

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas
(IIMAS)
Ciencia de Datos
Visualización de la Información

TAREA-1

04.02.2022

1. Informe de enfermedades del corazón

1.1. Introducción

De acuerdo con la Organización Mundial de la Salud (OMS) las enfermedades cardiovasculares son la principal causa de defunción en el mundo y, según estimaciones, se cobran 17,9 millones de vidas cada año. Estas enfermedades constituyen un grupo de trastornos del corazón y los vasos sanguíneos que incluyen cardiopatías coronarias, enfermedades cerebrovasculares y cardiopatías reumáticas. Más de cuatro de cada cinco defunciones por enfermedades cardiovasculares se deben a cardiopatías coronarias y accidentes cerebrovasculares, y una tercera parte de esas defunciones ocurren prematuramente en personas menores de 70 años.

1.2. Análisis Exploratorio

En el archivo **heart.csv** se presenta un conjunto de datos reales que incluye las características importantes de pacientes relacionados a enfermedades cardiovasculares. Hay trece características y un objetivo como se muestra a continuación:

- **age**: la edad de la persona en años
- **sex**: el sexo de la persona (1 = masculino, 0 = femenino)
- **cp**: El dolor torácico experimentado (Valor 1: angina típica, Valor 2: angina atípica, Valor 3: dolor no anginoso, Valor 4: asintomático)
- **trestbps**: la presión arterial en reposo de la persona (mm Hg al ingreso en el hospital)
- **chol**: la medida de colesterol de la persona en mg/dl
- **fbs**: azúcar en sangre en ayunas de la persona (>120 mg/dl, 1 = verdadero; 0 = falso)

- **resecg**: medición electrocardiográfica en reposo (0 = normal, 1 = con anormalidad de onda ST-T, 2 = mostrando hipertrofia ventricular izquierda probable o definitiva según los criterios de Estes)
- **thalach**: la frecuencia cardíaca máxima alcanzada por la persona
- **exang**: angina inducida por el ejercicio (1 = sí; 0 = no)
- **oldpeak**: depresión del ST inducida por el ejercicio en relación con el reposo
- **slope**: la pendiente del segmento ST del ejercicio máximo (Valor 1: pendiente ascendente, Valor 2: plano, Valor 3: pendiente descendente)
- **ca**: El número de vasos principales (0-3)
- **thal**: un trastorno de la sangre llamado Talasemia (3 = normal; 6 = defecto fijo; 7 = defecto reversible)
- **target**: enfermedad cardíaca (0 = no, 1 = sí)

1.3. Actividades a realizar

Utilice el archivo **heart.csv** en la versión de dataset para extraer los datos y determinar el porcentaje de personas que han sido diagnosticados con una enfermedad cardíaca (**target**, donde 0=no, 1=sí).

- Elabore un histograma o grafica de barras que permita visualizar la *edad* (**age**) comparada con el porcentaje de personas si diagnosticadas con una enfermedad cardíaca y las que no han sido diagnosticadas (**target**, donde 0=no, 1=sí).
- Elabore un histograma que permita visualizar la *presión arterial* (**trestbps**) comparada con el porcentaje de personas si diagnosticadas con una enfermedad cardíaca y las que no han sido diagnosticadas.