

K Means Clustering:

- * K means clustering is a unsupervised machine learning algorithm
- * It allow us to cluster the data into different groups and a convenient way to discover the categories of the groups in an unlabelled dataset on its own without the need for any training
- * It is a centroid based algorithm

STEPS TO BE FOLLOWED:

1. Select the number of K to decide the number of cluster
2. Select the random K points (or) centroid [It is from the input dataset]
3. Assign each datapoint to their closest centroid
4. Calculate the variance and place a new centroid of each cluster
5. (Repeat the 3rd step) ~ Reassign each datapoint to their new closest centroid of each cluster

In []:

In [1]:

```
install.packages("factoextra")
```

Warning message:

"dependency 'FactoMineR' is not available"also installing the dependencies 'rprojroot', 'fs', 'diffobj', 'rematch2', 'brio', 'callr', 'desc', 'pkgload', 'praise', 'processx', 'ps', 'waldo', 'evaluate', 'highr', 'yaml', 'xfun', 'Matrix', 'testthat', 'numDeriv', 'knitr', 'SparseM', 'MatrixModels', 'survival', 'boot', 'minqa', 'nloptr', 'RcppEigen', 'backports', 'carData', 'nnet', 'pbkrtest', 'quantreg', 'lme4', 'cli', 'pillar', 'broom', 'corrplot', 'car', 'glue', 'lifecycle', 'rlang', 'scales', 'vctrs', 'viridis', 'ggsci', 'purrr', 'dplyr', 'cowplot', 'ggsignif', 'gridExtra', 'polynom', 'rstatix', 'plyr', 'stringr', 'ellipsis', 'tidyselect', 'ggplot2', 'abind', 'cluster', 'dendextend', 'ggpubr', 'reshape2', 'ggrepel', 'tidyr'

There are binary versions available but the source versions are later:

	binary	source	needs_compilation
rprojroot	2.0.2	2.0.3	FALSE
fs	1.5.0	1.5.2	TRUE
diffobj	0.3.4	0.3.5	TRUE
brio	1.1.2	1.1.3	TRUE
callr	3.7.0	3.7.3	FALSE
desc	1.3.0	1.4.2	FALSE
pkgload	1.2.1	1.2.2	FALSE

In [2]:

```
install.packages("cluster")
```

There is a binary version available but the source version is later:

	binary	source	needs_compilation
cluster	2.1.2	2.1.4	TRUE

Binaries will be installed

package 'cluster' successfully unpacked and MD5 sums checked

The downloaded binary packages are in

C:\Users\nanth\AppData\Local\Temp\RtmpauCbeZ\downloaded_packages

In [3]:

```
install.packages("ggplot")
```

Warning message:

"package 'ggplot' is not available (for R version 3.6.1)"

In [4]:

```
# including the libraries
library(cluster)
library(ggplot)
library(factoextra)
```

Warning message:

"package 'cluster' was built under R version 3.6.3"

Error in library(ggplot): there is no package called 'ggplot'

Traceback:

1. library(ggplot)

In []:

```
# Group the clusters
```

In []:

```
# cluster
# 2 -> How many clusters
# 25 -> Random 25 values
Kclusters=kmeans(df,2,nstart=25)
Kclusters
```

In []:

```
# Visualising the cluster
fviz_cluster(Kcluster,)
```