

Documentation

Jeu de données sur le diabète chez des femmes

Cet ensemble de données fournit un aperçu complet des différents types de diabète, y compris le diabète induit par les stéroïdes, le diabète néonatal (NDM), le prédiabète, le diabète de type 1 et le syndrome de Wolfram. Il comprend un large éventail d'attributs médicaux, génétiques et de mode de vie qui sont importants pour comprendre les facteurs contribuant au diabète.

Source: <https://www.kaggle.com/datasets/ankitbatra1210/diabetes-dataset>

Nombre d'attributs:34

Nombre d'enregistrements: 70 000

Ce document regroupe les éléments essentiels pour l'analyse et la visualisation des données liées au diabète. Vous y trouverez d'abord un dictionnaire de données décrivant les attributs de notre jeu de données, suivi de la requête SQL nécessaire à la création de la table diabetes dans une base de données PostgreSQL. Ensuite, un ensemble de requêtes est proposé pour générer des graphiques sur Grafana. Pour finir, une liste des sources utilisées dans l'élaboration du rapport est fournie pour compléter ce travail.

Sommaire

1. Dictionnaire de données	2
2. Requête de création de la table diabetes (SQL dans une BD Postgresql)	5
3. Ensemble de requêtes pour la génération des graphiques sur Grafana	6
a. Nombre total de cas en fonction du type de diabète	6
b. Nombre de cas liés à antécédents familiaux connus de diabète par rapport au nombre total de cas	6
c. Niveau maximal de taux de glucose chez un patient	6
d. Niveau minimal de taux de glucose chez un patient	7
e. Niveau maximal d'insuline chez une patient	7
f. Niveau minimal d'insuline chez une patient	7
g. Type de diabète en fonction du groupe d'âge	7
h. Statut tabagique des patientes	8
i. Habitudes alimentaires des patientes	8
j. Proportion de femmes qui consomment de l'alcool tout en étant fumeuses	9
k. Niveau d'activité physique des patientes	9
l. Types de diabète subis par les femmes qui consomment de l'alcool tout en étant fumeuses	9
m. Types de diabète subis par les femmes fumeuses	9
n. Types de diabète subis par les femmes qui consomment de l'alcool (High & Moderate)	10
4. Sources (Rapport)	10

1. Dictionnaire de données

	Nom de l'attribut	Libellé	Type
1	Target	Le type de diabète ou de prédiabète	Chaîne de caractères
2	Genetic Markers	Indicateurs de prédisposition génétique au diabète	Booléen (positif/négatif)
3	Autoantibodies	Présence d'auto-anticorps communément associés au diabète auto-immun	Booléen (positif/négatif)
4	Family History	Informations sur l'existence d'antécédents familiaux connus de diabète	Booléen (Oui/Non)
5	Environmental Factors	Détails sur les influences environnementales qui peuvent contribuer au diabète	Booléen (Présent/Absent)
6	Insulin Levels	Taux d'insuline mesurés chez les patients	Entier
7	Age	Age du patient	Entier
8	BMI	Indice de masse corporelle du patient	Entier
9	Physical Activity	Niveau de l'activité physique du patient	Liste de valeurs (faible, moyen, élevé)
10	Dietary Habits	Habitudes alimentaires du patient	Booléen (Sain/Malsain)
11	Blood Pressure	Pression du sang	Entier

		dans les artères (Pression artérielle)	
12	Cholesterol Levels	Taux de cholestérol du patient	Entier
13	Waist Circumference	Tour de taille du patient	Entier
14	Blood Glucose Levels	Taux de glycémie du patient	Entier
15	Ethnicity	Impact de l'ethnicité du patient sur le risque d'apparition du diabète chez le patient	Booléen (Risque faible/élevé)
16	Socioeconomic Factors	Impact des facteurs socio-économiques sur le risque d'apparition du diabète chez le patient	Liste de valeurs (Faible, moyen, élevé)
17	Smoking Status	Statut tabagique	Booléen (Fumeur/Non fumeur)
18	Alcohol Consumption	Consommation d'alcool	Liste de valeurs (Faible, moyen, élevé)
19	Glucose Tolerance Test	Test de tolérance au glucose	Booléen (Normal/Anormal)
20	History of PCOS	Antécédents de SOPK (syndrome des ovaires polykystiques)	Booléen (Oui/Non)
21	Previous Gestational Diabetes	Antécédents de diabète gestationnel	Booléen (Oui/Non)
22	Pregnancy History	Antécédents de grossesse de la mère	Booléen (Normal/Complicqué)
23	Weight Gain During Pregnancy	Gain de poids pendant la grossesse	Entier
24	Pancreatic Health	Santé pancréatique	Entier

		(Chez l'adulte, la valeur normale du taux d'amylase se situe de 10 à 45 UI / L (à 30°C) et de 12 à 62 UI / L (à 37°C) dans le sang)	
25	Pulmonary Function	Fonction pulmonaire (Mesure de la quantité d'air que les poumons peuvent contenir et celle de la capacité à expulser l'air des poumons)	Entier
26	Cystic Fibrosis Diagnosis	Diagnostic de mucoviscidose	Booléen (Oui/Non)
27	Steroid Use History	Antécédents d'utilisation de stéroïdes	Booléen (Oui/Non)
28	Genetic Testing	Dépistage génétique des gènes liés au diabète	Booléen (Positif/Négatif)
29	Neurological Assessments	Évaluations neurologiques	Liste de valeurs (1, 2, 3)
30	Liver Function Tests	Tests de la fonction hépatique	Booléen (Normal/Anormal)
31	Digestive Enzyme Levels	Taux d'enzymes digestives	Entier
32	Urine Test	Test urinaire	Liste de valeurs (Normal, présence de glucose, présence de cétone, présence de protéine)
33	Birth Weight	Poids de naissance	Entier
34	Early Onset Symptoms	Symptômes précoces	Booléen (Oui/Non)

2. Requête de création de la table diabetes (SQL dans une BD Postgresql)

```
DROP TABLE IF EXISTS diabetes;
```

```
CREATE TABLE diabetes(  
  ID SERIAL PRIMARY KEY,  
  TARGET VARCHAR(255) NOT NULL,  
  AUTOANTIBODIES VARCHAR(10),  
  FAMILY_HIST VARCHAR(10),  
  ENVIRONMENT_FACT VARCHAR(10),  
  INSULIN_LEVELS INT,  
  AGE INT,  
  BMI INT,  
  PHYSICAL_ACT VARCHAR(10),  
  DIETARY_HABITS VARCHAR(10),  
  BLOOD_PRESSURE INT,  
  CHOLESTEROL_LEVEL INT,  
  BLOOD_GLUCOSE_LEVEL INT,  
  ETHNICITY VARCHAR(30),  
  SOCIOECONOMIC_FACTOR VARCHAR(30),  
  SMOKING_STATUS VARCHAR(30),  
  ALCOHOL_CONSUMPTION VARCHAR(30),  
  GLUCOSE_TOLERANCE VARCHAR(30),  
  HISTORY_PCOS VARCHAR(30),  
  PREVIOUS_GESTATIONAL_DIABETES VARCHAR(30),  
  PREGNANCY_HISTORY VARCHAR(30),  
  WEIGHT_GAIN_DURING_PREGNANCY INT,  
  PANCREATIC_HEALTH INT,  
  PULMONARY_FUNCTION INT,  
  CYSTIC_FIBROSIS_DIAGNOSIS VARCHAR(30),  
  STEROID_USE_HISTORY VARCHAR(30),  
  GENETIC_TESTING VARCHAR(30),  
  LIVER_FUNCTION_TESTS VARCHAR(30),  
  DIGESTIVE_ENZYME_LEVELS INT,  
  URINE_TEST VARCHAR(30),  
  EARLY_ONSET_SYMPTOMS VARCHAR(30));
```

3. Ensemble de requêtes pour la génération des graphiques sur Grafana

a. Nombre total de cas en fonction du type de diabète

```
SELECT

    COUNT(*) AS Nombre_total_de_cas,

    d1.target

FROM diabetes d1

GROUP BY d1.target

ORDER BY Nombre_total_de_cas DESC;
```

b. Nombre de cas liés à antécédents familiaux connus de diabète par rapport au nombre total de cas

```
SELECT

    COUNT(CASE WHEN d1.family_hist = 'Yes' THEN 1 END) AS
    Nombre_de_cas_liés_aux_antécédents,

    COUNT(*) AS Nombre_total_de_cas,

    d1.target

FROM diabetes d1

GROUP BY d1.target

ORDER BY Nombre_total_de_cas DESC;
```

c. Niveau maximal de taux de glucose chez un patient

```
select max(blood_glucose_level)

from diabetes ;
```

d. Niveau minimal de taux de glucose chez un patient

```
select min(blood_glucose_level)

from diabetes ;
```

e. Niveau maximal d'insuline chez une patient

```
select insulin_levels

from diabetes

order by insulin_levels desc ;
```

f. Niveau minimal d'insuline chez une patient

```
select insulin_levels

from diabetes

order by insulin_levels asc ;
```

g. Type de diabète en fonction du groupe d'âge

```
SELECT

CASE

    WHEN AGE BETWEEN 13 AND 19 THEN '13-19'

    WHEN AGE BETWEEN 20 AND 39 THEN '20-39'

    WHEN AGE BETWEEN 40 AND 59 THEN '40-59'

    WHEN AGE BETWEEN 60 AND 79 THEN '60-79'

    WHEN AGE >= 80 THEN '80+'

END AS age_group,

TARGET,

COUNT(*) AS count
```



```
FROM

    diabetes

GROUP BY

    age_group, TARGET

ORDER BY

    age_group, count DESC;
```

h. Statut tabagique des patientes

```
SELECT

    COUNT(smoking_status),

    smoking_status

FROM

    diabetes

GROUP BY

    smoking_status

ORDER BY

    COUNT(smoking_status) DESC
```

i. Habitudes alimentaires des patientes

```
select count(*) dtr_hbt , dietary_habits

from diabetes

GROUP BY dietary_habits

order by dtr_hbt asc;
```

j. Proportion de femmes qui consomment de l'alcool tout en étant fumeuses

```
SELECT count(alcohol_consumption) as cnt, alcohol_consumption
FROM diabetes
where smoking_status = 'Smoker'
group by alcohol_consumption
order by cnt desc;
```

k. Niveau d'activité physique des patientes

```
select count(*) cat_phy_act, physical_act
from diabetes
group by physical_act
order by cat_phy_act desc;
```

l. Types de diabète subis par les femmes qui consomment de l'alcool tout en étant fumeuses

```
SELECT count(target) as cnt, target
FROM diabetes
where smoking_status = 'Smoker' and (alcohol_consumption = 'High' or
alcohol_consumption='Moderate')
group by target
order by cnt desc;
```

m. Types de diabète subis par les femmes fumeuses

```
SELECT count(target) as cnt, target
FROM diabetes
```

```
where smoking_status = 'Smoker'  
  
group by target  
  
order by cnt desc;
```

**n. Types de diabète subis par les femmes qui consomment de l'alcool
(High & Moderate)**

```
SELECT count(target) as cnt, target  
  
FROM diabetes  
  
where alcohol_consumption = 'High' or alcohol_consumption='Moderate'  
  
group by target  
  
order by cnt desc;
```

4. Sources (Rapport)

Administrator, G. (s. d.). *Les bienfaits de l'activité sportive sur la santé physique et mentale*. Croix-Rouge Canadienne.

<https://www.croixrouge.ca/blogue/2021/2/les-bienfaits-de-l-activite-sportive-sur-la-sante-physique-et-mentale>

Diabetes - Symptoms and causes. (s. d.). Mayo Clinic.

<https://www.mayoclinic.org/diseases-conditions/diabetes/symptoms-causes/syc-20371444>

Hardy, K. (2023, 29 octobre). *Secondary diabetes - conditions, insulin resistance & glucagon*. Diabetes.

<https://www.diabetes.co.uk/secondary-diabetes.html>

L'activité physique et sportive : un atout essentiel pour le bien-être. (s. d.). ameli.fr | Assuré.

<https://www.ameli.fr/haute-garonne/assure/sante/themes/activite-physique-sante/sport-et-activite-physique-un-atout-pour-le-bien-etre#:~:text=Activit%C3%A9%20physique%20et%20bien%2D%C3%AAtre,niveau%20de%20la%20sant%C3%A9%20mentale>

News-Medical. (2022, 22 décembre). *Genetic Testing for Diabetes*.

<https://www.news-medical.net/health/Genetic-Testing-for-Diabetes.aspx>

Professional, C. C. M. (2024a, août 2). *Pancreas blood test*. Cleveland Clinic.

<https://my.clevelandclinic.org/health/diagnostics/23476-pancreas-blood-test>

Professional, C. C. M. (2024b, août 2). *Pancreas tests*. Cleveland Clinic.

<https://my.clevelandclinic.org/health/diagnostics/12029-pancreas-function-tests>

Wikipedia contributors. (2001, août 15). *Diabetes*. Wikipedia. <https://en.wikipedia.org/wiki/Diabetes>

Yang, B., Li, J., Haller, M. J., Schatz, D. A., & Rong, L. (2022). The progression of secondary diabetes : A review of modeling studies. *Frontiers In Endocrinology*, 13.
<https://doi.org/10.3389/fendo.2022.1070979>