# DATASET

La base de datos a analizar en esta tarea es *Diabetes.csv*, el cual contiene información sobre 768 mujeres de la raza Pima india en un estudio de diabetes. Para cada una de las pacientes se recogió información sobre los siguientes atributos:

* Pregnant: número de embarazos
* Glucose: concentración de glucosa en plasma en un test oral de tolerancia a la glucosa
* Boodpress: presión diastólica sanguínea (mm Hg)
* Bodymassindex: espesor del pliegue cutáneo del tríceps (mm)
* Insulin: insulina sérica de 2 horas (mu U/ml)
* Skinthickness: índice de masa corporal (peso en kg/ (altura en m)2)
* Pedigreefunc: resumen de la historia de diabetes mellitus en familiares y la relación genética de esos relativos con el sujeto
* Age: edad en años
* Output: Diabetes: variable clase (0 o 1)

1. EXPLORATORY DATA ANALYSIS

Previamente a la aplicación de cualquier técnica, conviene realizar un análisis exploratorio (EDA) en el que se realice una descripción numérica y gráfica de los datos de estudio recopilando la información necesaria que nos ayude a plantear las primeras hipótesis en la investigación.

Así, como punto de partida en este apartado, cabe preguntarse por la naturaleza de las variables y por la existencia de valores ausentes. En cuanto a esta última característica, no se observa ningún valor vacío y, en relación a las variables, se cuenta con un total de 9 atributos que se pueden clasificar en:

* Cuantitativas: Pregnant (Discreta), Glucose, Boodpress, Bodymassindex, Insulin, Skinthickness, Pedigreefunc, Age.
* Cualitativas: Diabetes

Dado que en esta base de datos todas las variables explicativas son cuantitativas,

resulta interesante llevar a cabo un resumen numérico sobre ellas:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Min | Q1 | Mediana | Media | Q3 | Max |
| PREGNANT | 0.000 | 1.000 | 3.000 | 3.845 | 6.000 | 17.000 |
| GLUCOSE | 0.0 | 99.0 | 117.0 | 120.9 | 140.2 | 199.0 |
| BLOODPRESS | 0.00 | 62.00 | 72.00 | 69.11 | 80.00 | 122.00 |
| SKINTHICKNESS | 0.00 | 0.00 | 23.00 | 20.54 | 32.00 | 99.00 |
| INSULIN | 0.0 | 0.0 | 30.5 | 79.8 | 127.2 | 846.0 |
| BODYMASSINDEX | 0.00 | 27.30 | 32.00 | 31.99 | 36.60 | 67.10 |
| PEDIGREEFUNC | 0.0780 | 0.2437 | 0.3725 | 0.4719 | 0.6262 | 2.4200 |
| AGE | 21.00 | 24.00 | 29.00 | 33.24 | 41.00 | 81.00 |