

ÉTUDE DE CAS – BLOC 1

Analyse du besoin et structuration des données pour une solution d'intelligence artificielle

Durée : 7h00

1. Contexte

La société **CulturaPass** développe une plateforme numérique permettant aux collectivités de proposer des **activités culturelles et sportives** (ateliers, spectacles, stages) à destination des habitants.

La plateforme est utilisée par plusieurs villes de tailles différentes, chacune disposant de ses propres règles de gestion, calendriers et partenaires.

Avec l'augmentation du nombre d'utilisateurs et d'événements proposés, CulturaPass rencontre des difficultés à :

- analyser les taux de participation,
- anticiper les pics de fréquentation,
- identifier les événements sous-utilisés ou surchargés.

Les données sont aujourd'hui stockées dans plusieurs fichiers CSV produits par différents partenaires (maisons de quartier, associations, services municipaux), sans structure harmonisée.

L'entreprise souhaite préparer une **base de données fiable** permettant, à terme, de développer des outils d'aide à la décision basés sur l'intelligence artificielle.

2. Mission confiée

Vous intervenez en tant que chargé de structuration des données. Votre rôle consiste à **clarifier les attentes métier, organiser les données existantes et mettre en place un environnement exploitable pour des analyses avancées.**

Vous devrez notamment :

- Reformuler les objectifs de CulturaPass en besoins fonctionnels et data.
- Identifier les jeux de données disponibles, leur origine et leurs caractéristiques.
- Concevoir une organisation de données cohérente à partir de fichiers CSV multiples.
- Mettre en place une base de données et définir les règles d'importation des fichiers.
- Analyser la qualité des données et proposer des traitements adaptés aux valeurs manquantes ou incohérentes.
- Harmoniser les formats (dates, catégories, identifiants).
- Structurer les données pour faciliter les analyses statistiques et prédictives futures.
- Tester l'efficacité de l'organisation retenue à l'aide de requêtes simples.

- Identifier les risques liés à l'exploitation des données utilisateurs.

3. Données mises à disposition (exemples)

a) Événements proposés – evenements.csv

event_id,city,category,start_date,end_date,max_capacity

E1001,Lille,Théâtre,2025-04-12,2025-04-12,120

E1002,Roubaix,Sport,2025-04-13,2025-04-13,50

E1003,Lille,Musique,2025-04-14,2025-04-15,

E1004,Tourcoing,Arts plastiques,2025-04-15,2025-04-15,30

b) Inscriptions – inscriptions.csv

registration_id,event_id,user_age,registration_date,attendance_status

R5001,E1001,34,2025-04-01,Confirmé

R5002,E1001,,2025-04-02,Confirmé

R5003,E1002,17,2025-04-05,Annulé

R5004,E1003,45,,Confirmé

c) Partenaires organisateurs – organisateurs.csv

organizer_id,city,type,years_active

O201,Lille,Association,12

O202,Roubaix,Service municipal,

O203,Tourcoing,Association,5

Les fichiers peuvent contenir :

- des champs vides,
- des incohérences de libellés,
- des doublons partiels,
- des formats hétérogènes selon les partenaires.

4. Livrables attendus

Vous rédigerez un **rapport de structuration et d'analyse des données** (10 pages maximum) comprenant :

- Une reformulation synthétique du besoin métier
- Une cartographie des fichiers CSV et de leurs relations
- Les principes d'organisation et de stockage retenus
- Les règles de nettoyage et d'harmonisation appliquées
- Une proposition de modèle de données
- Des exemples d'analyses rendues possibles par la structuration
- Une évaluation des performances de requêtes simples
- Une analyse des enjeux réglementaires et éthiques
- Un retour critique sur les limites des données disponibles