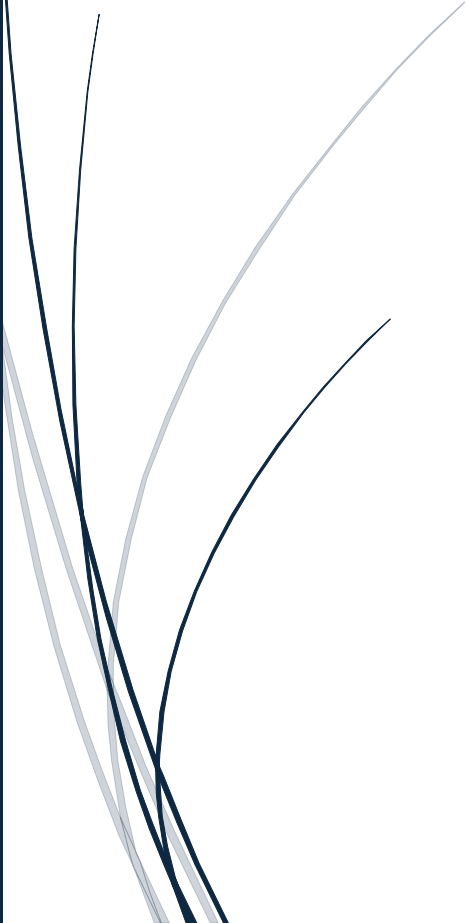


27/01/2026

# Data Gouvernance

Analyse de la qualité des données et  
plan de redressement priorisé



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

## Table des matières

1. Introduction .....	2
2. Analyse de la qualité de la base de données hospitalière .....	3
2.1 Analyse de la table Patients .....	3
2.1.1 Vue d'ensemble de la table .....	3
2.1.2 Règles de qualité des données appliquées .....	4
2.1.3 Analyse globale de la qualité des données .....	4
2.1.4 Anomalies détectées et impact métier .....	5
2.2 Analyse de la table Consultations .....	5
2.2.1 Vue d'ensemble de la table .....	5
2.2.2 Règles de qualité des données appliquées .....	6
2.2.3 Analyse globale de la qualité des données .....	6
2.2.4 Anomalies détectées et impact métier .....	7
2.3 Analyse de la table Staff .....	7
2.3.1 Vue d'ensemble de la table .....	7
2.3.2 Règles de qualité des données appliquées .....	7
2.3.3 Analyse globale de la qualité des données .....	8
2.3.4 Anomalies détectées et impact métier .....	9
2.4 Analyse de table Staff Schedule .....	10
2.4.1 Vue d'ensemble de la table .....	10
2.4.2 Règles de qualité des données appliquées .....	10
2.4.3 Analyse globale de la qualité des données .....	10
2.4.4 Anomalies détectées et impact métier .....	10
3. Plan de redressement priorisé .....	11

# 1. Introduction

L'objectif de cette phase est d'analyser la qualité des données de la base hospitalière, d'identifier les anomalies dans les tables patients, staff, consultations et staff schedule, et de proposer des actions correctives pour garantir l'intégrité et la fiabilité des données.

Périmètre :

- Table patients : informations personnelles, coordonnées
- Table staff : informations sur le personnel
- Table consultations : rendez-vous patients / staff
- Table staff\_schedule : planning du personnel

Méthodologie :

- Analyse des formats de données
- Définition des règles de validation des données
- Vérification de la cohérence inter-colonnes
- Détection des doublons et valeurs manquantes
- Analyse des anomalies métier

## 2. Analyse de la qualité de la base de données hospitalière

Dans cette section, nous analysons la qualité des données présentes dans la base hospitalière afin d'identifier les anomalies, les incohérences et les valeurs manquantes ainsi que leur impact métier.

### 2.1 Analyse de la table Patients

#### 2.1.1 Vue d'ensemble de la table

Overview

Alerts 14

Reproduction

Dataset statistics

Number of variables	14
Number of observations	50000
Missing cells	26699
Missing cells (%)	3.8%
Duplicate rows	0
Duplicate rows (%)	0.0%

Variable types

Text	6
Numeric	3
DateTime	3
Categorical	2

Alerts

name has 1516 (3.0%) missing values

Missing

age has 1474 (2.9%) missing values

Missing

arrival\_date has 1528 (3.1%) missing values

Missing

departure\_date has 1467 (2.9%) missing values

Missing

service has 1483 (3.0%) missing values

Missing

satisfaction has 1499 (3.0%) missing values

Missing

adresse\_postale has 1448 (2.9%) missing values

Missing

code\_postal has 1523 (3.0%) missing values

Missing

genre has 1446 (2.9%) missing values

Missing

date\_naissance has 1523 (3.0%) missing values

Missing

telephone has 5096 (10.2%) missing values

Missing

email has 5210 (10.4%) missing values

Missing

numero\_dossier\_medical has 1486 (3.0%) missing values

Missing

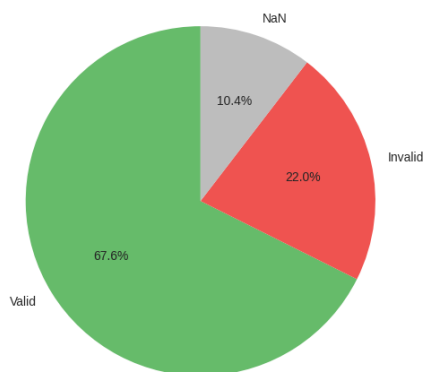
## 2.1.2 Règles de qualité des données appliquées

Catégorie	Champ(s) concerné(s)	Description de la règle
Format	email	L'adresse email doit respecter un format standard, incluant un caractère @ et un nom de domaine valide.
Format	phone	Le numéro de téléphone doit être composé de 10 chiffres et respecter le format français.
Format	postal_code	Le code postal doit contenir exactement 5 caractères numériques, conformément au format français.
Cohérence inter-colonnes	arrival_date, departure_date	La date de départ doit être strictement postérieure à la date d'arrivée.
Cohérence inter-colonnes	birth_date, age	L'âge déclaré doit être cohérent avec la date de naissance du patient.

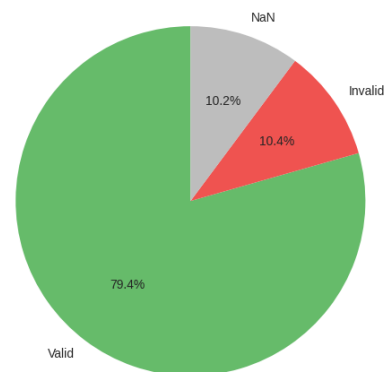
## 2.1.3 Analyse globale de la qualité des données

Vue d'ensemble de la qualité des données des patients

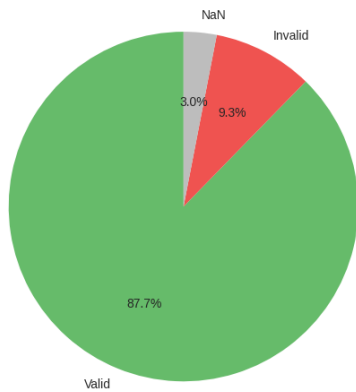
Statut de validation des e-mails des patients



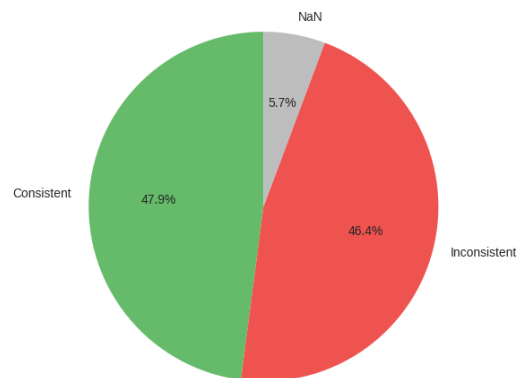
Statut de validation des numéros de téléphone des patients



Statut de validation des codes postaux des patients



Statut de cohérence âge-date de naissance des patients



### 2.1.4 Anomalies détectées et impact métier

Anomalie	Impact métier
Email invalide	Communication impossible
Téléphone invalide	Suivi patient incomplet
departure_date < arrival_date	Planning incohérent, facturation
Âge incohérent avec birth_date	Risque prescription incorrecte

## 2.2 Analyse de la table Consultations

### 2.2.1 Vue d'ensemble de la table

#### Dataset statistics

Number of variables	5
Number of observations	50000
Missing cells	7492
Missing cells (%)	3.0%
Duplicate rows	0
Duplicate rows (%)	0.0%

#### Variable types

Text	3
DateTime	1
Categorical	1

#### Alerts

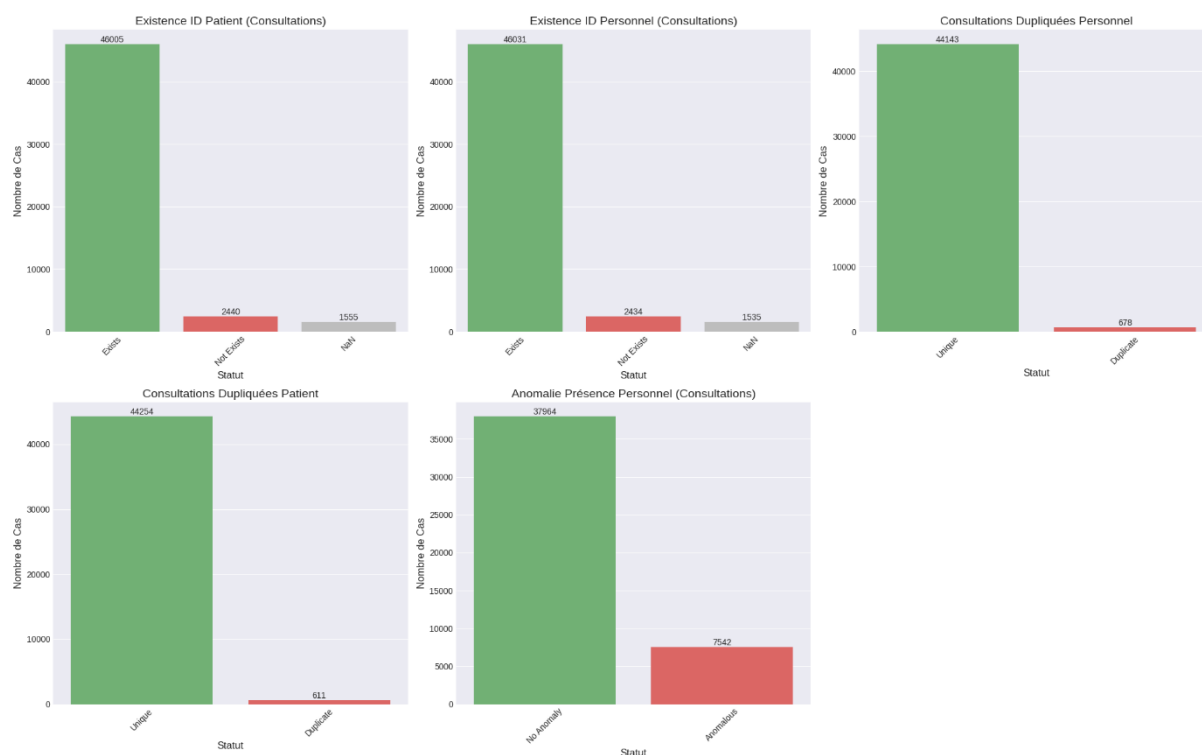
patient_id has 1555 (3.1%) missing values	Missing
staff_id has 1535 (3.1%) missing values	Missing
consultation_date has 1500 (3.0%) missing values	Missing
consultation_time has 1459 (2.9%) missing values	Missing
description has 1443 (2.9%) missing values	Missing

## 2.2.2 Règles de qualité des données appliquées

Catégorie	Champ(s) concerné(s)	Description de la règle
Intégrité référentielle	patient_id	Chaque patient_id présent dans la table consultations doit exister dans la table patients.
Intégrité référentielle	staff_id	Chaque staff_id présent dans la table consultations doit exister dans la table staff.
Unicité	staff_id, consultation_date, consultation_time	Un même membre du personnel ne peut pas avoir deux consultations enregistrées à la même date et à la même heure.
Unicité	patient_id, consultation_date, consultation_time	Un même patient ne peut pas avoir deux consultations enregistrées à la même date et à la même heure.
Cohérence métier	staff_id, consultation_date	Une consultation ne doit pas être enregistrée si le membre du personnel est indiqué comme absent (present = 0) dans la table staff_schedule à la date concernée.

## 2.2.3 Analyse globale de la qualité des données

Vue d'ensemble de la qualité des données : Consultations et Présence du Personnel



## 2.2.4 Anomalies détectées et impact métier

Anomalie	Impact métier
patient_id inexistant	Consultation non rattachée
staff_id inexistant	Impossible de valider consultation
Chevauchements de consultation incohérents	Surbooking, double facturation
Staff absent selon staff_schedule	Fraude possible, incohérence planning
Valeurs manquantes	Planning incomplet

## 2.3 Analyse de la table Staff

### 2.3.1 Vue d'ensemble de la table

Dataset statistics		Variable types	
Number of variables	12	Text	5
Number of observations	50000	Categorical	4
Missing cells	27563	Numeric	2
Missing cells (%)	4.6%	DateTime	1
Duplicate rows	0		
Duplicate rows (%)	0.0%		

#### Alerts

staff_name has 1789 (3.6%) missing values	Missing
role has 1923 (3.8%) missing values	Missing
service has 1846 (3.7%) missing values	Missing
adresse_postale has 1779 (3.6%) missing values	Missing
code_postal has 1812 (3.6%) missing values	Missing
genre has 1754 (3.5%) missing values	Missing
date_naissance has 1871 (3.7%) missing values	Missing
age has 1846 (3.7%) missing values	Missing
telephone has 5575 (11.2%) missing values	Missing
email has 5493 (11.0%) missing values	Missing
departement has 1875 (3.8%) missing values	Missing

### 2.3.2 Règles de qualité des données appliquées

Catégorie	Champs concernés	Règle de qualité
Identité	staff_name, genre, date_naissance, age	Complétude des informations d'identité
Organisation	role, service, departement	Complétude des informations organisationnelles



Coordonnées	adresse_postale, code_postal	Complétude des informations de localisation
Contact	telephone, email	Complétude des informations de contact

### 2.3.3 Analyse globale de la qualité des données

Vue d'ensemble de la qualité des données du personnel



### 2.3.4 Anomalies détectées et impact métier

Catégorie	Champs concernés	Anomalie détectée	Impact métier
Identité	staff_name, genre, date_naissance, age	Informations d'identité manquantes	Difficulté d'identification du personnel, risques d'erreurs administratives et limitations dans le suivi RH.
Organisation	role, service, département	Informations organisationnelles manquantes	Mauvaise affectation des ressources, erreurs de planification et difficultés dans l'analyse organisationnelle.
Localisation	adresse_postale, code_postal	Informations de localisation manquantes	Problèmes administratifs, difficultés de gestion des dossiers du personnel.
Contact	telephone, email	Informations de contact manquantes	Impossibilité de contacter rapidement le personnel, notamment en cas d'urgence ou de modification de planning.

## 2.4 Analyse de table Staff Schedule

### 2.4.1 Vue d'ensemble de la table

#### Dataset statistics

Number of variables	6
Number of observations	6552
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%

#### Variable types

Numeric	1
Text	2
Categorical	3

### 2.4.2 Règles de qualité des données appliquées

Catégorie	Champ concerné	Règle de qualité	Description de la règle
Intégrité référentielle	staff_id	Existence du staff	Chaque staff_id doit exister dans la table staff afin de garantir la cohérence entre les plannings et le personnel enregistré.
Complétude	date / période	Champ obligatoire	La date ou la période de planning doit être renseignée pour permettre une planification correcte des ressources.
Validité	present	Valeurs autorisées	Le champ present doit contenir une valeur valide : 1 si le membre du personnel est présent, 0 s'il est absent.

### 2.4.3 Analyse globale de la qualité des données

Vue d'ensemble de la qualité des données du planning du personnel



### 2.4.4 Anomalies détectées et impact métier

Cette table ne contient pas d'anomalie particulière mais sera importante à lier avec les autres tables pour l'impact métier si jamais certains médecins ont des rendez-vous sur des semaines où ils sont absents.

### 3. Plan de redressement priorisé

Anomalie	Action corrective	Priorité
patient_id inexistant	Vérifier patient → corriger ou supprimer ligne	Critique
staff_id inexistant	Vérifier staff → corriger ou supprimer ligne	Critique
Staff absent lors d'une consultation	Vérification manuelle / audit	Critique
Doublon consultation	Fusion ou suppression doublons	Haute
departure_date < arrival_date	Corriger date ou alerter planning	Haute
Âge incohérent	Vérifier date de naissance ou âge	Haute
Email/téléphone invalide	Standardisation ou contact patient	Moyenne
Valeurs manquantes	Compléter ou signaler pour saisie future	Haute