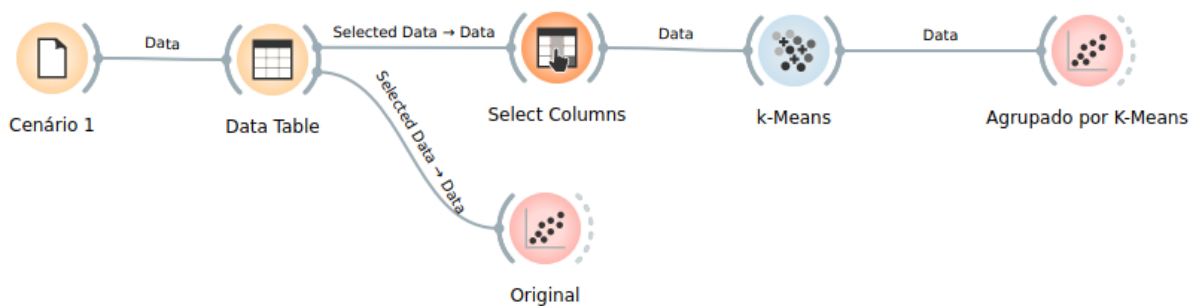


Caroliny Cardoso de França, 5º DSM, Fatec Araras

Gustavo Henrique Pinto, 5º DSM, Fatec Araras

Agrupamento

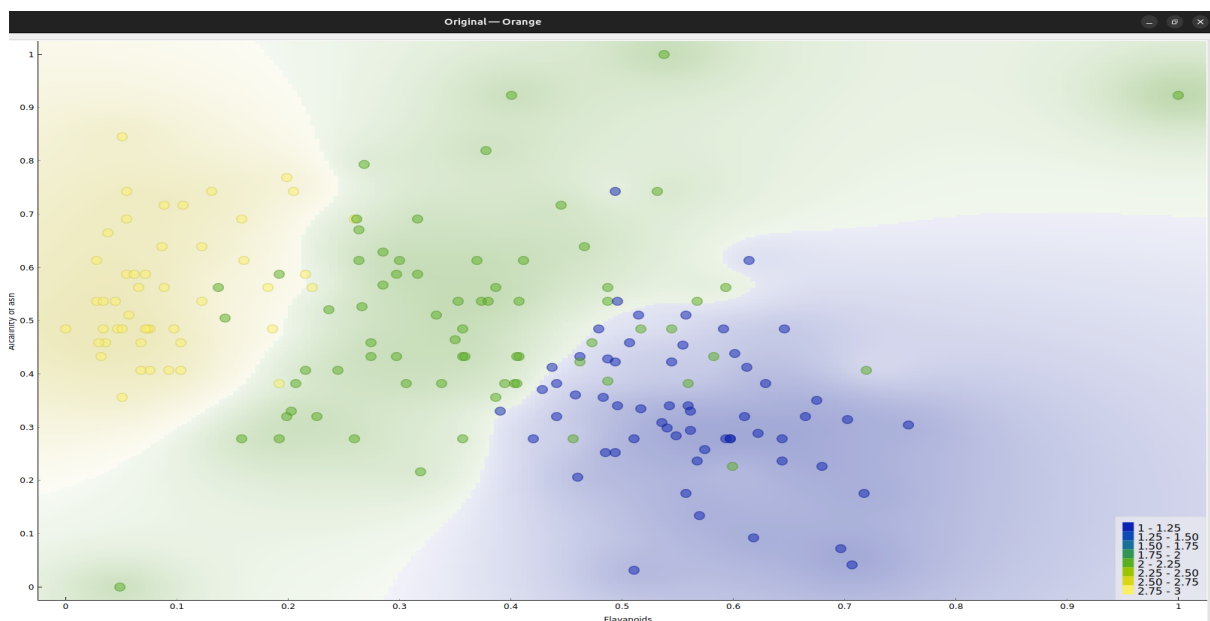
As análises de agrupamento foram realizadas através da ferramenta Orange Data Mining, baseando-se em um algoritmo K-Means para a realização automática do agrupamento dos dados em clusters (subconjuntos dos dados originais), implementando o seguinte fluxo em cada um dos conjuntos estudados.



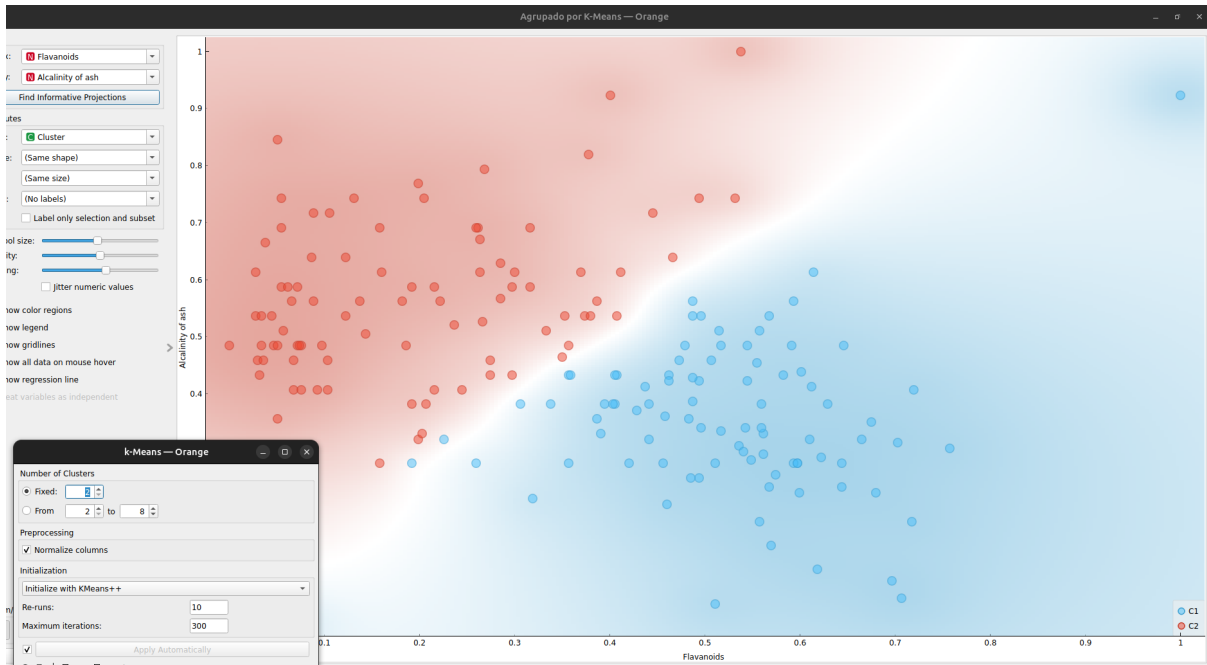
Cenário 1 – Três classes

O conjunto apresentado no primeiro cenário refere-se a uma coletânea de dados de três classes de vinhos, com a realização de quatro estudos de agrupamento:

- Conjunto original, sem a aplicação do K-Means:



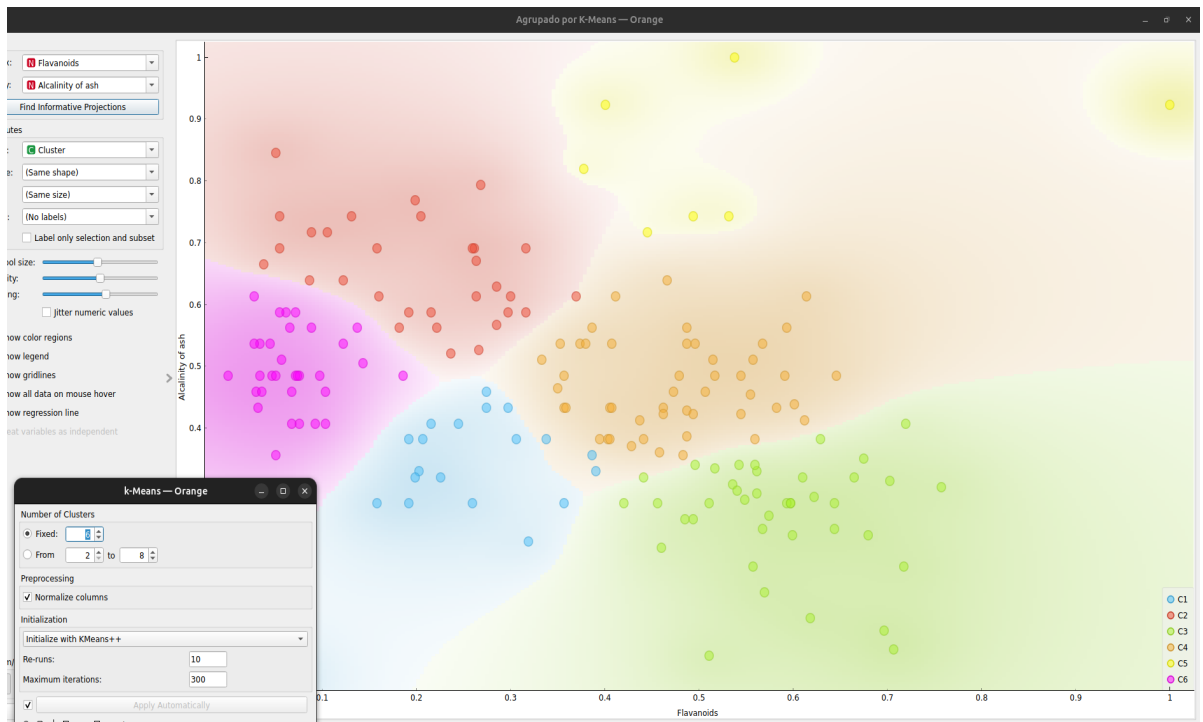
- Conjunto agrupado em dois grupos com K-Means, onde se notou uma clara divisão entre duas possíveis classes de vinho, com uma régua de divisão marcante, mas ainda contando com elementos espalhados:



- Conjunto agrupado usando K-Means com três grupos, onde foi possível perceber a redução do número de membros de cada conjunto, porém com mais uniformidade entre eles, também notou-se uma régua de divisão mais tênue entre os grupos, com bem menos elementos periféricos em relação aos demais.



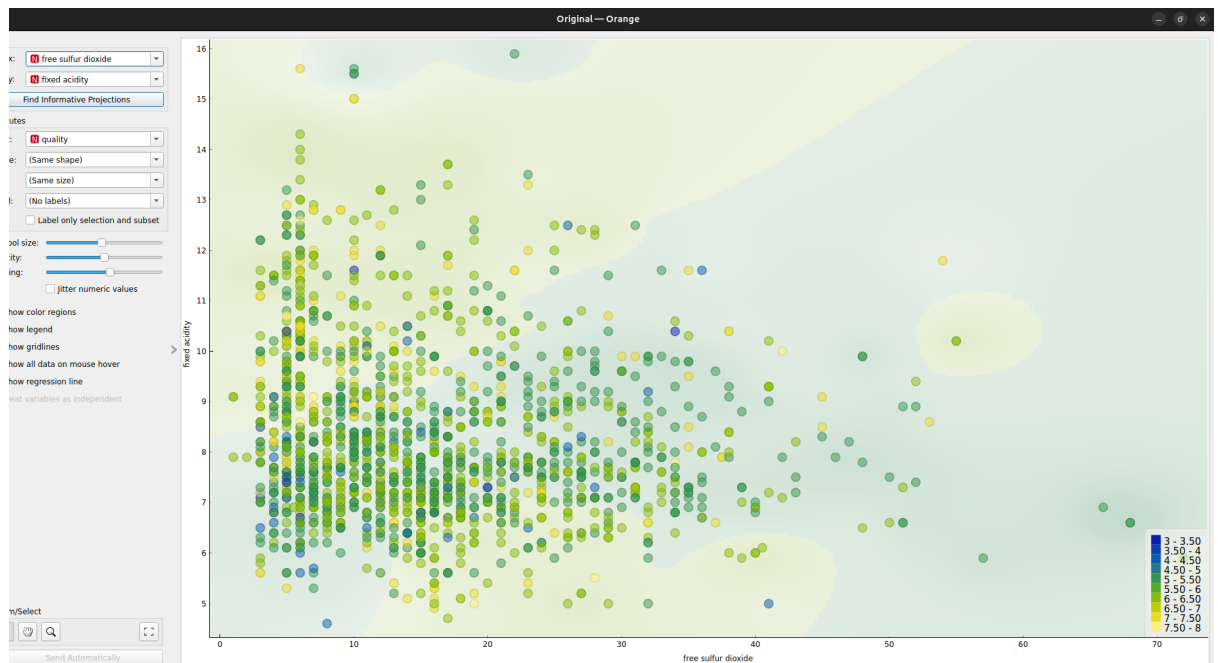
- Conjunto agrupado usando K-Means com seis grupos, com esse agrupamento os conjuntos e os dados tornaram-se granulares, e com divisões menos perceptíveis:



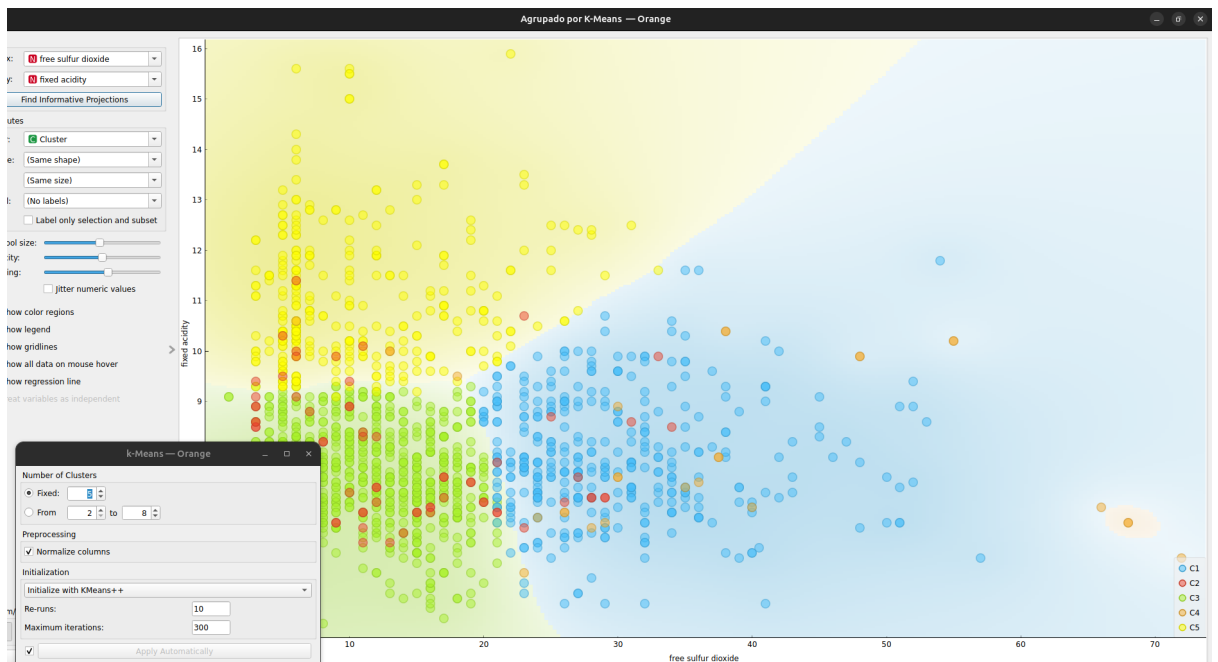
Cenário 2 - Dez classes

Nesse segundo cenário apresenta-se uma coletânea de dados com dez classes de vinhos, com base nas propriedades de vinhos, como a acidez e o nível de açúcar, assim como no primeiro conjunto foram realizados quatro estudos de agrupamento:

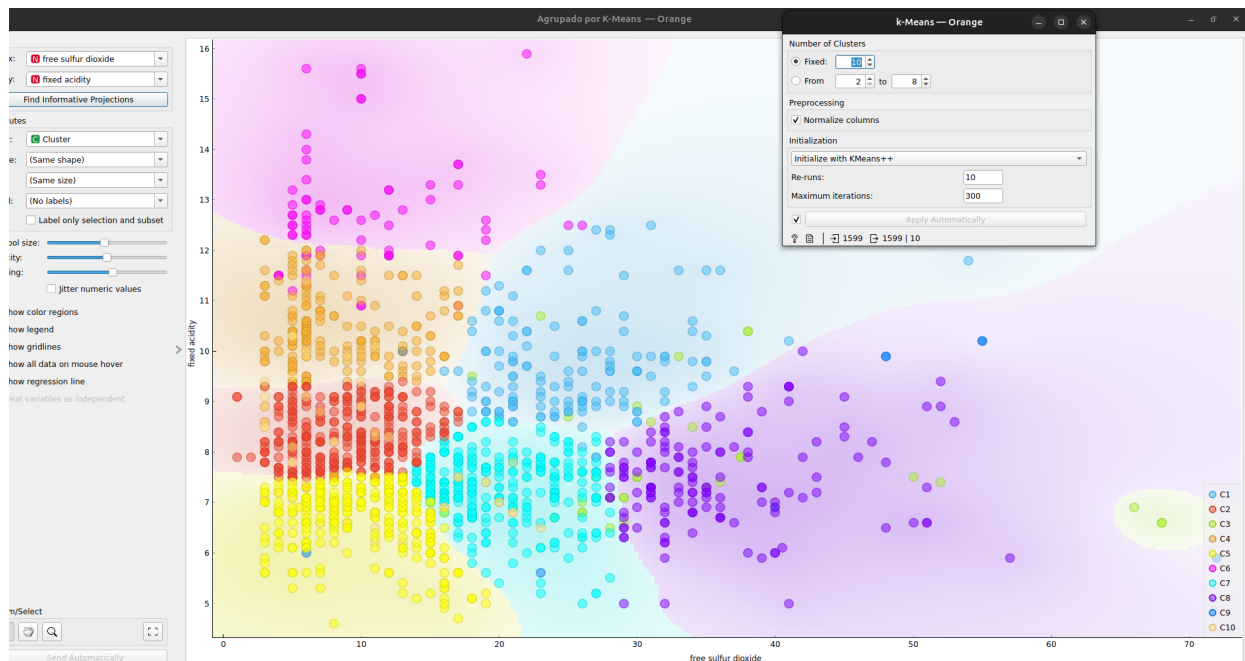
- Conjunto original, sem a aplicação do K-Means, nota-se que esse segundo conjunto de dados apresenta menos definição de seus agrupamentos, não sendo possível demarcar suas classes e subconjuntos:



- Conjunto agrupado usando K-Means com cinco grupos, nota-se a formação de três grandes subconjuntos marcantes, com outras duas classes presentes dissolvidas e esparsas:



- Conjunto agrupado por K-Means com dez grupos, nota-se a formação de sete grandes grupos, com outras três classes esparsas e dissolvidas, semelhante ao conjunto com cinco grupos:



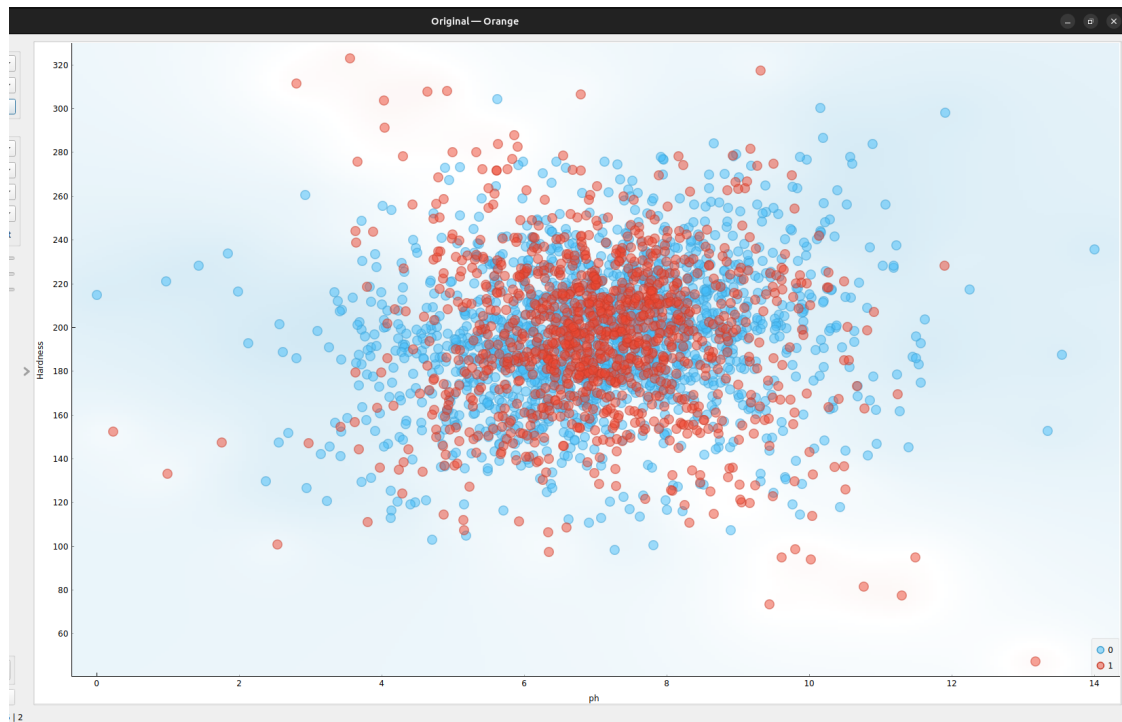
- Conjunto agrupado usando K-Means com quinze grupos, os subconjuntos formados apresentam características parecidas com o agrupamento de dez grupos, porém com uma divisão mais tênue entre cada um dos grupos:



Cenário três - Duas classes

O conjunto estudado no terceiro cenário apresenta dados sobre a potabilidade da água com base em suas propriedades químicas, classificando-os em duas classes, a de água potável e não potável.

- Conjunto original, sem a aplicação do K-Means, nota-se que esse terceiro conjunto apresenta uma versão original menos definida de seus agrupamentos, não sendo possível demarcar suas classes e subconjuntos:



- Conjunto agrupado com K-Means usando dois conjuntos, nota-se a formação inicial de dois grupos melhor definidos, porém ainda contando com elementos transitantes entre eles:



- Conjunto agrupado com K-Means utilizando seis subconjuntos, nota-se uma redução das separações entre os conjuntos, tornando quase impossível o discernimento de suas classes:

