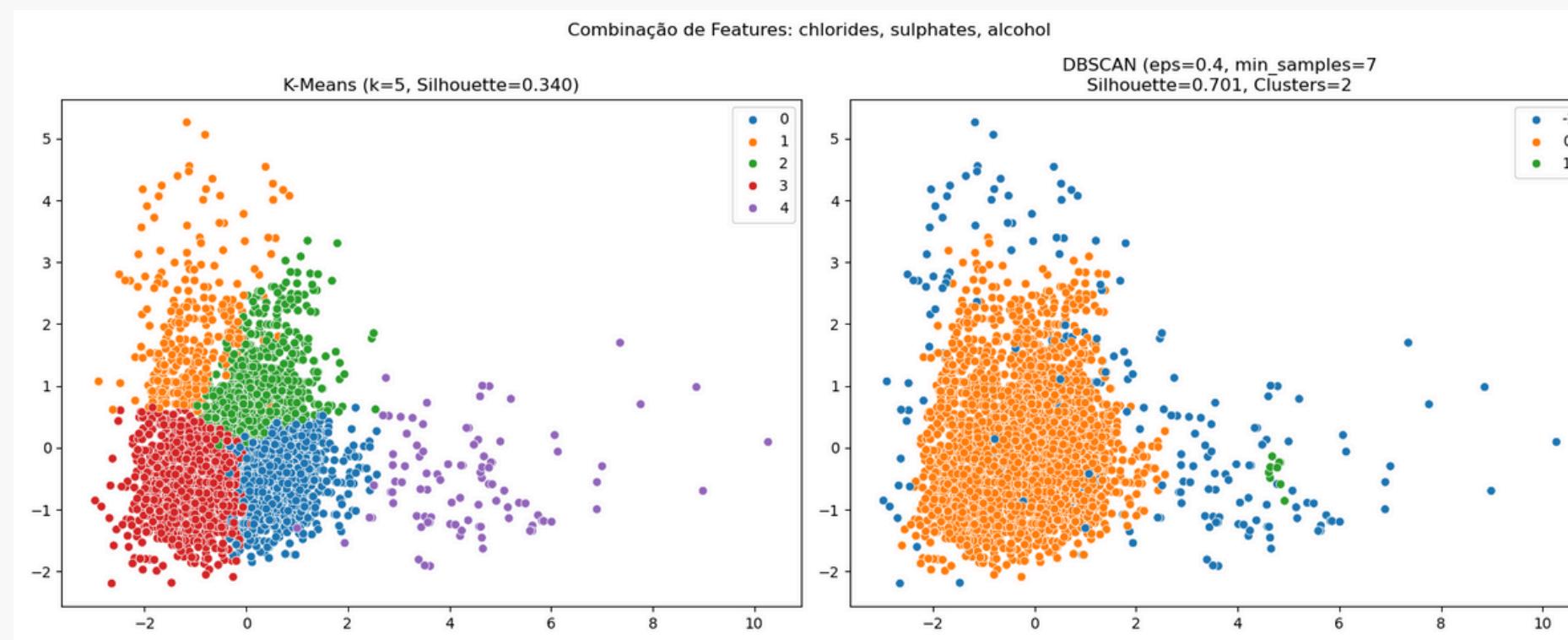
# Resultados

Combinação de todas as colunas (menos a quality)



# Conclusão

- Melhor agrupamento obtido (com base no agrupamento disponibilizado na tabela pela coluna quality que agrupa em valores inteiros de 3 à 8)



- Foram testados diversos valores de eps e min\_samples para o algoritmo de DBSCAN, mas os resultados de agrupamento não foram próximos ao esperado pela tabela usando a coluna quality como referência.
- K-Means deve um bom desempenho e bem próximo ao esperado



# Obrigado

