

# K-MEANS CLUSTERING

## K-MEANS CLUSTERING THEORY

K-Means Clustering is an unsupervised (it takes in unlabeled data) learning algorithm that will attempt to group similar clusters together in your data. The overall goal is to divide data into distinct groups such that observations within each group are similar.

The K-Means Algorithm,

- (i). Choose a number of clusters “k”.
- (ii). Randomly assign each point to a cluster.
- (iii). Until clusters stop changing, repeat the following:

For each cluster, compute the cluster centroid by taking the mean vector of points in the cluster.

Assign each data point to the cluster for which the centroid is the closest.

## K-MEANS CLUSTERING WITH PYTHON

**Data Source:** <https://www.kaggle.com/kushal1996/customer-segmentation-k-means-analysis>

Mall Customer Segmentation Data - The objective of the work is to perform customer segmentation and analysis with k-means clustering algorithm.

### Columns

**CustomerID:** Unique ID assigned to the customer

**Gender:** Gender of the customer

**Age:** Age of the customer

**Annual Income (k\$):** Annual Income of the customer

**Spending Score (1-100):** Score assigned by the mall based on customer behavior and spending nature