**Clustering** is exactly what we want from unsupervised learning, but exactly how can we determine the clusters?

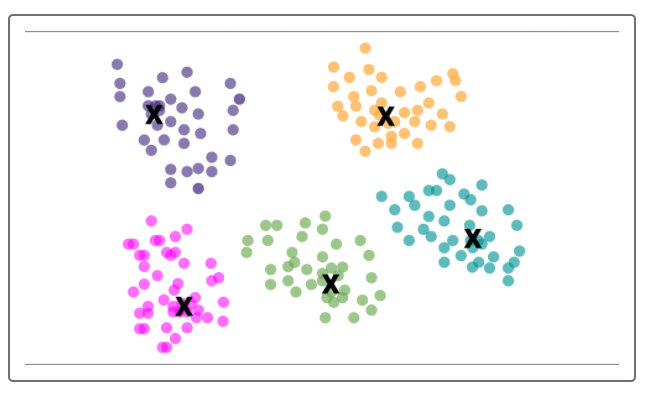
One of the most popular ways to cluster is by using the K-means algorithm.

K-means is an unsupervised learning algorithm used to identify and solve clustering issues.

**K** represents how many clusters there will be. These clusters are then determined by the **means** of all the points that will belong to the cluster.

The K-means algorithm groups the data into K clusters, where belonging to a cluster is based on some similarity or distance measure to a centroid.

A **centroid** is a data point that is the arithmetic mean position of all the points on a cluster:



The centroid is found by taking the mean of all the x values in a cluster, and the mean of all the y values in a cluster.