

Requêtes : agrégations

Statistiques, ensemble et sous-ensemble



Sommaire :

- Les fonctions MIN et MAX
 - La fonction Moyenne
 - La fonction Somme
 - La fonction de comptage
 - La fonction de calcul de la variance
 - La fonction écart type
 - La commande GROUP BY
 - La clause HAVING
-
- Mise en pratique

Les fonctions MIN et MAX

Les fonctions **MIN** et **MAX** permet de retrouver les valeurs minimum et maximum d'un champ.

```
select max(prix) as prix_max, min(prix) as prix_min  
from jeux_video;
```

prix_max	prix_min
60	2

Attention :

```
select min(prix) as prix_min, nom, possesseur  
from jeux_video;
```

prix_min	nom	possesseur
2	Super Mario Bros	Florent

```
select max(prix) as prix_max, nom, possesseur  
from jeux_video;
```

prix_max	nom	possesseur
60	Super Mario Bros	Florent

Pas de
correspondance
entre le nom, le
possesseur et le
prix

La fonction Moyenne

- La fonction **MOY** ou **AVG** permet de retrouver la valeur moyenne d'une colonne.

Pour connaitre le prix moyen des jeux vidéo :

```
select avg(prix) as prix_moyen  
from jeux_video;
```

prix_moyen
28.34

La fonction Somme

La fonction **SUM** retour la somme des valeurs d'un champ.

Pour connaitre le prix total de la valeur des jeux vidéo :

```
select sum(prix) as valeur_total  
from jeux_video;
```

valeur_total
1417

La fonction de comptage

La fonction **COUNT** permet de retourner le nombre d'enregistrements d'une requête.

Pour connaître le nombre total de jeux :

```
select count(nom) as nb_jeux  
from jeux_video;
```

nb_jeux
50

La fonction de calcul de la variance

Rappel :

En statistique et en théorie des probabilités, la **variance** est une mesure servant à caractériser la **dispersion** d'un échantillon ou d'une distribution.

Elle indique de quel manière la série statistique ou la variable aléatoire se **disperse autour de ça moyenne**.

- Une variance de 0 signifie que toutes les valeurs sont identiques.
- Une petite variance est signe que toutes les valeurs sont proches les unes des autres.
- Une variance élevée est signe que celles-ci sont très disparates.

La fonction de calcul de la variance

La fonction **VAR** ou **VARIANCE** permet de retourner la variance d'une colonne donnée et d'en déduire **l'amplitude par rapport à la moyenne**.

Pour connaître la variance des prix des jeux vidéos :

```
select variance(prix) as variance  
from jeux_video;
```

variance
317.42440000000005

La fonction écart type

Rappel :

- En probabilité, l'écart type est une mesure de la **dispersion** d'une variable aléatoire.
- En statistique il est une mesure de dispersion de données.
- Il est défini comme la racine carré de la variance.

La fonction **STD** ou **STDDEV** retourne l'écart type d'une série de valeurs et permet d'en déduire leur **dispersion** ou leur **homogénéité**.

La fonction écart type

La fonction **STD** ou **STDDEV** retourne l'écart type d'une série de valeurs et permet d'en déduire leur **dispersion** ou leur **homogénéité**.

Pour connaître l'écart type du prix des jeux vidéo :

```
select variance(prix) as variance  
from jeux_video;
```

ecart_type
17.81640816775368

La commande GROUP BY

Les fonction agrégative peuvent d'appliquer à l'intégralité ou à des sous-ensembles du résultat d'une requête.

Exemple :

Pour connaitre le nombre de jeux par console :

```
select count(nom) as nb_jeux, console  
from jeux_video;
```

nb_jeux	console
50	NES

Le résultat obtenue est **faux**. Il n'y a qu'une seul valeur qui concerne toutes les console et non seulement la 'NES'.

La commande GROUP BY

Pour répondre à cette question il faut ajouter un critère de regroupement. Pour ce faire on utilise la clause **GROUP BY**.

- Elle permet de constituer des **sous-ensembles** d'enregistrements l'or de la requête.
- Chaque attribut cité dans le GROUP BY doit être inclus dans le SELECT.
- Les résultats des clauses WHERE et FROM sont regroupés selon les critères du GROUP BY.

La commande GROUP BY

```
select count(nom) as nb_jeux, console  
from jeux_video  
group by console;
```

Dans le résultat chaque group est réduit à une occurrence et toutes les opérations qui suivent sont appliquées sur cette dernière.

nb_jeux	console
3	NES
3	Megadrive
4	Nintendo 64
2	GameCube
10	Xbox
6	PC
3	SuperNES
9	PS2
2	GBA

La commande GROUP BY

Il est possible de spécifier plusieurs champs dans la commande GROUP BY.

```
select count(nom) as nb_jeux, console, possesseur  
from jeux_video  
group by console, possesseur;
```

nb_jeux	console	possesseur
1	NES	Florent
2	Megadrive	Patrick
2	Nintendo 64	Florent
1	GameCube	Michel
5	Xbox	Patrick

La clause HAVING

La clause **HAVING** est liée à la clause GROUP BY et ne peut être utilisé sans cette dernière. Elle permet de préciser un critère de restriction sur les résultats de l'agrégat.

```
select count(nom) as nb_jeux, possesseur  
from jeux_video  
group by possesseur;
```

```
select count(nom) as nb_jeux, possesseur  
from jeux_video  
group by possesseur  
having nb_jeux > 5;
```

nb_jeux	possesseur
17	Florent
13	Patrick
12	Michel
3	Mathieu
3	Sebastien
2	Corentin

nb_jeux	possesseur
17	Florent
13	Patrick
12	Michel

Mise en pratique :

1. Afficher le nombre de films dans les quels à joué l'acteur « JOHNNY LOLLOBRIGIDA », regroupé par catégorie.
2. Ecrire la requête qui affiche les catégories dans les quels « JOHNNY LOLLOBRIGIDA » totalise plus de 3 films.
3. Afficher la durée moyenne d'emprunt des films par acteurs.
4. L'argent total dépensé au vidéos club par chaque clients, classé par ordre décroissant.