**Feature Engineering**

When you use domain knowledge to choose which data metrics to use as input features into a machine learning algorithm, you are using feature engineering. Feature engineering plays a key role in k-means clustering by capturing the variability of the data that is essential for the algorithm to find all of the natural-occurring groups.

**Density-Based Spatial Clustering of Applications with Noise (DBSCAN)**

DBSCAN begins with an arbitrary starting data point that has not been visited. If there are a sufficient number of points with this neighborhood, then the clustering process starts and the current data points becomes the first point of the new cluster. Otherwise, the point will be labelled as noise. ~~~~~ expand

One of the major advantages of DBSCAN is that is does not require a pre-set number of clusters. Another advantage is that it identifies outliers as noise. Points in low-density regions are classified as noise and omitted.

The main drawback of DBSCAN is that it doesn’t perform as well as others when the clusters are of varying density.