

Données – Tri et filtrage d'une base de données / Classement des villes et sports

I. Importation de la base de donnée.

La base de donnée est donnée brute en fichier csv.

Définition : Un fichier CSV (comma-separated values) désigne un fichier informatique de type texte, dont les valeurs sont séparées par des virgules. C'est en général le format de sortie brute d'une base de donnée exportée.

Par exemple lorsque les données des élèves sont exportées depuis la base de donnée académique, on obtient un fichier avec Nom, prénom, date de naissance, numéro, etc. Chaque donnée est séparée par des virgules.

Activité 1 : Ouvrir le fichier **SPORT_BDD.csv** avec un éditeur de texte type notepad et observer la façon dont sont présentées les données :

1. Que représente la première ligne ?

2. Combien de données a-t-on sur chaque ligne ?

3. Comment sont séparées les différentes données sur chaque ligne ?

Activité 2 : Ouvrir le fichier **SPORT_BDD.csv** avec un logiciel de tableur-grapheur (Excel, Libre office Calc, googlesheet...)

4. Où se trouvent les données présentes sur une ligne ?

L'objectif va être d'obtenir une donnée par cellule. Pour cela nous allons **convertir** les données disponibles.

EXCEL	DONNEES / CONVERTIR
LIBRE OFFICE	DONNEES / TEXTE EN COLONNE

Choisir comme séparateur « VIRGULE » et décocher le reste. On voit un aperçu de ce à quoi va ressembler le fichier une fois converti. Vérifier que chaque donnée sera dans une colonne différente. Chaque colonne est maintenant utilisée pour un descripteur. Sauvegarder le fichier (au format CSV ou XLS ou ODS)

II. Filtrage et tri simple / multicritère

Définition :

- Filtrage : Le filtrage s'effectue à partir de critères qui permettent de n'afficher que les informations y correspondant, en masquant les autres.

Activité 3 : Sélectionner toutes les colonnes de votre document et activer le filtrage :

EXCEL	DONNEES / FILTRER
LIBRE OFFICE	DONNEES / AUTO FILTRE

Des « triangles » apparaissent dans la première ligne qui fera office de choix de descripteur. EN cliquant sur ces triangles on peut maintenant choisir un filtrage particulier :

5. Afficher uniquement les clubs ayant participé au championnat de Basket-Ball.
6. Pour chaque sport, afficher la ville ayant fini première en 2014
7. Afficher les 3 premières villes en 2016 pour le football
8. Afficher les villes ayant fini dans les 10 premiers chaque année en rugby.

III. Tri simple / multiple

Définition :

- Tri : Permet de ranger (par ordre alphabétique, par ordre croissant ...) les éléments d'une liste.

Activité 4 :

9. Trier par ordre alphabétique les noms des villes inscrites en Handball.
10. Trier par ordre de classement les villes inscrites en Handball en 2015. Noter les 4 premières :

11. Trier les villes par classement en 2018, puis 2017, puis 2016, puis 2015 et enfin 2014. Noter les 4 premières.
12. Afficher les villes ayant terminées dans les 10 premiers chaque année en rugby.

IV. Détermination d'un système de classement.

Le but ici est de déterminer quelle est le club le « meilleur » tout sport confondu.

Activité 5 : Dans la colonne I nous allons essayer de créer la donnée sur laquelle nous allons trier les clubs pour déterminer le meilleur.

13. Donner votre définition du « meilleur » club
14. A l'aide d'une formule insérée dans la colonne I, calculer l'indicateur qui vous permettra de classer les clubs.
15. Trier alors les clubs du meilleur au « moins bon ». Indiquer les 4 meilleurs clubs