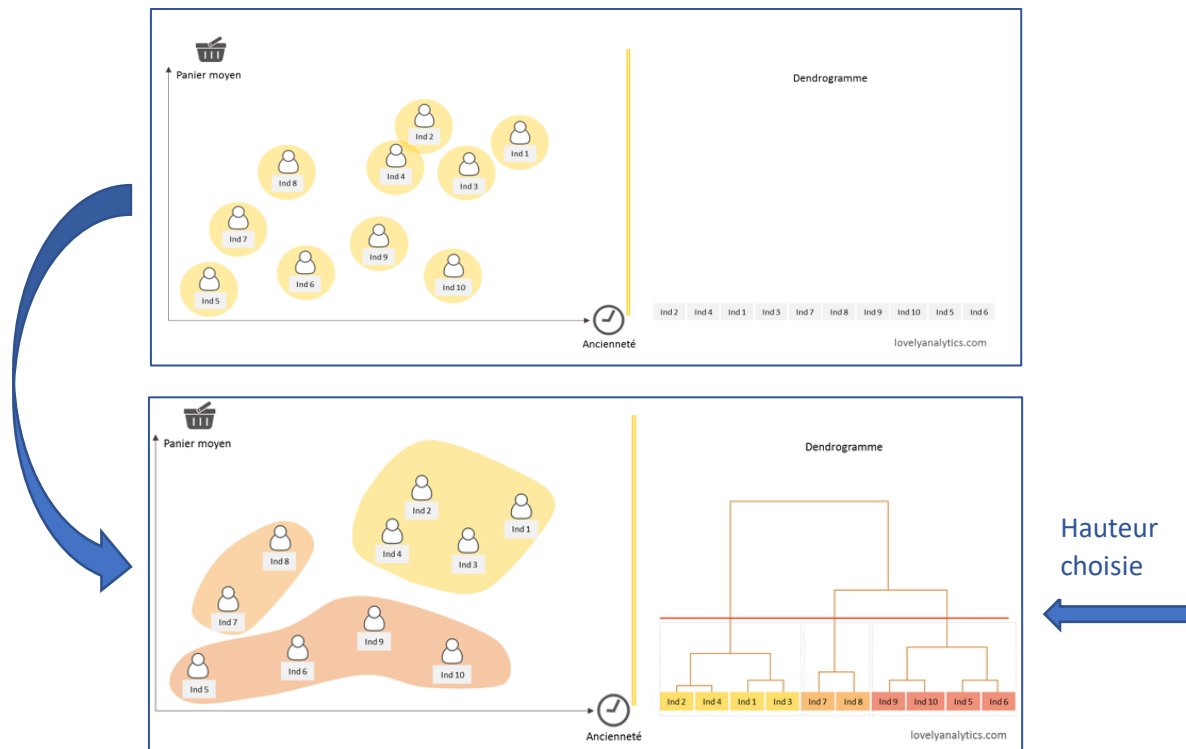


### Classification Ascendante Hiérarchique ou Regroupement Hiérarchique (Hierarchical Clustering) :

Le principe de la CAH est de rassembler des individus selon un critère de ressemblance défini au préalable qui s'exprimera sous la forme d'une matrice de distances, exprimant la distance existante entre chaque couple d'individus. Deux observations identiques auront une distance nulle. Plus les deux observations seront dissemblables, plus la distance sera importante.

La CAH va ensuite rassembler les individus de manière itérative afin de produire **un dendrogramme ou arbre de classification**. La classification est ascendante car elle part des observations individuelles ; elle est hiérarchique car elle produit des classes ou groupes de plus en plus vastes, incluant des sous-groupes en leur sein. **En découpant cet arbre à une certaine hauteur choisie, on produira la partition désirée.**



#### Algorithme :

**Entrées** : un ensemble E de N données;

**Sorties** : Hiérarchie de groupes (dendrogramme) ;

1. Chaque donnée définit un cluster ;
2. tant que (nombre de clusters > 1) faire
  1. Calcul des distances entre tous les groupes issus de l'itération précédente ;
  2. Regroupement des 2 groupes ayant la plus petite distance;

fin tant que

**Avantages** : il est possible de

- observer l'ordre des agrégations/fusions de groupes
- examiner les rapports des similarités entre groupes
- obtenir plusieurs partitionnements à des niveaux de granularité différents

**Inconvénients** :

- Coûteux en temps de calcul

#### Paramètres importants :

Le choix de la distance entre les groupes (la mesure de la dissimilarité entre groupes ou indice d'agrégation) a un impact parfois significatif sur les résultats obtenus.

Exemples : le lien minimum (single linkage) , le lien maximum (complete linkage) ,le lien moyen (average linkage) ...