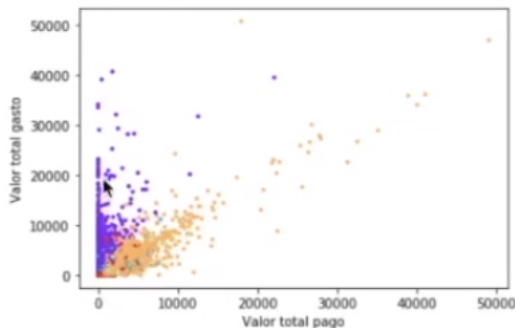


Interpretação visual

Transcrição

Conseguimos gerar o gráfico com sucesso, com duas dimensões.



Ao analisarmos o cluster roxo, notaremos que há uma tendência a desenhar uma linha reta na vertical, o que indica que o valor total gasto dos clientes aumenta, mas o valor total pago não acompanha essa ascendência. De maneira geral, o que conseguimos perceber é que o cluster representa os clientes que gastaram mais, contudo não pagaram a fatura ou a pagaram parcialmente.

Já o cluster representado pela cor laranja notaremos uma diagonal, o que indica que a medida que o valor gasto aumenta o valor pago também. Esse cluster representa os clientes que pagaram a fatura ou quase seu valor total.

Temos informações interessantes, mas isso não é o suficiente para compreender todo o conjunto. Plotaremos outras dimensões, mas sempre em pares, e assim conseguiremos interpretar outros atributos de maneira clara, visível.

Utilizaremos a biblioteca `seaborn` para plotar os atributos par a par. Atribuiremos uma nova coluna para o dataframe em que inseriremos os labels, afinal trata-se de um requisito para a função.

```
import seaborn as sns
dataframe["cluster"] = labels
sns.pairplot(dataframe[0:], hue="cluster")
```

Feito isso teremos um grande conjunto de gráficos, todos divididos em pares, cerca de 240. São muitos, e isso dificulta nosso processo de validação as informações com a visualização dessa forma.

A visualização nem sempre é uma técnica que irá nos auxiliar na interpretação de clusters. Nas próximas aulas vamos aprender algumas estratégias para guiar nossa interpretação.