

```
In [1]: from IPython.display import Image  
Image(filename='logo.PNG', height=340, width=900)
```

Out[1]:



## Introduction to Clustering

- Clustering is “the process of organizing objects into groups whose members are similar in some way”.
- A cluster is therefore a collection of objects which are “similar” between them and are “dissimilar” to the objects belonging to other clusters.

## Applications of Clustering

- **Business** : Businesses collect large amounts of information about current and potential customers. **Clustering can be used to segment customers into a small number of groups for additional analysis and marketing activities**

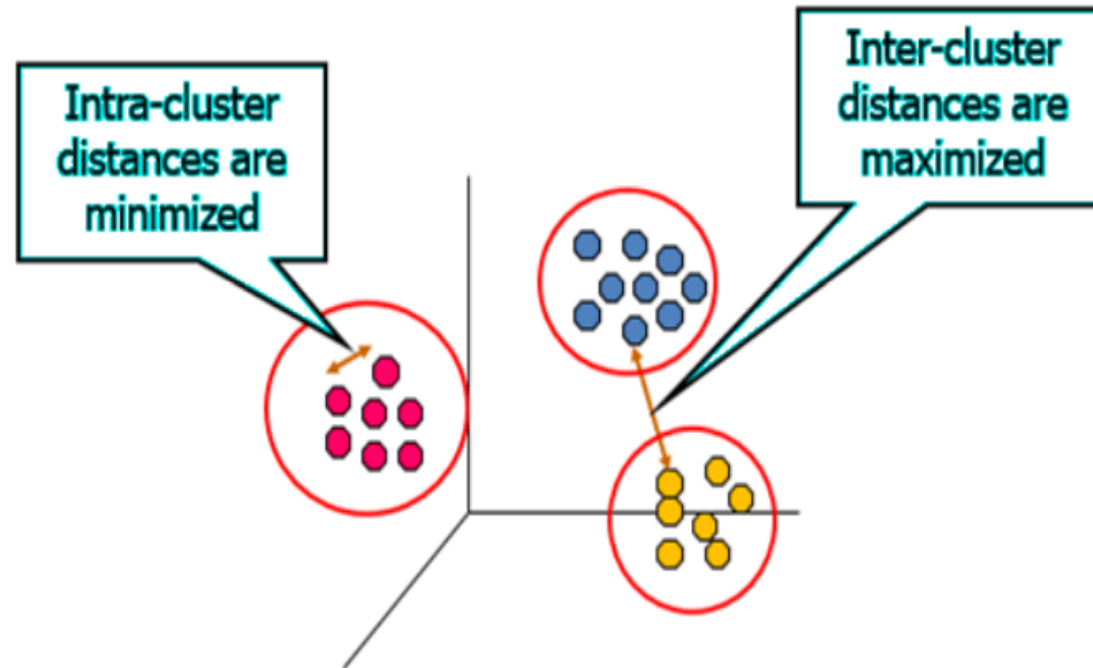
- **Marketing** : Finding groups of customers with similar behavior given a large database of customer data containing their properties and past buying records.
- **Climate** : Understanding the Earth's climate requires finding patterns in the atmosphere and ocean. To that end, cluster analysis has been applied to **find patterns in atmospheric pressure and ocean temperature that have a significant impact on climate**.
- **Psychology and Medicine** : An illness or condition frequently has a number of variations, and cluster analysis can be used **to identify these different subcategories**. For example, clustering has been used to identify different types of depression. Cluster analysis can also be used to detect patterns in the spatial or temporal distribution of a disease

## OBJECTIVE OF CLUSTER ANALYSIS

- Intra cluster distance is the sum of distances between objects in the same cluster
- This distance should always be **minimized**.
- Inter cluster distance is the distance between objects in the different cluster.
- This distance should always be **maximized**.

```
In [2]: from IPython.display import Image  
Image(filename='objective.PNG', height=340, width=900)
```

Out[2]:



In [ ]: