

La chaos de données peu structurées

Dans cette activité, tu vas manipuler un fichier réel contenant des informations de billetterie. Ton objectif est d'observer les données et de comprendre pourquoi une simple liste ne suffit pas toujours.

Dans cette activité, tu vas manipuler un fichier réel contenant des informations de billetterie. Ton objectif est d'observer les données et de comprendre pourquoi une simple liste ne suffit pas toujours.

Objectifs

À la fin de cette activité, tu seras capable de :

- repérer des problèmes dans un tableau de données
- identifier des répétitions et des incohérences
- comprendre qu'un tableau peut contenir plusieurs types d'informations mélangées
- proposer une organisation plus claire des données

Fichier à utiliser

Ouvre le fichier CSV fourni par ton professeur dans un tableur (Excel, LibreOffice, Google Sheets...).

Partie 1 – Observation des données

Observe attentivement le contenu du fichier.

Réponds aux questions suivantes :

- Combien de fois le client **Marie Dupont** apparaît-il dans le fichier ?
- Les informations de ce client sont-elles toujours identiques (nom, téléphone...) ?
- Combien de fois l'événement **Soirée Quiz Pop Culture** apparaît-il ?
- Les noms des événements sont-ils toujours écrits de la même manière ?
- Les dates sont-elles toutes écrites avec le même format ?
- Les prix sont-ils toujours stockés de la même façon ?
- Trouves-tu des informations manquantes ? Donne un exemple.
- Ce fichier te semble-t-il fiable et facile à utiliser ? Explique pourquoi.

Partie 2 – Manipulation

Effectue les actions suivantes directement dans le fichier ou en expliquant comment tu ferais :

- Retrouver tous les tickets du client **Lucas Martin**
- Retrouver tous les participants du **Quiz Cinéma**

- Retrouver tous les tickets pour des événements qui se passent au Forum de Liège
- Modifier le numéro de téléphone de **Marie Dupont**
- Changer le lieu du **Blind Test Années 90**
- Supprimer un ticket annulé
- Compter le nombre de tickets vendus pour chaque événement

Partie 3 – Analyse

Réponds aux questions suivantes :

- Pourquoi certaines informations apparaissent-elles plusieurs fois ?
- Que se passe-t-il si tu modifies une information à un seul endroit ?
- Quels problèmes peuvent apparaître si plusieurs personnes modifient ce fichier ?
- Est-ce que toutes les informations présentes dans une ligne concernent la même chose ?

Partie 4 – Comprendre les données

Dans le tableau, on retrouve plusieurs types d'informations.

Classe les colonnes dans les trois catégories suivantes :

Informations sur le client

→

Informations sur le ticket

→

Informations sur l'événement

→

Informations sur la salle / le lieux

→

Partie 5 – Proposition d'amélioration

Propose une meilleure organisation des données.

Tu dois :

- regrouper les informations qui vont ensemble
- donner un nom à chaque groupe
- expliquer pourquoi cette organisation est plus claire

Tu peux répondre sous forme de schéma simple ou de liste.



À retenir

Un tableau unique peut devenir difficile à gérer lorsque :

- les mêmes informations sont répétées plusieurs fois
- les données ne sont pas toujours cohérentes
- plusieurs types d'informations sont mélangés

Dans ce cas, il est nécessaire de réfléchir à une meilleure structure des données.
