

## B - Analyse de la conversion en DataFrame

Pour chaque type de fichier mentionné, nous analysons la possibilité de le convertir en un DataFrame , en expliquant les raisons et en proposant des alternatives si nécessaire.

### 1. Fichier Audio ( librosa)

- Conversion en DataFrame ? Non
- Justification : Un fichier audio contient une onde sonore sous forme d'un signal numérique. Ce type de données n'est pas tabulaire.
- Alternative : Il est possible d'extraire certaines caractéristiques ( durée , fréquence , intensité) et de les stocker dans un DataFrame sous forme de colonnes.

### 2. Fichier Vidéo ( OpenCV)

- Conversion en DataFrame ? Non
- Justification : Une vidéo est une séquence d'images avec des métadonnées associées ( FPS , résolution , durée ) . Format CEUne vidéo est une séquence d'images avec des métadonnées associées (FPS, résolution, durée). Ce format n'est pas directement tabulaire.
- Alternative : On peut extraire les métadonnées (FPS, durée, dimensions) et les stocker sous forme d'un DataFrame.

### 3. Fichier Parquet ( pandas)

- Conversion en DataFrame ? Oui
- Justification : Le format Parquet est conçu pour stocker des données tabulaires de manière optimisée. Il est parfaitement compatible avec un DataFrame.

### 4. Fichier Avro ( fastavro)

- Conversion en DataFrame ? Oui
- Justification : Le format Avro stocke des données structurées sous forme d'objets, ce qui le rend facilement convertible en DataFrame.

### 5. Fichier YAML ( yaml)

- Conversion en DataFrame ? Parfois
- Justification : YAML est un format de stockage hiérarchique (clé-valeur). Si les données sont structurées sous forme de liste d'objets, elles peuvent être transformées en DataFrame.
- Alternative : Extraire uniquement les éléments clés sous forme d'un dictionnaire et le convertir en DataFrame.

### 6. Fichier ZIP ( zipfile)

- Conversion en DataFrame ? Non

- Justification : Un fichier ZIP est une archive contenant plusieurs fichiers compressés, ce qui n'a pas de structure tabulaire.
- Alternative : Extraire la liste des fichiers contenus dans l'archive avec leurs tailles et dates de modification, puis créer un DataFrame avec ces informations.

## 7. Fichier PDF ( PyPDF2)

- Conversion en DataFrame ? Non
- Justification : Un fichier PDF contient généralement du texte et des images, sans structure tabulaire explicite.
- Alternative : Si le PDF contient des tableaux, on peut utiliser `pdfplumber` pour les extraire et les convertir en DataFrame.

## 8. Fichier Markdown ( .md)

- Conversion en DataFrame ? Non
- Justification : Le format Markdown est utilisé pour la mise en forme de texte avec des balises. Il ne convient pas à une structure tabulaire.
- Alternative : Convertir le fichier Markdown en HTML, puis extraire des tableaux s'ils existent avec `BeautifulSoup`.

## 9. Fichier TXT ( .txt)

- Conversion en DataFrame ? Non
- Justification : Un fichier TXT est un fichier de texte brut sans structure spécifique.
- Alternative : Si le fichier suit un format spécifique (exemple : valeurs séparées par des tabulations), il peut être transformé en DataFrame