This dataset explores the subjective quality assessment of digital colposcopies, studied at Hospital Universitario de Caracas.

The dataset has three modalities (i.e. Hinselmann, Green, Schiller).

Number of Attributes: 69 (62 predictive attributes, 7 target variables).

**Non-supervised mining - Cluster**

**Clustering** is the task of grouping a set of objects in such a way that objects in the same group (called a **cluster**) with similar traits.

* **KMeans**

The goal of KMeans Clustering is to find groups in the data, where the number of groups is given by the K variable. The algorithm assigns each data point to one of the K groups based on the distance to the cluster center. To discover the K variable, we used the Elbow Curve Method, assigning different values to K and using the Euclidean distance between each data point with the cluster center. Plotting a graph with the variation of K and the distance, we can notice that the distance decreases as K gets larger. This happens because when you have more clusters, they will be smaller, so the distance between the data points and the cluster center will also be smaller.