**K-Means Algorithm:**

K-means clustering is a type of unsupervised learning used for grouping unlabelled data (i.e., data without defined categories or groups). The goal is to find groups in the data and ‘K’ represents the number of groups in that data. The algorithm works iteratively to assign each data point to one of K groups based on the features in the dataset. Data points are clustered based on feature similarity.

**Input:** No of clusters and the dataset

**Steps:**

#1 - Randomly assigns values to the centroid (No of centroids = K) from the dataset

#2 - : each data point is assigned to its nearest centroid, based on the squared Euclidean distance.

#3 - In this step, the centroids are recomputed. This is done by taking the mean of all data points assigned to that centroid's cluster.

The algorithm iterates between steps one and two until a stopping criteria is met (i.e., no data points change clusters, the sum of the distances is minimized, or some maximum number of iterations is reached).

**Result:**

1. The centroids of the *K* clusters, which can be used to label new data
2. Labels for the training data (each data point is assigned to a single cluster)

**Choosing K:**

The optimal K for a given problem statement can be found using the *Elbow Method* – In this method the graph between the average within-cluster distance to the centroid vs K is plotted and the point (K) where the graph sharply drops is chosen as K

**Use Case:**

To find the number of clusters/ no of hackers using the data retrieved from the hacking attacks

**Dataset: hack\_data.csv**

* Session\_Connection\_Time (How long the session lasted in minutes)
* Bytes Transferred (Megabytes transferred during session)
* Kali\_Trace\_Used (Whether the hacker was using Kali Linux)
* Servers\_Corrupted (Number of server corrupted during the attack)
* Pages\_Corrupted (Number of pages illegally accessed)
* Location (Location attack came from)
* WPM\_Typing\_Speed (Estimated typing speed based on session logs)

**Result: scaled\_data.csv**