

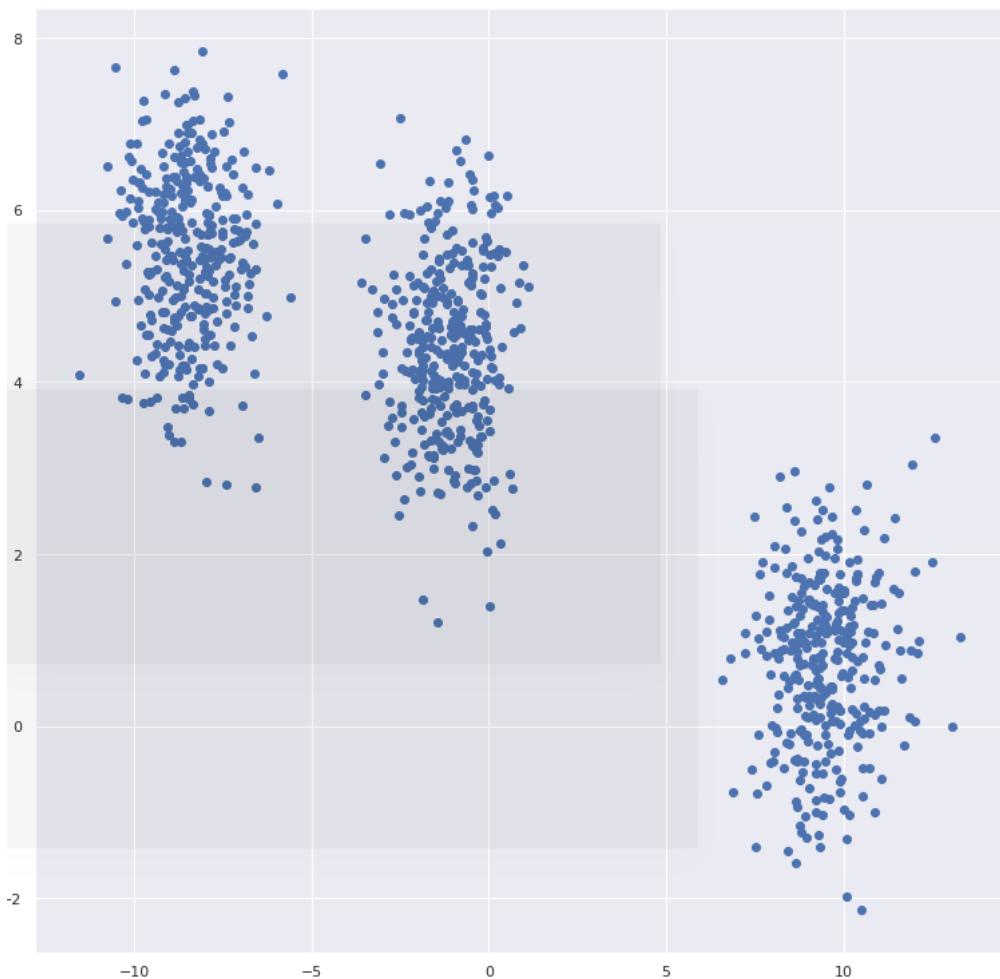
Outra forma de agrupamento

No curso vimos duas formas de agrupamento, o particionado e o hierárquico. Porém, existem outra forma de agrupar os dados.

Vamos lá. Em todos os exemplos precisamos passar o número de clusters, mas o que acontece quando não temos esse número?

O agrupamento por densidade tenta resolver isso. Nele, não passamos o número de clusters e ele tentará agrupar os dados pela densidade dos pontos.

Vamos pegar os mesmos blobs que geramos no exercício de visualizar os centróides no K-Means:



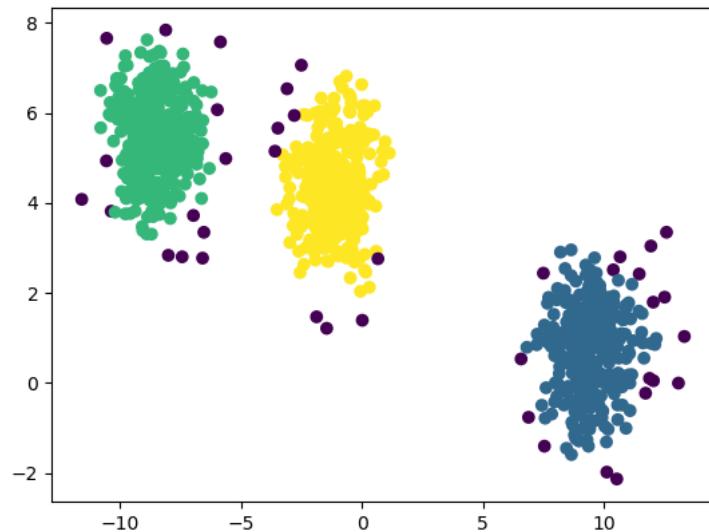
Vamos pedir para o Python importar a classe DBSCAN do módulo de cluster.

```
from sklearn.cluster import DBSCAN
```

Vamos criar o modelo e falar para ele agrupar os dados:

```
modelo = DBSCAN()  
grupos = modelo.fit_predict(dados)
```

Plotando e colorindo os dados com o valor do grupo, temos o seguinte resultado:



Podemos ver que ele conseguiu agrupar em três grupos distintos, mas o que são esses pontos que não estão nas cores dos grupos?

Esses são os outliers. São os pontos que o DBSCAN achou que não pertenciam aos grupos que estavam próximos.

Para quem tiver interesse em aprender mais sobre esse algoritmo, no mesmo site que usamos para ver o funcionamento do K-Means, também podemos ver o DBSCAN em ação: <https://www.naftaliharris.com/blog/visualizing-dbscan-clustering/> (<https://www.naftaliharris.com/blog/visualizing-dbscan-clustering/>).