

# Analyse Détaillée du Dataset de Diabète

## 1. Structure du Dataset

Colonne	Description
id_patient	Identifiant unique pour chaque patient.
Age	Âge du patient en années.
Pregnancies	Nombre de grossesses (pour les femmes).
BMI	Indice de masse corporelle, calculé à partir du poids et de la taille.
Glucose	Taux de glucose dans le sang (mg/dL).
BloodPressure	Pression artérielle (mmHg).
HbA1c	Taux d'HbA1c, qui donne une indication du contrôle glycémique sur le long terme.
LDL	Cholestérol LDL (lipoprotéines de basse densité).
HDL	Cholestérol HDL (lipoprotéines de haute densité).
poids	Poids du patient (en kg).
Triglycerides	Taux de triglycérides dans le sang (mg/dL).
sexe	Sexe du patient (Homme ou Femme).
WaistCircumference	Circonférence de la taille (en cm).
HipCircumference	Circonférence des hanches (en cm).
WHR	Rapport taille-hanches (Waist-Hip Ratio).
FamilyHistory	Antécédents familiaux de diabète (0 = Non, 1 = Oui).
DietType	Type de régime alimentaire (0, 1, ou 2 selon la classification).
Hypertension	Antécédents d'hypertension (0 = Non, 1 = Oui).
MedicationUse	Utilisation de médicaments (0 = Non, 1 = Oui).
Outcome	État de diabète du patient (0 = Pas de diabète, 1 = Diabète).

## 2. Statistiques Descriptives

Pour mieux comprendre le dataset, voici quelques statistiques descriptives pour les colonnes numériques :

- **Nombre total d'entrées** : 1123
- **Âge** :
  - Moyenne : 53.5 ans
  - Minimum : 18 ans
  - Maximum : 89 ans
- **Nombre de grossesses** :
  - Moyenne : 6.75
  - Minimum : 0
  - Maximum : 16
- **BMI** :
  - Moyenne : 27.1
  - Minimum : 15.0
  - Maximum : 41.8
- **Glucose** :
  - Moyenne : 109.1 mg/dL
  - Minimum : 50.0 mg/dL
  - Maximum : 207.2 mg/dL
- **Pression artérielle** :
  - Moyenne : 85.2 mmHg
  - Minimum : 60.0 mmHg
  - Maximum : 138.0 mmHg
- **HbA1c** :
  - Moyenne : 4.9
  - Minimum : 4.0
  - Maximum : 6.9
- **LDL** :
  - Moyenne : 100.4 mg/dL
  - Minimum : -12.0 mg/dL (valeurs aberrantes possibles)
  - Maximum : 202.2 mg/dL
- **HDL** :
  - Moyenne : 49.4 mg/dL
  - Minimum : -9.2 mg/dL (valeurs aberrantes possibles)
  - Maximum : 107.8 mg/dL
- **Triglycérides** :

- Moyenne : 145.2 mg/dL
- Minimum : 50.0 mg/dL
- Maximum : 345.8 mg/dL

### 3. Analyse des Variables Catégorielles

- **Sexe** : Répartition entre hommes et femmes.
- **Antécédents familiaux** : Pourcentage de patients ayant des antécédents familiaux de diabète.
- **Hypertension** : Proportion de patients ayant des antécédents d'hypertension.
- **Utilisation de médicaments** : Proportion de patients utilisant des médicaments.

### 4. Corrélations

Il est pertinent d'explorer les corrélations entre les différentes variables, notamment entre les niveaux de glucose, l'IMC et le résultat du diabète. Cela peut donner des indications sur les facteurs de risque associés au diabète.

### 5. Valeurs Manquantes et Doublons

Il est important de vérifier la présence de valeurs manquantes ou de doublons dans le dataset, car cela peut affecter les analyses et les modèles prédictifs.

### 6. Conclusion

Ce dataset contient des informations précieuses pour l'analyse et la modélisation des risques de diabète. L'exploration et le prétraitement des données sont des étapes cruciales avant d'appliquer des techniques d'apprentissage automatique ou d'effectuer des analyses statistiques.

Si vous avez besoin d'une analyse plus approfondie ou d'une visualisation spécifique, n'hésitez pas à demander !