



27/01/2026

Data Gouvernance

Présentation de la base de données



Yousra Bouhanna, Abdeljebbar Abid, Elias Ait Hassou
MASTER 2 BUSINESS INTELLIGENCE & ANALYTICS – LYON2

Présentation de la base de données

Dans le cadre de ce projet, nous avons choisi de travailler sur une base de données hospitalière. La gestion d'un hôpital est sujet important, une bonne gestion d'un hôpital passe également par une bonne gestion de ses données. La fiabilité et la cohérence des données sont essentielles, c'est pourquoi nous avons choisi d'appliquer la Data Gouvernance à ce sujet.

La base de données étudiée offre une vision globale de la gestion hospitalière et contient cinq tables.

La table **patients** contient les informations relatives aux patients, telles que le nom, l'âge, l'adresse et les coordonnées de contact. Elle comprend un total de 50 000 lignes et 14 colonnes.

La table **staff** regroupe les informations concernant le personnel hospitalier, incluant le personnel médical et administratif. Elle contient notamment les informations d'identification, l'âge, l'adresse, les coordonnées de contact ainsi que le service d'affectation. Cette table est composée de 50 000 lignes et 12 colonnes.

La table **consultations** recense l'ensemble des rendez-vous réalisés entre les patients et le personnel médical. Elle contient les identifiants du patient et du membre du personnel, ainsi que la date et l'heure de la consultation. Cette table comporte 50 000 lignes et 5 colonnes.

La table **staff_schedule** décrit le planning du personnel médical. Pour chaque semaine de l'année, elle indique si un membre du personnel est présent ou absent. Cette table est la plus volumineuse de la base, avec environ 2 millions de lignes et 6 colonnes.

Enfin, la table **services_weekly** regroupe des informations hebdomadaires par service hospitalier, notamment la gestion des lits disponibles ainsi que des indicateurs de satisfaction des patients. Cette table contient 209 lignes et 10 colonnes.

Source des données et enrichissement par anomalies

Les données utilisées dans ce projet proviennent de la plateforme Kaggle, une plateforme reconnue pour la mise à disposition de jeux de données à des fins pédagogiques et analytiques. Lien du dataset :

<https://www.kaggle.com/datasets/jaderz/hospital-beds-management>

Afin de rendre l'étude plus réaliste et adaptée aux problématiques de data gouvernance, des anomalies ont volontairement été introduites dans la base de données à l'aide de techniques basées sur l'intelligence artificielle.

Ces anomalies concernent notamment des incohérences de formats, des valeurs manquantes, des doublons et des violations de règles métier. Cette démarche permet de simuler des situations proches de la réalité et d'évaluer l'efficacité des mécanismes de détection et de gouvernance des données mis en place dans le cadre du projet.