Analyse des données

Définitions:

- **Données aggrégées** : données regroupées en classes (clusters), éventuellement organisées de façon hiérarchiques. Possibilité d'appartenances à de multiples hiérarchies (cubes de données).
- **Analyse des données** : le but est de dégager des indicateurs à partir d'un grand ensemble de données, afin de faciliter la prise de décision.

cas d'utilisation :

- quels sont les magasins les plus rentables? doit-on ouvrir / fermer des magasins?
- où doit-on implanter un nouveau magasin?
- y a-t-il une corrélation entre le lancement d'une campagne publicitaire et les chiffres de vente? quels sont les supports les plus efficaces?
- quelle est la liste des clients à fidéliser?
- de quelle quantité doit-on approvisionner les magasins en fonction de la période de l'année?

analyse:

- quels sont les catégories de films/livres les plus fréquemment empruntés?
- réussite / taux d'embauche / salaire en fonction de la prépa d'origine / sexe / profession des parents

Les opérateurs d'aggrégation permettent de réaliser des statistiques sur les données, sous forme d'histogrammes (ou camemberts) organisés selon des catégories définies par les valeurs de certains attributs:

principe :

- opérateur d'aggrégation : comptage, somme, moyenne, ecart-type (count, sum, mean, avg, ...)
- les données aggrégées sont de type quantitatif
- les attributs définissant les classes sont de type qualitatif.

Exemples de requêtes faisant appel aux fonctions d'aggrégation :

Nombre d'élèves par groupe de TD / par prepa d'origine etc..:

```
SELECT groupe_TD , COUNT(num_eleve)
FROM Eleve
GROUP BY groupe_TD
```

Donner les chiffres des ventes du magasin pour chaque mois de l'année

```
SELECT mois, SUM(montant)
FROM Vente
GROUP BY mois
```

Donner le nombre de ventes d'un montant > à 1000 euros pour chaque mois de l'année

```
SELECT mois, COUNT(num_vente)
FROM Vente
GROUP BY mois
HAVING montant >= 1000
```

Tester les disparités salariales entre hommes et femmes

```
SELECT sexe, avg( salaire )
FROM Employé
GROUP BY sexe
```

Tester les disparités salariales selon le niveau d'éducation

```
SELECT niveau_educatif, avg( salaire )
FROM Employé
GROUP BY niveau_éducatif
```

From:
https://wiki.centrale-med.fr/informatique/ - **WiKi informatique**

Permanent link:
https://wiki.centrale-med.fr/informatique/public:std-3:cm2:interrogation_des_bases_de_donnees:analyse_des_donnees

Last update: **2016/10/03 13:12**