DBSCAN ALGORITHM

The Noisy Application Density-Based Spatial Clustering Algorithm or DBSCAN is a density-based clustering algorithm that finds the number of clusters, estimates the distribution of the remaining nodes.

It is extremely popular what is sought is to find a density function that estimates the clusters.

DBSCAN, unlike K-means, does not need the desired number of clusters. Also, those clusters are not biased to specific geometric shapes like other clustering algorithms. Also, it has the notion of noise and is not amenable to the order of the points in the database

The problem with this algorithm is that it depends on the notion of distance, the Euclidean norm, but other types can be used. When they have a high dimension, the distance tends to be useless due to the curse of dimensionality. When working with densities DBSCAN cannot group well sets with difference in densities.

In this case we use a sales database with gender, age, annual income, and spending score parameters. For that, since we are going to use the standard scaler since we are going to use the notion of distance, it is important to standardize the data. To do this we implement the dbscan algorithm in python and compare it with the dbscan from the scikitlearn library.

Metric	score
homogeneity score:	0.5428869121773657
rand score:	0.8540201005025125
completeness score:	0.8668435836885163

Metrics, these metrics compare homogeneity, completeness, and randomness, with 0 being the worst and 1 being the best. We see that the score of rand and completeness is close to 1, so both algorithms obtain similar solutions, but the homogeneity is close to 0.5, therefore, it is not very homogeneous with respect to the other. therefore, it can be concluded that this algorithm created in this laboratory is a quite competent solution for the DBSCAN algorithm.