

**IESETyFP - Instituto de Educación Superior de Educación Técnica y Formación Profesional.**

Tecnicatura Superior en Ciencia de Datos e Inteligencia Artificial.

* Actividad: Trabajo Práctico N° 1.
* Espacio Curricular: Base y Procesamiento de Datos.
* Profesor: Lucas RIOS.
* Alumnos: Emanuel Horacio QUIÑONEZ.

Sebastián Alcides RIVERO.

Mapik Guido GODOY.

Julián SANTÁGATA.

Lucas BAZÁN.

Ciclo Lectivo 2025

**Introducción**

Este informe presenta el desarrollo completo de un análisis exploratorio y predictivo sobre una base de datos de pacientes con potencial riesgo de enfermedad cardíaca. El propósito es identificar patrones relevantes, visualizar comportamientos por grupo etario, género y síntomas clínicos, y realizar una modelización preliminar con enfoque predictivo.

La base de datos analizada cuenta con **1000 registros de pacientes**, cada uno con variables como edad, género, presión arterial, colesterol, síntomas de dolor en el pecho, frecuencia cardíaca, resultados de electrocardiograma y diagnóstico clínico final.

**Metodologia**

Cómo grupo utilizamos las siguientes etapas para desarrollar el informe:

* **Carga inicial** de datasets de pacientes
* **Cálculos estadísticos descriptivos** para identificar promedios, extremos, medianas y modas.
* **Segmentación de variables categóricas** (género, tipo de dolor en el pecho, diagnóstico)
* **Visualización de datos** mediante gráficos (barras, histogramas, boxplots, heatmaps, gráficos de dispersión y torta).
* **Modelado predictivo básico**, a través de regresión lineal, para observar tendencias en las variables vinculadas al diagnóstico.

## ***RESULTADOS DEL ANÁLISIS DESCRIPTIVO***

### ***Edad de pacientes***

* ***Edad mínima:*** *20 años*
* ***Edad máxima:*** *80 años*
* ***Edad promedio:*** *49.24 años*
* ***Mediana:*** *49 años*
* ***Moda:*** *20 año*s

Este rango amplio permite identificar patrones en diferentes grupos etarios, siendo los grupos entre 40 y 60 años los que presentan mayor concentración de pacientes.

### **Distribución por género**

Los registros marcan:

***Mujeres: 235 casos (23.5% del total)***

***Hombres: 765 casos (76.5% del total)***

*\*Son los dos géneros sin separar por diagnóstico final*

### 

### Tipos de dolor en el pecho

La siguiente variable es clave para indicar señales tempranas. La interpretación de valores es bajo nuestra comprensión:

***Típico*** [***anginoso***](https://medlineplus.gov/spanish/angina.html#:~:text=La%20angina%20es%20un%20dolor,Se%20trata%20con%20medicinas.) ***(0): 420 casos (42%)***

[***Angina atípica***](https://pmc.ncbi.nlm.nih.gov/articles/PMC7270897/#:~:text=La%20angina%20at%C3%ADpica%20se%20define,con%20una%20causa%20isqu%C3%A9mica%20card%C3%ADaca.) ***(1): 224 casos (22.4%)***

***Dolor no anginoso (2): 312 casos (31.2%)***

[***Asintomático***](https://medlineplus.gov/spanish/ency/article/002217.htm) ***(3): 44 casos (4.4%)***

*Los pacientes que presentan dolor* ***tipo anginoso o son asintomáticos*** *en general reflejan mayor* ***asociación con un diagnóstico positivo****.*

### Presión en reposo

* **Presión mínima:** 90 mmHg
* **Presión promedio:** 151.75 mmHg
* **Presión máxima:** 200 mmHg

*Estos datos son* ***claves*** *para identificar riesgos cardiovasculares ocultos.*

### Glucemia en ayunas

Bajo nuestra interpretación, 1 mayor a 120 mg/dL y 0 menor o igual a 120 mg/dL.

**Menor o igual a 120 mg/dL: 704 pacientes (70.4%)**

**Mayor a 120 mg/dL: 296 pacientes (29.6%)**

Este parámetro ayuda a entender la presencia de [comorbilidades](https://www.cancer.gov/espanol/publicaciones/diccionarios/diccionario-cancer/def/comorbilidad) como la diabetes en relación con enfermedades cardíacas.

### **Diagnostico final**

***Pacientes sin enfermedad cardíaca (taget=0): 420 pacientes (42%)***

***Pacientes con diagnóstico de enfermedad cardíaca (1): 580 pacientes (58%)***

*Más de la mitad presenta algún grado de riesgo clínico, lo que subraya la importancia de una intervención temprana.*

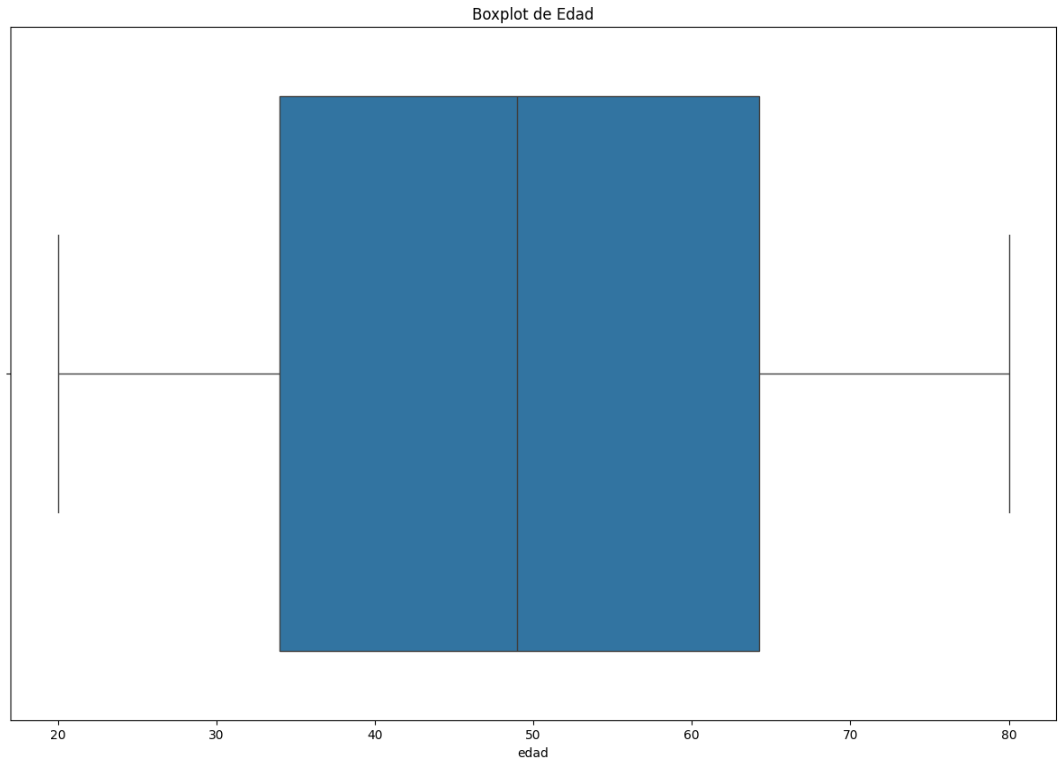
#### ***Conclusión***

*En conjunto, los resultados sugieren que una proporción significativa de los pacientes presenta factores de riesgo acumulativos: edad avanzada, presión elevada, glucemia fuera de rango y tipos de dolor en el pecho preocupantes. Esta combinación de indicadores justifica intervenciones preventivas y un monitoreo clínico activo.*

## VISUALIZACIÓN DE DATOS

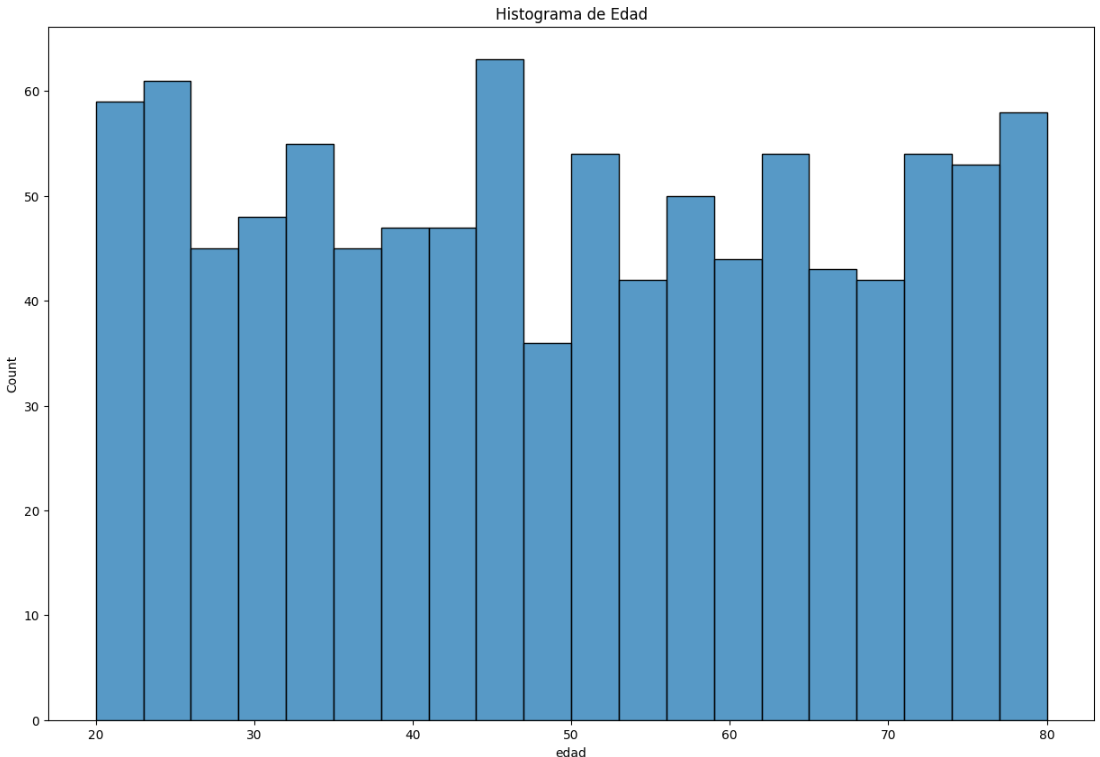
### Boxplot de edad

*Muestra la distribución de las edades, incluyendo la mediana, los cuartiles y posibles valores atípicos, lo que permite visualizar la variabilidad de la edad en el conjunto de datos.*



### 

### Histograma por edad

*Representa la frecuencia de las edades agrupadas en rangos, permitiendo observar cómo se distribuyen los pacientes según su edad.*  


### 

### 

### 

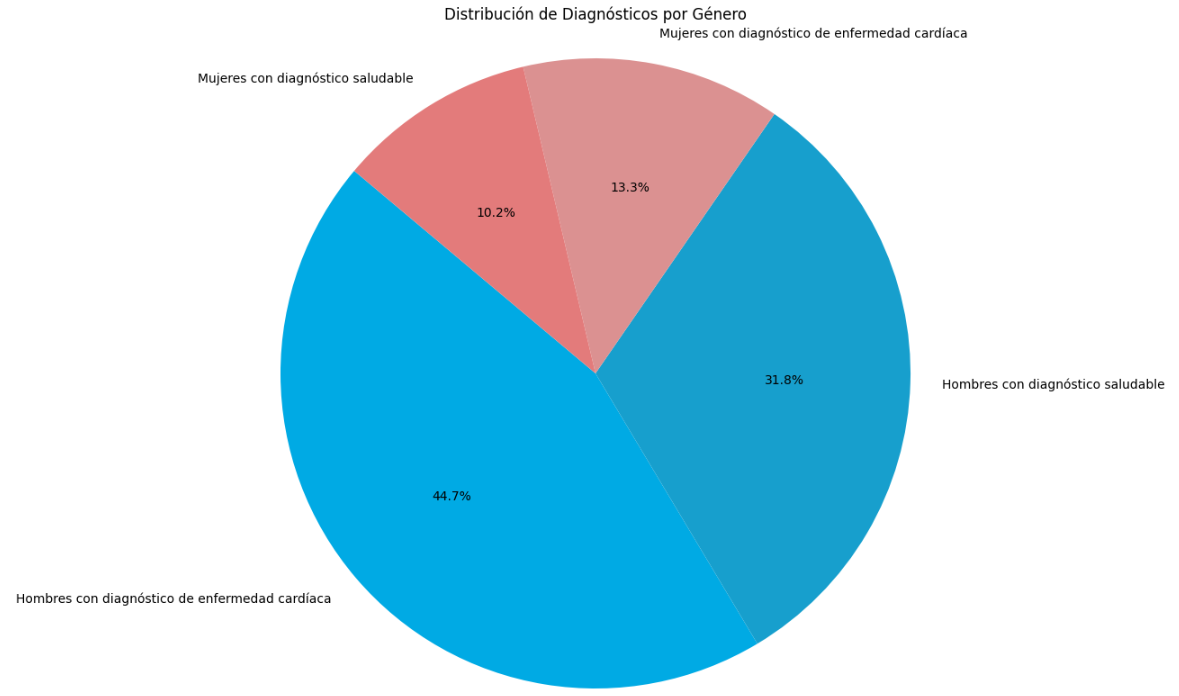
### 

### 

### 

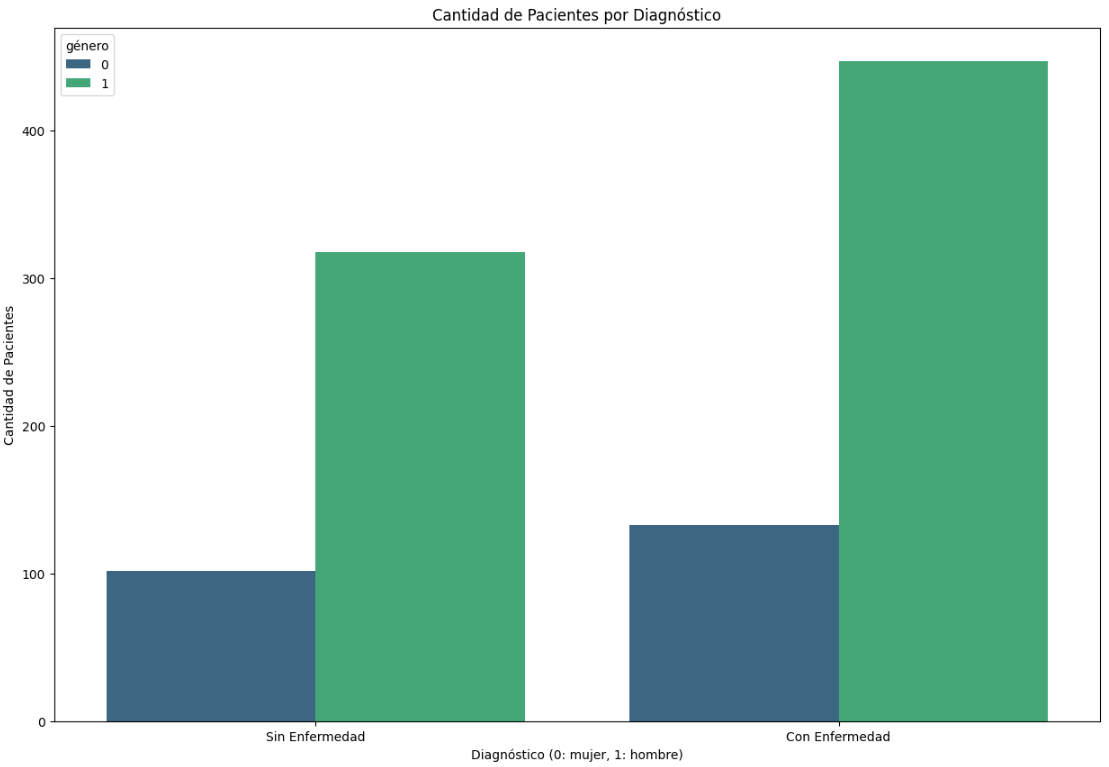
### Gráfico de torta (Género y diagnóstico)

*Visualiza la proporción de diagnósticos dentro de cada género, facilitando la comparación del porcentaje de casos entre hombres y mujeres.*



### Gráfico de columnas (género y diagnóstico)

*Compara el número de diagnósticos entre géneros de forma clara y cuantitativa, permitiendo ver qué género presenta más casos en cada categoría.*



### 

### Histograma 2D (edad x frecuencia)

*Muestra la densidad conjunta entre la edad y otra variable (por ejemplo, frecuencia cardíaca), facilitando la identificación de patrones entre ambas variables.*

### 

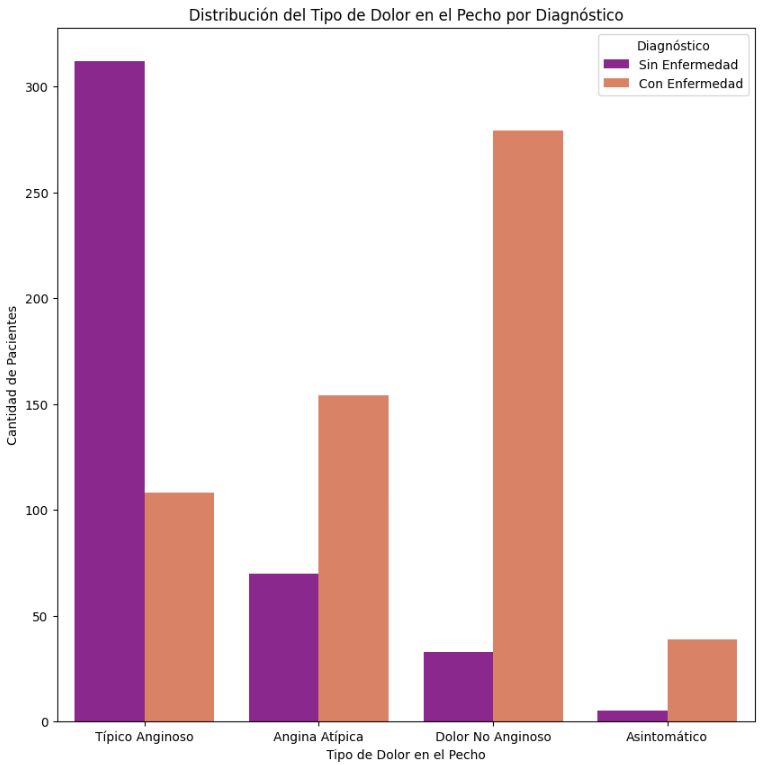
### 

### 

### 

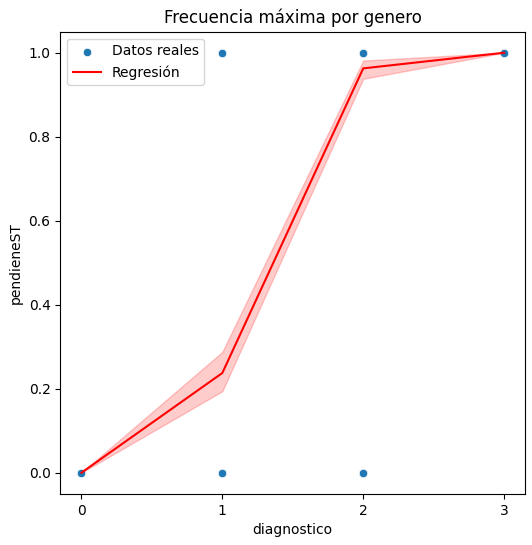
### Gráfico de columnas (tipo de dolor en el pecho y diagnóstico)

*Permite observar la relación entre los distintos tipos de dolor torácico y los diagnósticos, mostrando cuántos casos de cada diagnóstico se asocian a cada tipo de dolor.*

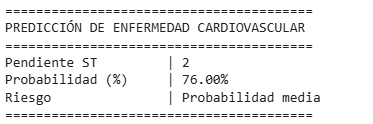


### Entrenamiento del Modelo

*Este gráfico muestra la relación entre la pendiente del segmento ST (pendienteST) y el diagnóstico de enfermedad cardíaca. Los puntos representan los datos reales del conjunto, mientras que la línea roja indica la tendencia estimada por el modelo de regresión lineal. Esta visualización permite observar cómo varía el diagnóstico en función de los distintos valores de pendienteST.*



### Interfaz para usuarios



Por último, proponemos una interfaz (en estos momentos, se encuentra en la primera etapa de desarrollo), la cuál permite al usuario ingresar el número de pendientes ST que sale en los papeles de su análisis, con dicho valor, la interfaz arroja la probabilidad y el riesgo de presentar una enfermedad cardiovascular. Cabe aclarar que los resultados de predicción están basados en 1000 casos de pacientes entre ellos, personas con o sin resultados de la enfermedad.

### 

### Enlace al GitHub para visualizar el código.

<https://drive.google.com/file/d/10Q4eN_dDwHxf4aAX7T1iMC7pvd2NQDgy/view?usp=sharing>