

Evidencia N° 2 Analista de Datos

Análisis preliminar del Dataset médico sobre infarto de miocardio

1. Origen de los datos

Los datos provienen del hospital Zheen en Erbil, Iraq, y fueron recopilados entre enero y mayo de 2019. El dataset está alojado en Kaggle y fue publicado por la Universidad de Kurdistán Hewlêr.

Link:

<https://www.kaggle.com/datasets/fatemehmohammadinia/heart-attack-dataset-tarik-a-rashid>

2. Descripción de cada columna

Columna	Tipo de dato	Descripción
Age	Numérico (float)	Edad del paciente.
Gender	Numérico (float)	Género: 1 = Masculino, 0 = Femenino.
Heart rate	Numérico (int)	Frecuencia cardíaca del paciente (latidos por minuto).
Systolic blood pressure	Numérico (int)	Presión sistólica (presión máxima durante el latido).
Diastolic blood pressure	Numérico (int)	Presión diastólica (presión entre latidos).
Blood sugar	Numérico (float)	Nivel de glucosa en sangre.
CK-MB	Numérico (float)	Enzima cardíaca usada para detectar infarto.
Troponin	Numérico (float)	Proteína que se eleva ante daño al miocardio.
Result	Categorico (string)	Diagnóstico: "positive" (1) o "negative" (0).

3. Valores posibles por columna

- **Age:** valores numéricos positivos (edad del paciente).
- **Gender:** 1 (hombre), 0 (mujer).
- **Heart rate:** valor numérico entero positivo.
- **Systolic blood pressure y Diastolic blood pressure:** valores numéricos enteros positivos.
- **Blood sugar:** valor continuo, sin normalizar en esta versión (no binario).
- **CK-MB y Troponin:** valores numéricos flotantes, representando niveles en sangre.
- **Result:** “positive” o “negative”.

4. Formato de los datos

Los datos se presentan en **formato RAW**, lo que implica que se encuentran en su forma original, tal como fueron recopilados en el entorno hospitalario, sin haber sido sometidos a procesos avanzados de limpieza, transformación o normalización. Si bien algunas columnas muestran una codificación básica (por ejemplo, el género está representado como 1 para masculino y 0 para femenino), el conjunto en general conserva su estructura inicial. Esto permite un análisis exploratorio más profundo, pero también requiere una etapa previa de depuración para detectar y corregir posibles inconsistencias, valores nulos o errores de registro.