



BNDMR

Banque Nationale de Données
Maladies Rares

Banque Nationale de Données Maladies Rares (BNDMR)

***Test technique pour un stage en
Data Science à la BNDMR***

1. Introduction:

Ce test pour la position d'un stage de data science à la BNDMR est à faire chez soi dans le temps imparti. Il n'y a pas de contraintes sur l'utilisation de langages ou de méthodes. Les données sont fictives.

Le travail de data science à la BNDMR nécessite des compétences en SQL et la maîtrise d'au moins d'un langage de traitement de données comme Python ou R.

2. Contexte

Une association de patients souhaite obtenir un rapport descriptif sur les patients souffrant d'Epidermolyse Bulleuse Héréditaire (EBH), une maladie dermatologique rare.

D'après la nomenclature Orphanet des maladies rares, il y a plusieurs groupes d'EBH dont les EBHs Simples et les EBHs Jonctionnelles (voir le diagramme en information complémentaire).

La population d'intérêt concerne tous les patients avec les codes orpha 257, 304, 305, 79403, 158684, 251393 et avec au moins une activité entre 2017 et 2021.

Données

Pour cet exercice, deux fichiers de données sont fournis sous format .csv:

- Fichier patients: données démographiques des patients
- Fichier diagnostics: données de diagnostic des patients

3. Objectifs

- 1) En vous basant sur le schéma de la base de données ci-dessous, écrire la requête sql qui permet d'obtenir les 2 fichiers .csv en sortie.
Pour information, la date_derniere_activité du fichier diagnostic est la date d'activité la plus récente connue pour chaque patient.
- 2) Etablir un rapport (.pdf si possible) décrivant les patients par sexe, classe d'âge (<18 et > 18 ans), ville de prise en charge, durée de suivi ainsi que la distribution des cas par groupe d'EBHs (EBH simple vs EBH Jonctionnelle).

4. Informations complémentaires

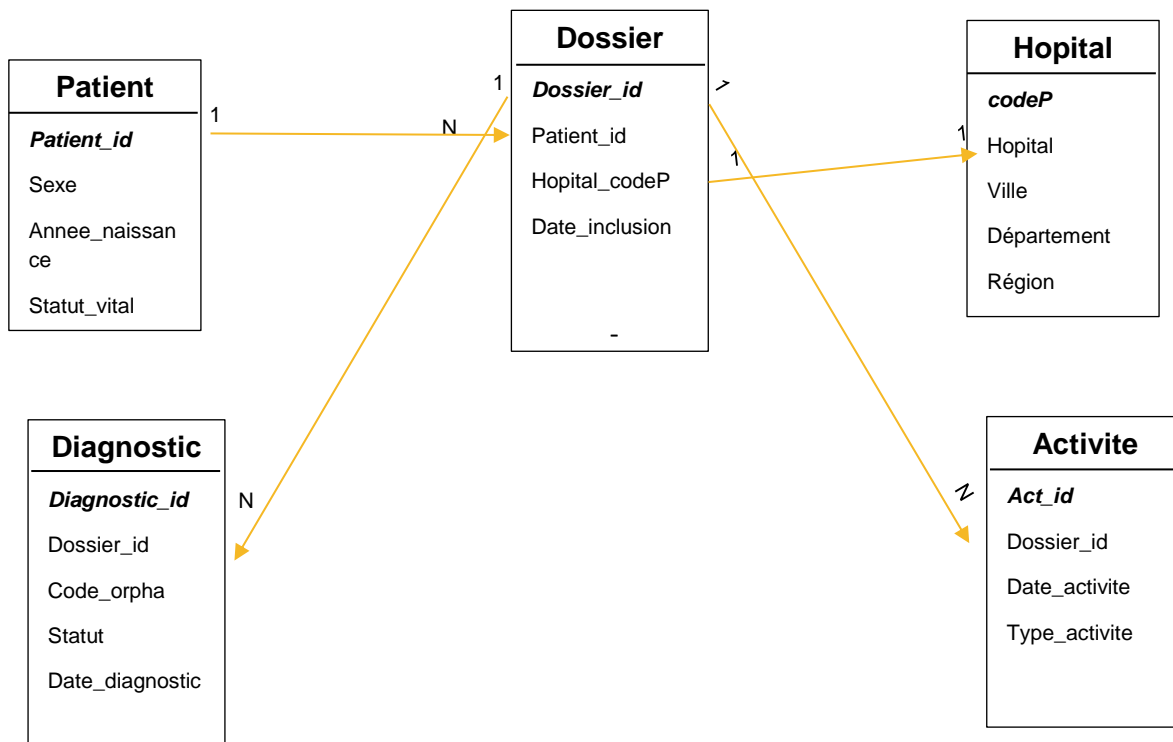
1- Schéma de la base de données

Les 2 fichiers extraits sont issus d'un entrepôt de données « BD_MR_2024 » sous le système de gestion de base de données PostgreSQL.

Nom du serveur de base de données: BNDMR_BD_host

Votre username : stagiaire

Votre motdepasse : jesuisstagiaire?2024



2- Relations entre les codes diagnostics

La lecture du nœud parent vers le nœud fils signifie : 'inclut' ou 'contient' et de fils vers le parent se lit : 'est' 'fait partie'

