**Kmeans Algorithm**

Questions:

* How is similarity defined?

Definition:

* K-means is simple: group similar data points together and discover underlying patterns. To achieve this objective, K-means looks for a fixed number (k) of clusters in a dataset.”
* K-means is an unsupervised learning model, which means that it has no target variable and does not require labeled data.
* Define target number K, which refers to the number of centroids.
* Every data point is allocated to each of the clusters through reducing the in-cluster sum of squares.
* In other words, the K-means algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible.
* The ‘means’ in the K-means refers to averaging of the data; that is, finding the centroid.

**References**:

Kmodes:

<https://analyticsdefined.com/using-k-modes-clustering-categorical-data/>

Kmeans (python)

<https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1>

Kmeans (R)

<https://www.datacamp.com/community/tutorials/k-means-clustering-r>