

TD 8.10 - Analyse de données

Le tri croisé consiste à considérer une donnée avec plusieurs dimensions :

- des factures avec leurs dates, leurs clients, les produits considérés...
- des publicités avec leurs nombres de vues, leur durée, le produit concerné, le nombre de réalisation...
- des matches sportifs avec leur résultat, les joueurs impliqués...

On tente d'analyser les données en fonction des dimensions d'analyse pour voir lesquelles sont pertinentes et permettent de voir des différences sur la donnée. Par exemple, on va peut être constater, pour des publicités, que la probabilité d'actes d'achat est corrélée à la durée de vue/lecture. On va peut être constater que les factures sont moins honorés par certains types de clients...

Le but est de produire de l'information à partir de données, cette information permettant ensuite de prendre des décisions stratégiques pertinentes pour l'organisation. Cette démarche est l'un des aspects du Big Data qui prend de plus en plus d'importance et nous vaut des formules saisissantes qui voient les données comme le "pétrole du XXIème siècle".

Excel offre de faire du tri croisé dynamique au travers des tableaux croisés dynamiques. Pour faire un tableau croisé dynamique, on sélectionne les données que l'on veut analyser puis on utilise l'onglet insertion → Tableau croisé dynamique. On peut ensuite se laisser guider par l'interface qui est plutôt ergonomique.

Ex 1.

A partir du fichier TD 8.10_1.xls qui contient des données sur des factures de vente,

- 1) Créer un tableau croisé dynamique.
- 2) Faire évoluer le tableau pour avoir en abscisse les catégories de produits, en ordonnée les clients et pour chaque client et catégorie de produit, le montant total des ventes.
- 3) Filtrer les résultats en ne les gardant que pour une journée particulière. On ôtera ensuite le filtre.
- 4) Modifier la donnée qui apparaît pour avoir la moyenne des ventes par client et catégorie de produit et non plus le montant total.
- 5) Rajouter une dimension d'analyse.

Ex 2.

A partir du fichier TD 8.10_2.xls qui contient des données sur les SICAV. Les SICAV sont des sociétés réunissant des investisseurs pour qu'ils aient une masse critique pour investir en bourse. Il existe beaucoup de SICAV, gérées par différents organismes financiers (banque, assurances, autres...), il peut être intéressant de mener une analyse pour trouver celles qui sont les plus appropriées pour les investisseurs.

- 1- Procéder au filtrage des données : on sélectionne la ligne 3 sur l'ensemble des colonnes remplies et on clique sur Filtres depuis le menu Données.
- 2- Afficher seulement les sicav telles que le signal d'achat est supérieur à 70 pour fin juillet 2006 et fin aout 2006.
- 3- Afficher seulement les SICAV françaises en utilisant la colonne "Code ISIN". Combien de valeurs sont

sélectionnées ?

4- Effacer les filtres précédents. Afficher les valeurs telles que leur signal d'achat est toujours supérieur à 30 à chaque fin de mois. Combien de valeurs sont sélectionnées ?

5- Copier **en valeur** les données qui concernent le mois de juillet 2006 dans une nouvelle feuille, reporter également les noms des sicav et leurs codes. Dans une nouvelle colonne étiquetées "pays", on affiche les deux premières lettres de chaque code ISIN. Les colonnes dont on dispose : Sicav / Code ISIN / rang / valeur / signal d'achat / pays. Cette dernière colonne contient un indice pays ("FR" pour France, "LU" pour Luxembourg etc...)

6- Créer un tableau croisé dynamique sur l'ensemble des colonnes et des lignes..

7- Afficher la moyenne des indices d'achat et le nombre des valeurs pour chaque indice pays disponible.

Ex 3.

A partir du fichier TD 8.10_3.xls , on propose de faire quelques recherches en utilisant un tableau croisé dynamique :

1- Quel est l'âge moyen des agents par site

2- Quel est le salaire moyen par âge ?

3- Quel est le salaire moyen par âge et par site ?

4- Quel est le salaire moyen des femmes par âge et par site ?

Ex 4.

Le fichier TD 8.10_4.xls contient des données publiques sur les financements accordés par l'État à diverses associations (récupération depuis le site data.gouv.fr)

Analyser la donnée de manière à faire apparaître des éléments intéressants/pertinents.