

# TP : Analyse et Nettoyage de Données avec Tableau Software

## Contexte

Vous travaillez pour un festival artistique international. Trois fichiers contenant des informations sur les artistes, les spectacles et la billetterie vous sont fournis. Cependant, ces fichiers nécessitent un nettoyage avant toute analyse. Votre objectif est de nettoyer, analyser et visualiser les données pour extraire des insights utiles.

## Fichiers fournis

1. **Artists.csv** : Informations sur les artistes.
2. **Shows.csv** : Détails des spectacles prévus.
3. **Tickets.csv** : Historique des ventes de billets.

## Objectifs du TP

1. Identifier et corriger les problèmes dans les données.
2. Réaliser une analyse exploratoire avec Tableau.
3. Créer des visualisations pour répondre aux questions spécifiques du festival.

## Étapes à suivre

### 1. Préparation et exploration des données

1. **Importer les fichiers dans Tableau Software** :
  - Importez les trois fichiers CSV dans Tableau.
  - Identifiez les relations entre les tables à l'aide de jointures ou de relations basées sur les clés suivantes :
    - ArtistID (dans **Artists** et **Shows**)
    - ShowID (dans **Shows** et **Tickets**)
1. **Explorer les données brutes** :
  - Observez les données dans Tableau :
  - Y a-t-il des valeurs manquantes ?
  - Voyez-vous des doublons ?
  - Les types de données sont-ils corrects (dates, nombres, etc.) ?

### 2. Nettoyage des données

1. **Valeurs manquantes** :
  - Utilisez des filtres ou remplacez les valeurs manquantes pour :
    - Les noms des artistes (remplacez par "Inconnu" si manquant).
    - Les dates des spectacles (investiguez les spectacles associés).
1. **Doublons** :
  - Identifiez et supprimez les doublons dans les tables **Artists** et **Shows**.
1. **Erreurs logiques** :

- Vérifiez la cohérence des valeurs :
  - Les âges des artistes doivent être raisonnables (entre 18 et 70 ans).
  - Les prix des billets doivent être  $> 0$  €.
1. **Nettoyage des formats :**
    - Uniformisez les noms de pays (par ex., "USA" et "United States" doivent être identiques).

### 3. *Analyse exploratoire*

1. **Créer des visualisations :**
  - **Répartition des artistes :**
    - Par pays.
    - Par genre artistique.
  - **Analyse des spectacles :**
    - Quels types de spectacles attirent le plus de participants ?
    - Quel est le revenu total par spectacle ?
  - **Analyse des ventes de billets :**
    - Moyenne des prix des billets par type de spectacle.
    - Nombre de billets vendus par méthode de paiement.
1. **Identifier les tendances :**
  - Existe-t-il une corrélation entre le genre de spectacle et le prix moyen des billets ?
  - Quels artistes participent à plusieurs spectacles ?

### 4. *Réponses aux questions spécifiques*

Répondez aux questions suivantes en créant des visualisations pertinentes :

1. Quel pays a fourni le plus grand nombre d'artistes ?
2. Quel type de spectacle génère le revenu moyen le plus élevé ?
3. Quelle méthode de paiement est la plus utilisée ?

### 5. *Présentation*

1. Préparez un tableau de bord interactif avec Tableau :
  - Intégrez vos visualisations dans un tableau de bord clair et organisé.
  - Ajoutez des filtres interactifs pour explorer les données par pays, artiste ou spectacle.

## Livrables

1. Tableau de bord interactif avec :
  - Au moins **5 visualisations**.
  - Une organisation claire avec des titres et des explications.
1. Un rapport décrivant :
  - Les problèmes identifiés dans les données et comment ils ont été corrigés.
  - Les principaux insights tirés des visualisations.

## **Bonus**

- Ajoutez une visualisation temporelle pour montrer l'évolution des ventes de billets au fil du temps.
- Comparez les revenus générés par spectacle dans différentes villes.