

Proyecto 1:

# Predicción de Enfermedades Cardiacas

Santiago González  
Juliana Cárdenas

# Contenido:



## Introducción

Motivación de la realización del proyecto.



## Datos

Análisis exploratorio de los datos.



## Modelo

Desarrollo del modelo por medio de Redes Bayesianas.



## Producto

Aplicación desarrollada en Dash.



## Conclusiones

Que se pudo aprender del proyecto.

# Introducción:

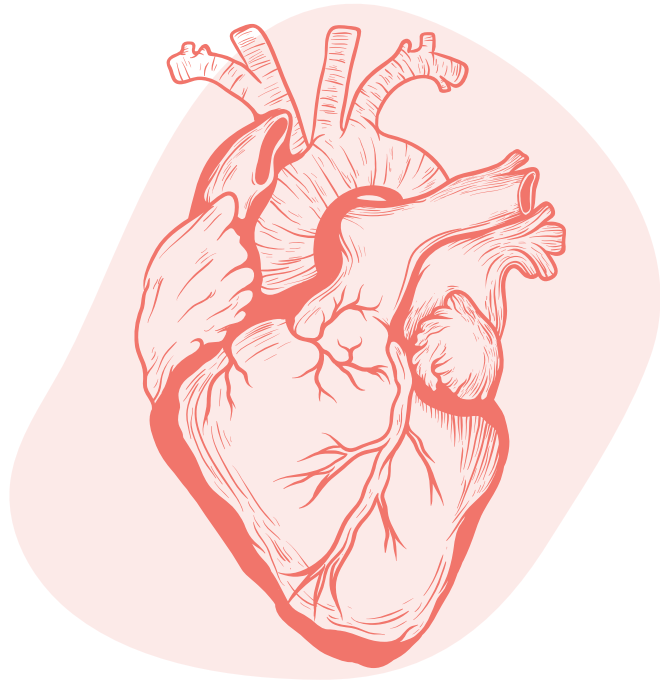
1

El corazón es uno de los órganos más importantes del cuerpo humano y las enfermedades cardíacas son una de las principales causas de muerte a nivel mundial.

2

El **objetivo principal** de este proyecto es determinar que factores influyen en la aparición de enfermedades cardíacas y desarrollar una aplicación que cuente con un modelo predictivo por medio de redes bayesianas para identificar a los pacientes en riesgo.





# Análisis exploratorio de los datos

# Variables:

Se explorarán los datos para Enfermedades Cardíacas, el cual es compuesto por **14 atributos** para los que se cuenta con **303 observaciones**.

## Variables Cuantitativas

- Edad
- Presión arterial
- Colesterol sérico
- Frecuencia cardíaca máxima
- Depresión segmento ST

## Variables Categóricas

- Sexo
- Tipo de dolor en el pecho
  - Azúcar en sangre
  - Resultados electrocardiográficos en reposo
- Angina inducida por ejercicio
  - Pendiente segmento ST
- Numero de vasos coloreados
  - Talasemia
- Presencia enfermedad cardíaca

# Tipos de análisis:

## Análisis Univariado

Una vez se tienen claras las variables y la categoría a la cual pertenece cada una, se procede a realizar un análisis de las distribuciones de estas.

- Variables cuantitativas
- Variables categóricas

## Análisis Multivariado

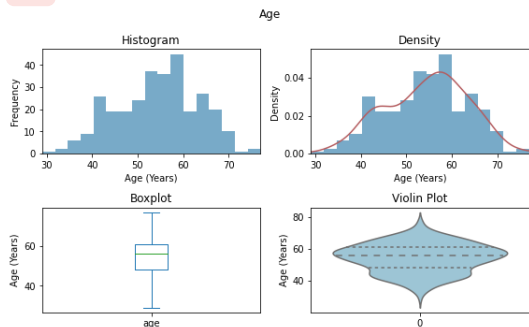
Se busca entender la relación entre las variables

- Categóricas vs Categóricas
- Categóricas vs Numéricas
- Numéricas vs Numéricas

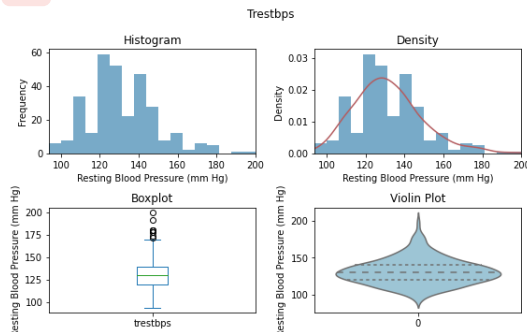
# Variables cuantitativas:

El análisis para las variables numéricas consiste en cuatro graficas: un histograma con  $k=\sqrt{n}$  numero de clases, una grafica de densidad ajustada a la distribución, un diagrama de caja y un diagrama de violín.

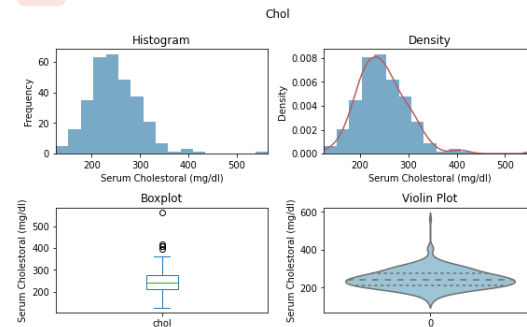
## Edad



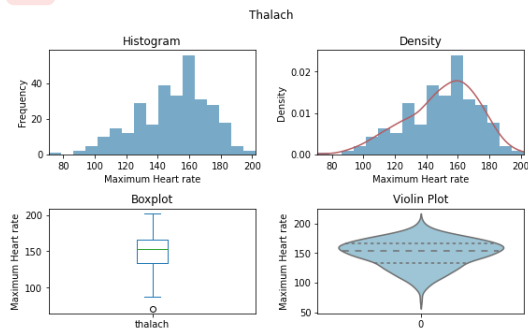
## Presión arterial



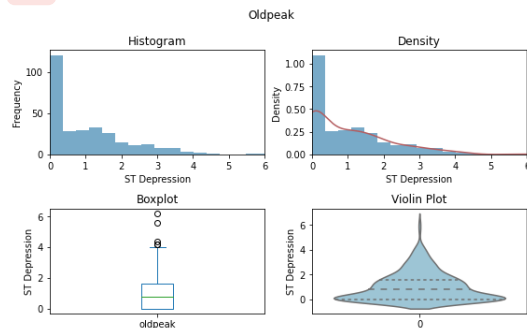
## Colesterol sérico



## Frecuencia cardiaca máxima

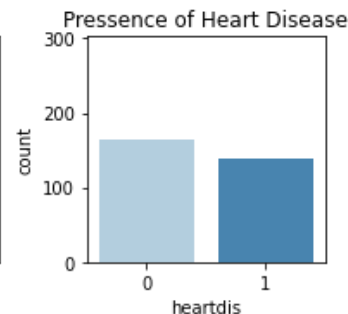
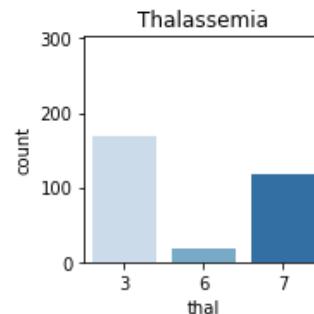
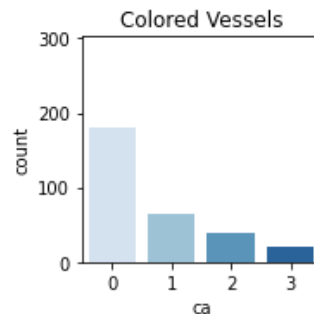
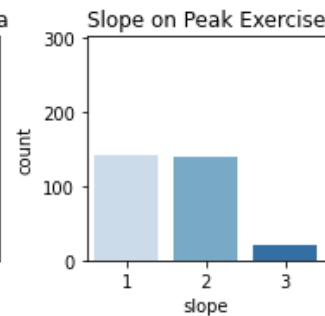
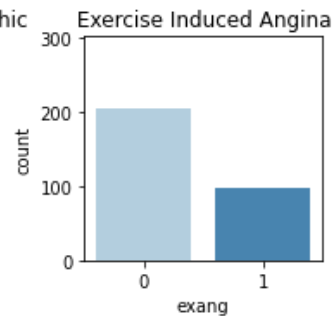
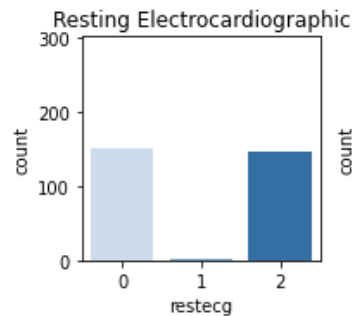
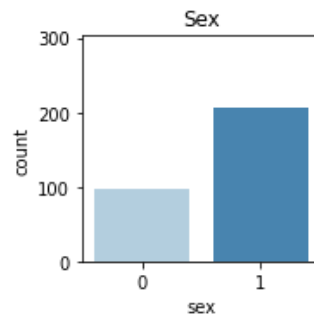


## Depresión segmento ST



# Variables categóricas:

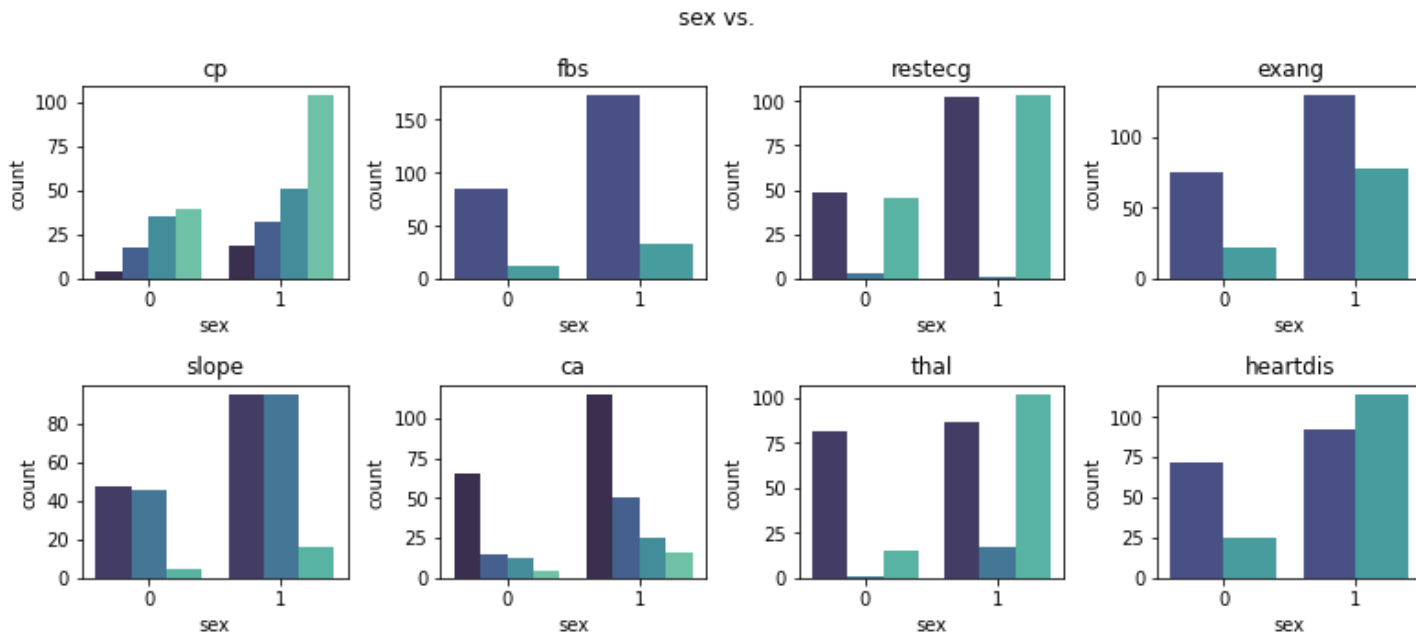
Para las variables categóricas se propondrá un análisis a partir de las cuentas para cada uno de los valores que puede tomar la variable categórica, esto permitirá saber la distribución de la variable y ver si los datos se encuentran distribuidos uniformemente o si hay prelación por ciertos atributos de la variable.





# Catégoricas vs Catégoricas: Sexo

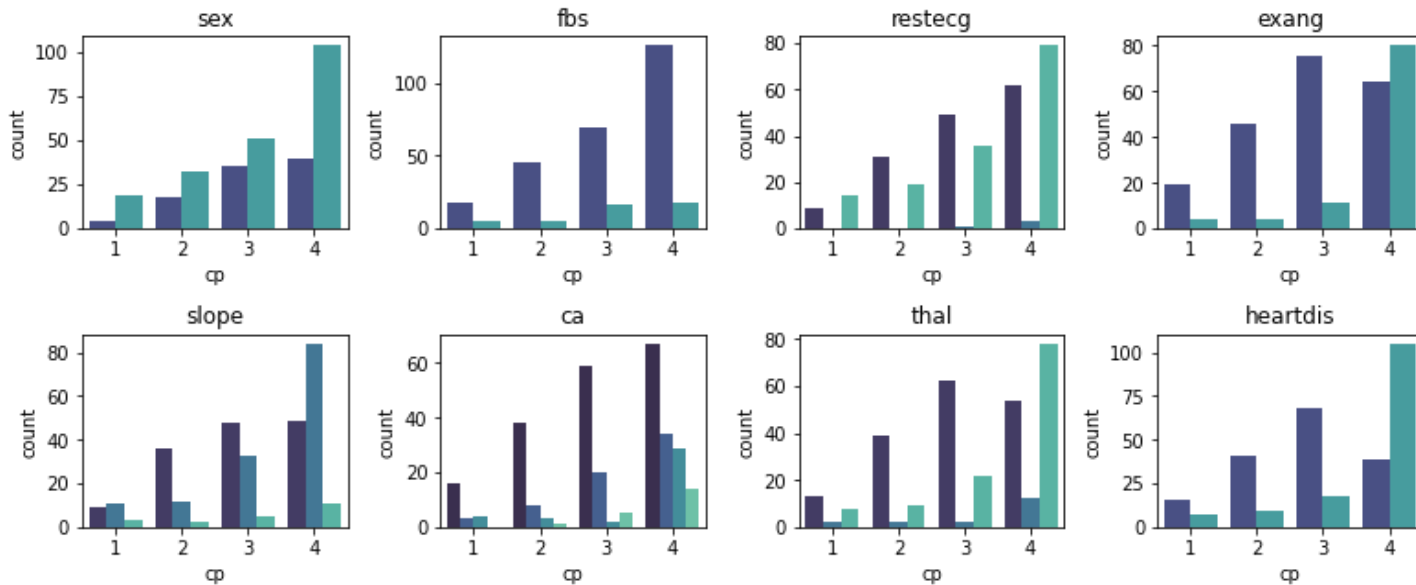
- El sexo si afecta o se ve afectado por las variables catégoricas **tal y heartdis**. Esto dado que hay una diferencia en la distribución de estas variables catégoricas si el sexo es femenino a si es masculino.



# Categóricas vs Categóricas: Tipo Dolor de Pecho

El tipo de dolor de pecho si afecta o se ve afectado por las variables categóricas **exang** y **heartdis**. Esto dado que hay una diferencia en la distribución de estas variables categóricas si el dolor es anginal típico, atípico, no anginal o asintomático.

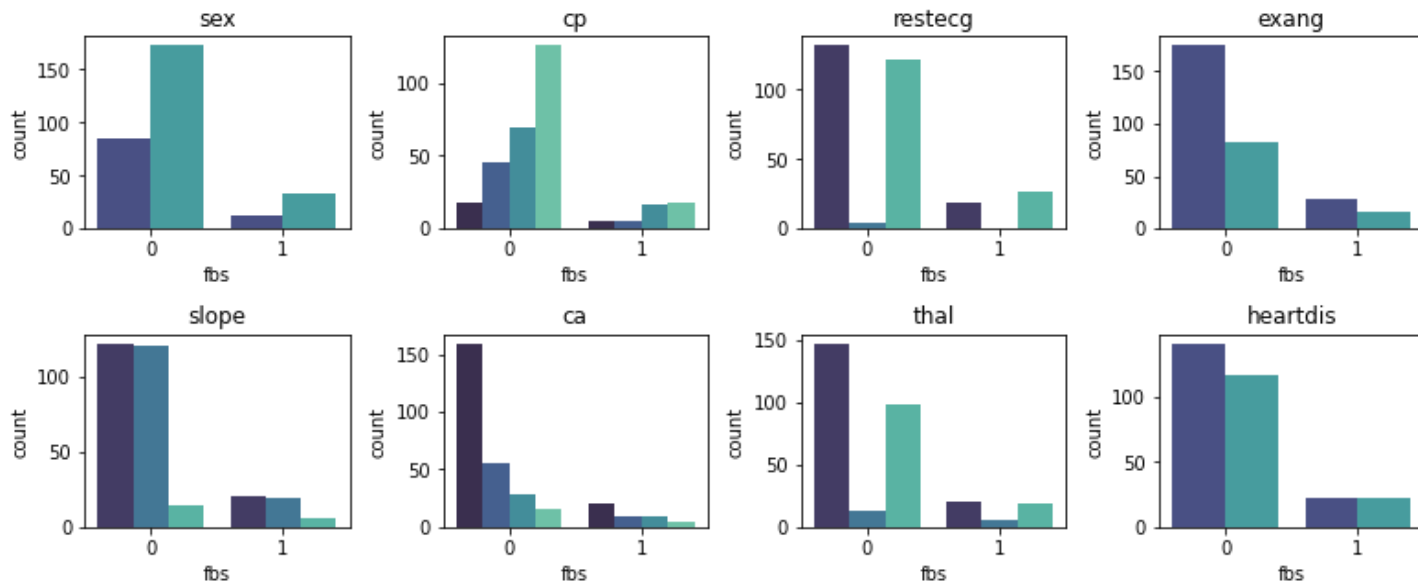
cp vs.



# Categorías vs Categorías: Azúcar en sangre

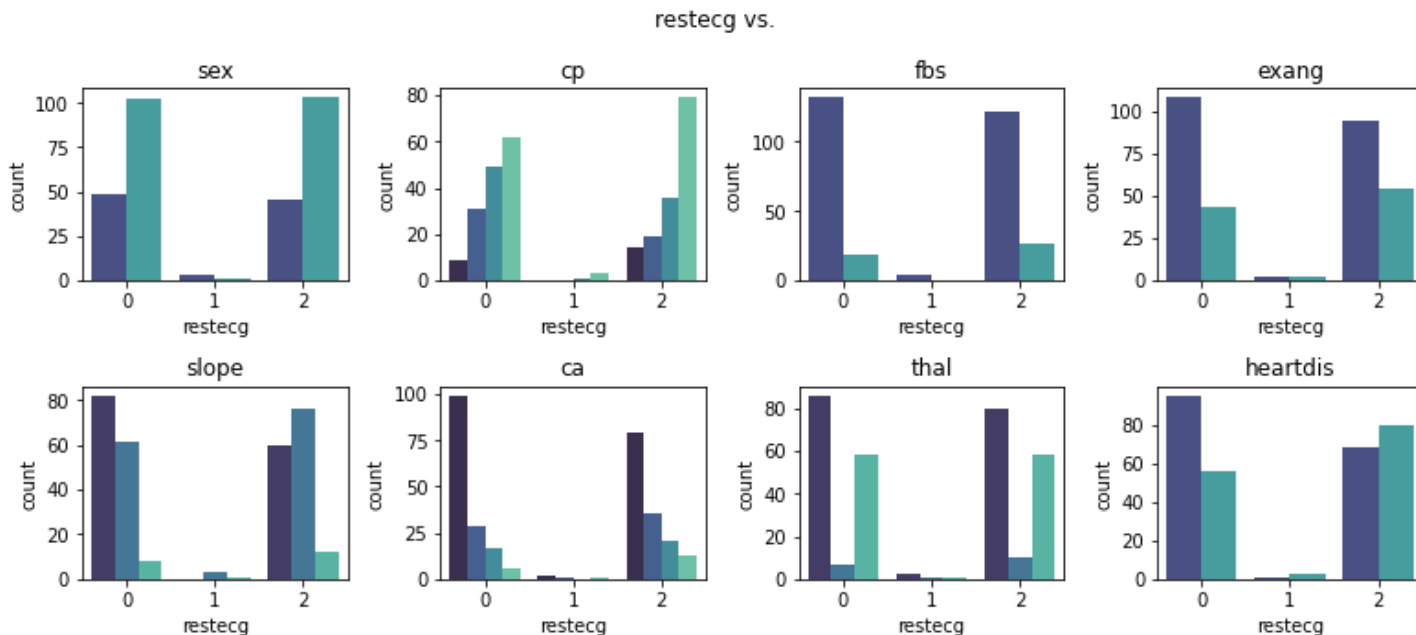
Si el azúcar en sangre en ayunas es mayor a 120 mm Hg **no afecta o se ve afectado** por las variables categóricas. Esto dado que hay no una diferencia en la distribución de estas variables categóricas si es mayor o no el azúcar.

fbs vs.



# Categorías vs Categorías: Resultados electrocardiográficos en reposo

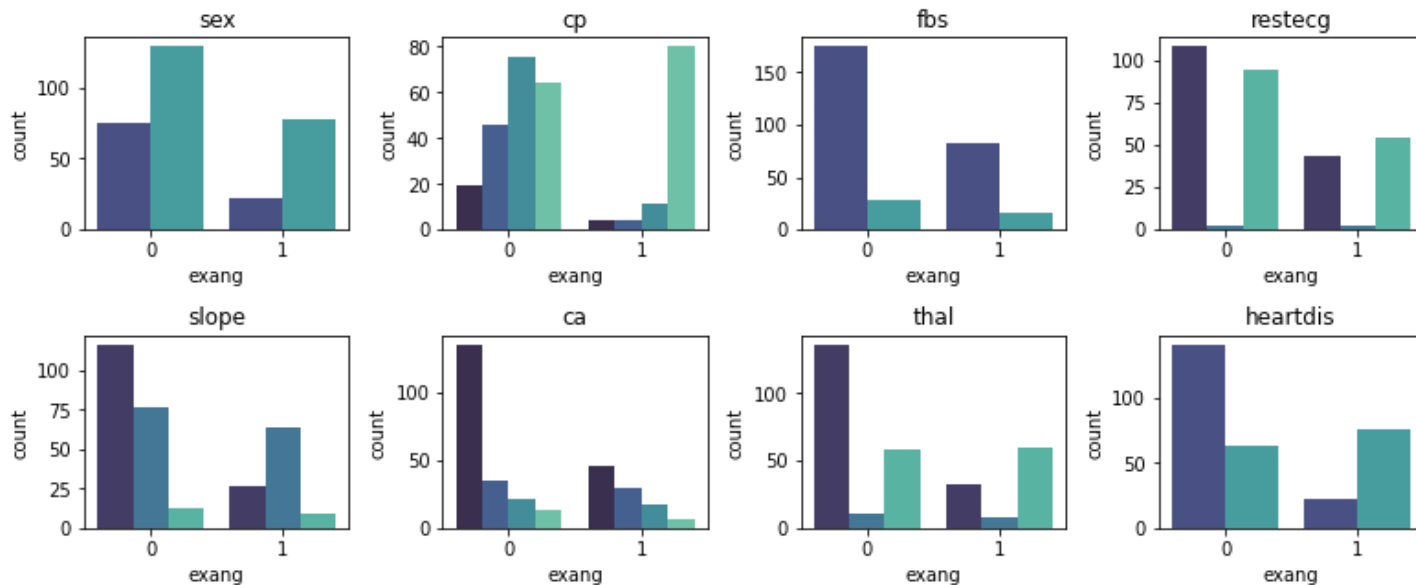
El resultado electrocardiográfico en reposo **no afecta o se ve afectado** por las variables categóricas. Esto dado que hay no hay diferencia en la distribución de las variables categóricas dado el resultado del examen.



# Catg3ricas vs Catg3ricas: Angina inducida por ejercicio

Si la angina se produce por hacer ejercicio si afecta o se ve afectado por las variables catg3ricas **thal** y **heartdis**. Esto dado que hay una diferencia en la distribuci3n de estas variables si la angina es producida o no.

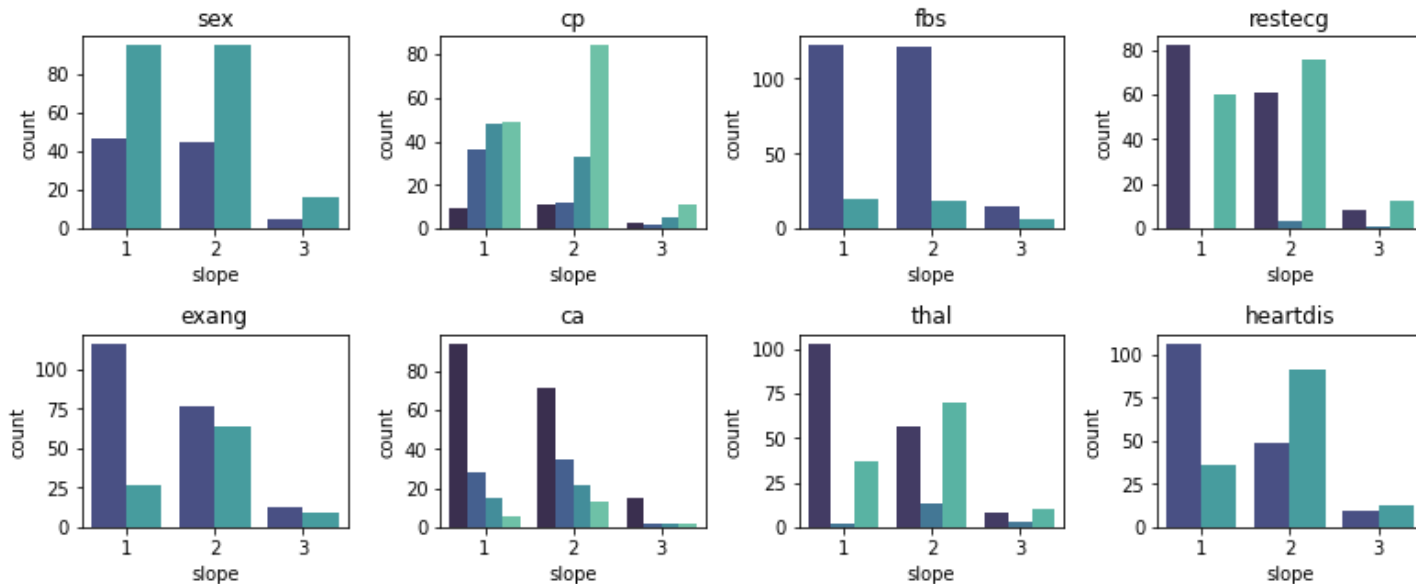
exang vs.



# Categorías vs Categorías: Pendiente segmento ST

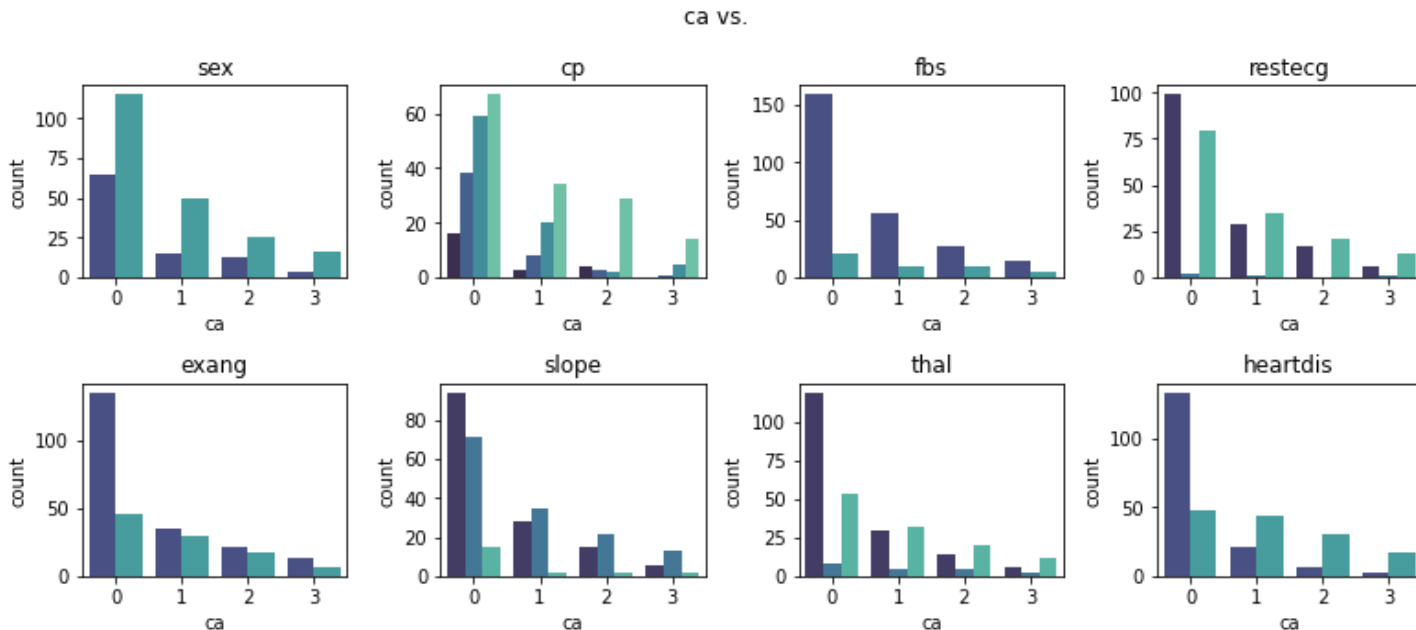
La pendiente del segmento ST en el pico de ejercicio si afecta o se ve afectado por la variable categórica **heartdis**. Esto dado que hay una diferencia en la distribución de esta variable categórica si la pendiente cambia.

slope vs.



# Catégoricas vs Catégoricas: Numero vasos coloreados

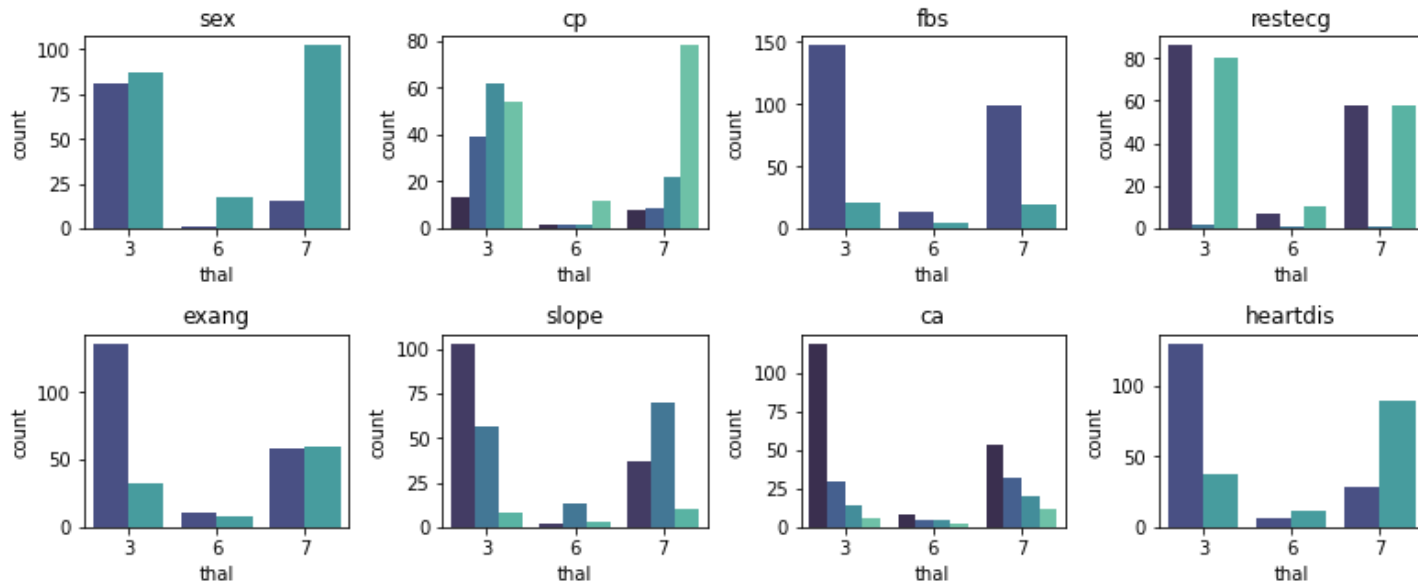
El número de vasos coloreados por fluoroscopia si afecta o se ve afectado por la variable categórica **heartdis**. Esto dado que hay una diferencia en la distribución de esta variable categórica si el número cambia.



# Catégoricas vs Catégoricas: Talasemia

La enfermedad talasemia si afecta o se ve afectado por las variables catégoricas **sex**, **cp**, **exang**, **slope**, **ca** y **heartdis**. Esto dado que hay una diferencia en la distribución de estas variables catégoricas según el tipo de talasemia.

thal vs.

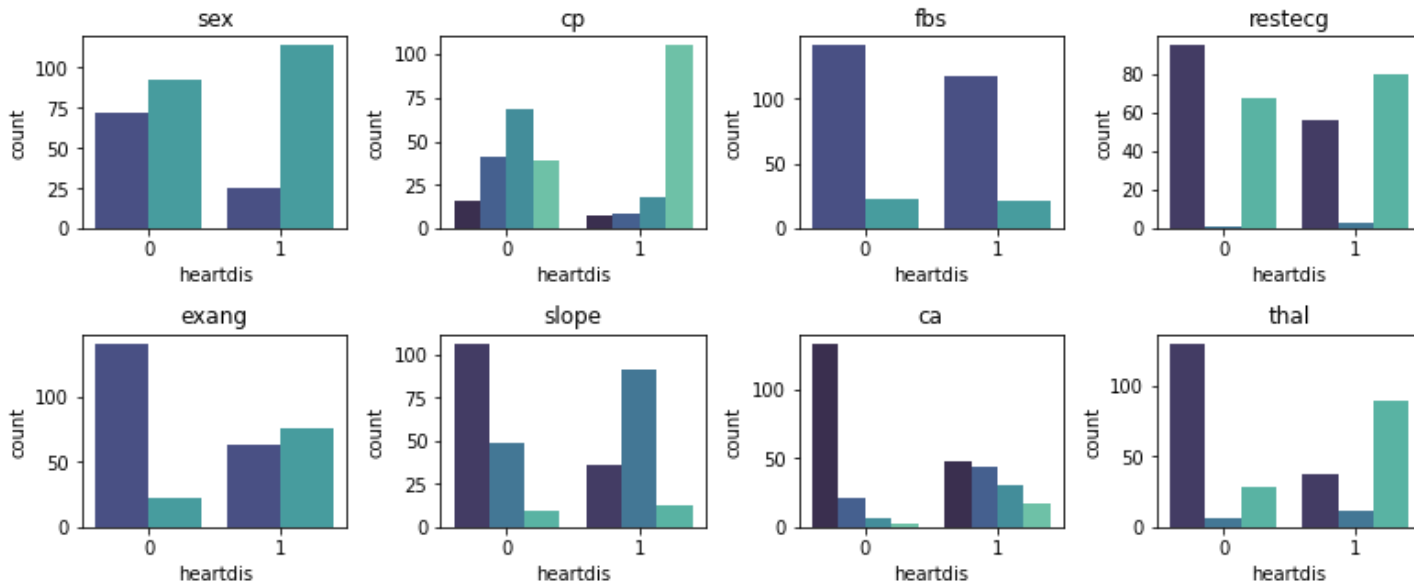




# Categorías vs Categoricals: Presencia enfermedad cardíaca

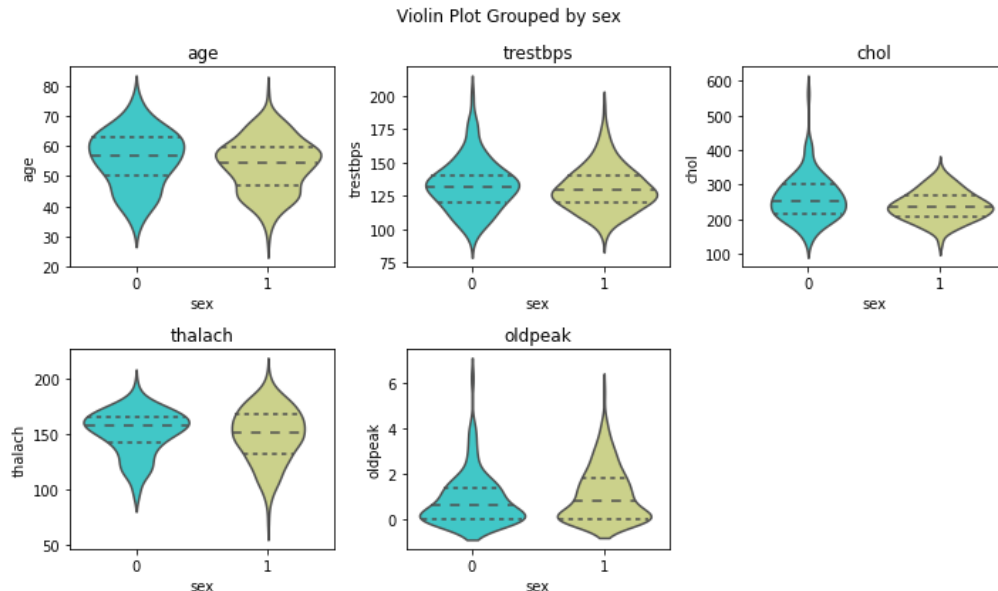
La presencia de enfermedades cardíacas si afecta o se ve afectado por las variables categóricas **sex**, **exang**, **slope**, **thal**. Esto dado que hay una diferencia en la distribución de estas variables si la pendiente cambia.

heartdis vs.



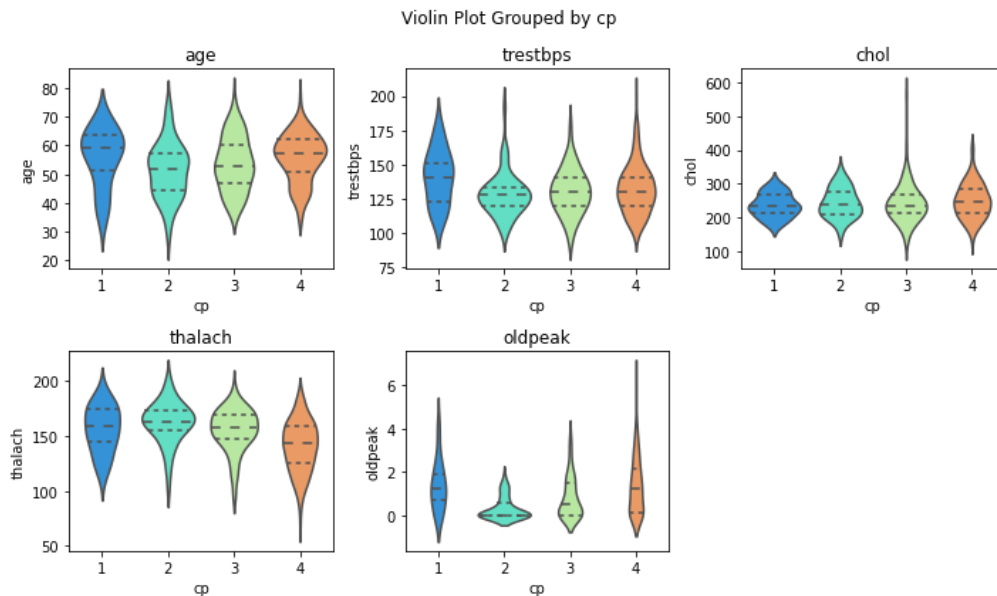
# Categóricas vs Numéricas: Sexo

El sexo de la persona **no afecta o se ve afectado** por ninguna variable numérica. Esto se debe a que no hay diferencia en la distribución de las variables numéricas si la persona es de sexo femenino o masculino. Este fenómeno sucede a pesar de que se tienen más datos de personas de sexo masculino lo que reafirma que el sexo no afecta variables numéricas.



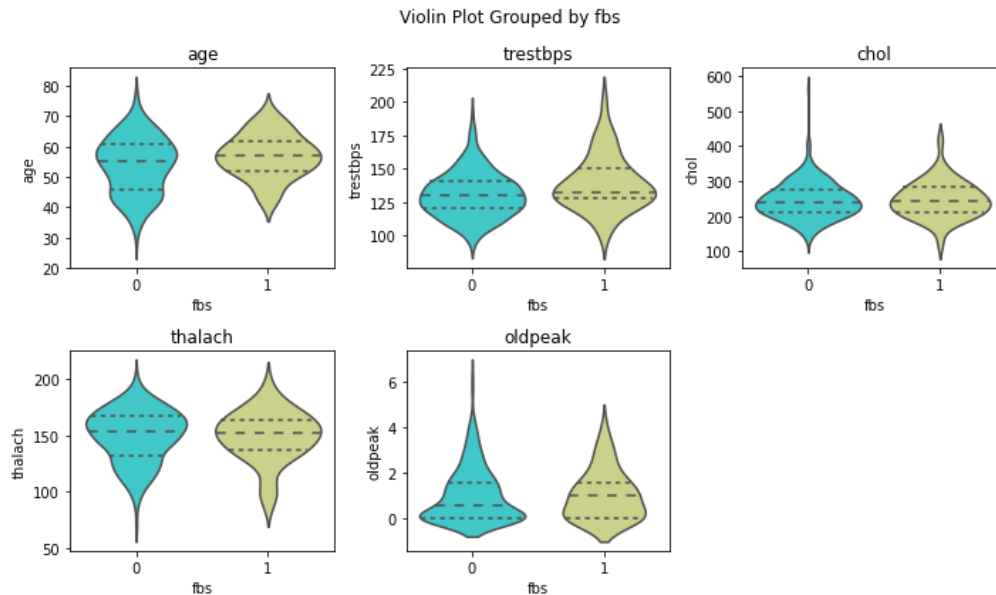
# Categóricas vs Numéricas: Tipo Dolor de Pecho

- El tipo de dolor de pecho **no afecta o se ve afectado** por ninguna variable numérica. Esto se debe a que no hay diferencia en la distribución de las variables numéricas si el dolor de pecho es de angina típica, atípica, no anginal o sin dolor. Este fenómeno sucede a pesar de que se tienen más datos de personas sin dolor de pecho o asintomáticos.



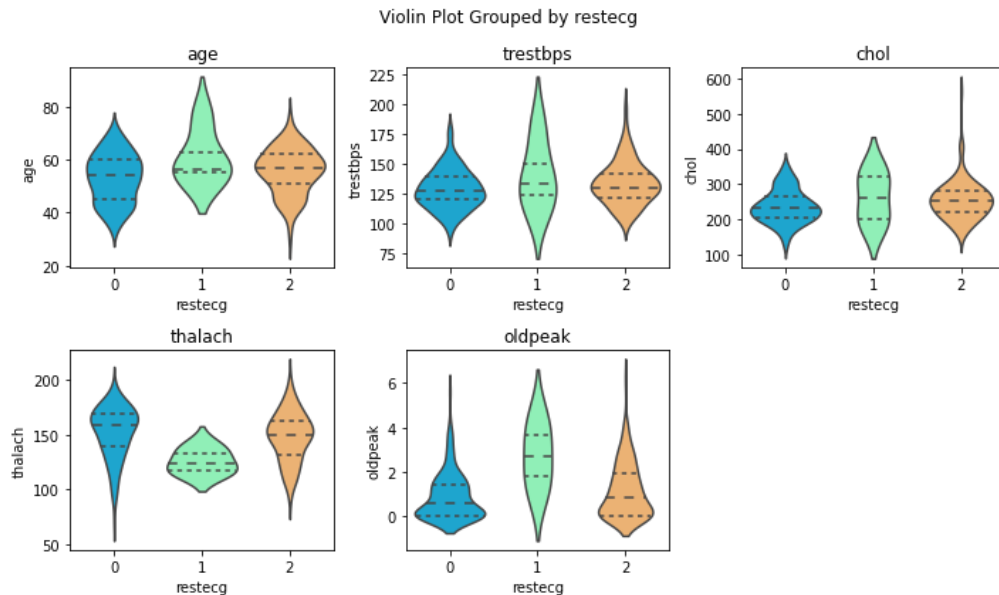
# Categorías vs Numéricas: Azúcar en sangre

Si el azúcar en sangre en ayunas es mayor a 120 mm Hg **no afecta o se ve afectado** por ninguna variable numérica. Esto se debe a que no hay diferencia en la distribución de las variables numéricas si es mayor o no el azúcar en sangre en ayunas a 120 mm Hg. Este fenómeno sucede a pesar de que se tienen más datos de personas con nivel de azúcar en sangre menor a 120 mm Hg.



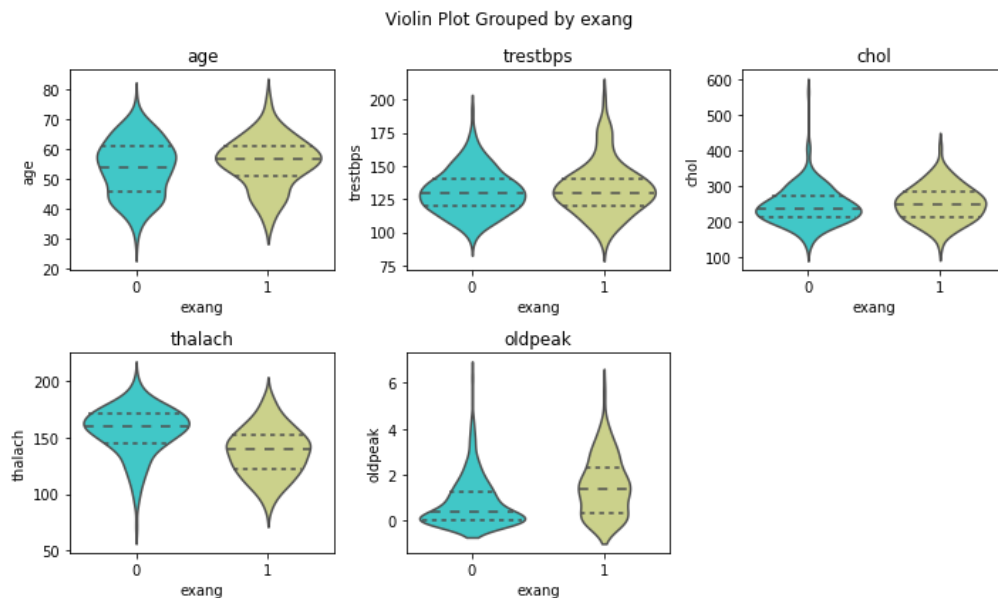
# Categorías vs Numéricas: Resultados electrocardiográficos en reposo

Los resultados electrocardiográficos en reposo si afectan o se ven afectados por las variables numéricas **thalach** y **oldpeak**. Esto dado que hay una diferencia en la distribución de estas variables numéricas si los resultados son normales, si hay anomalía en la onda ST o si hay hipertrofia ventricular. Sin embargo, esta diferencia se da cuando los resultados son de anomalía en la onda ST y esto se puede deber a que se cuenta con muy pocos datos (4 de 303) para este valor de la categórica. Por lo tanto, toca tomar estos hallazgos con cuidado.



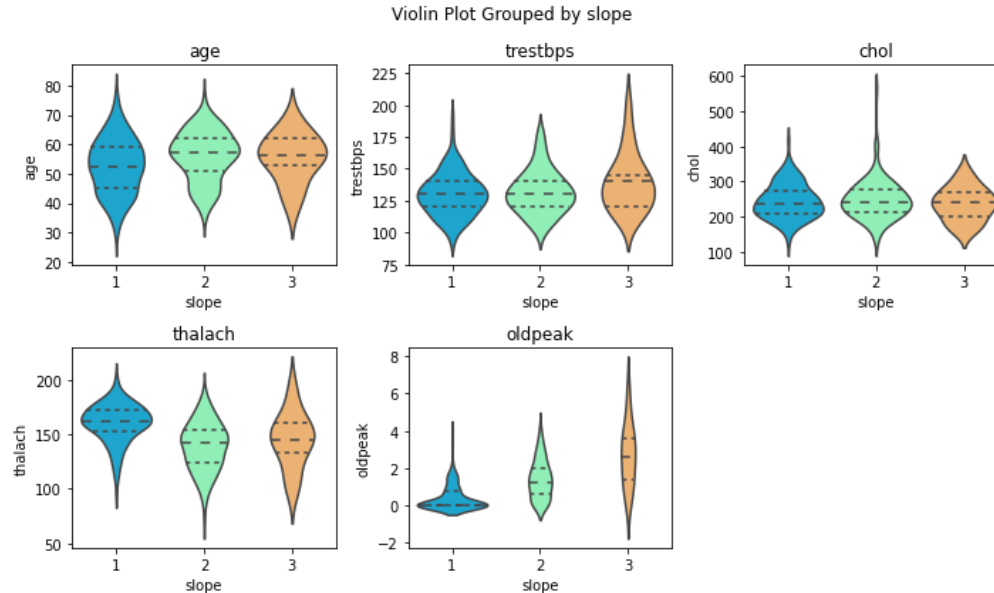
# Categóricas vs Numéricas: Angina inducida por ejercicio

Si la angina es provocada o inducida por el ejercicio si afecta o se ve afectada por las variables numéricas **thalach** y **oldpeak**. Esto dado que hay una diferencia en la distribución de esta variable numérica si la angina es provocada por el ejercicio o no.



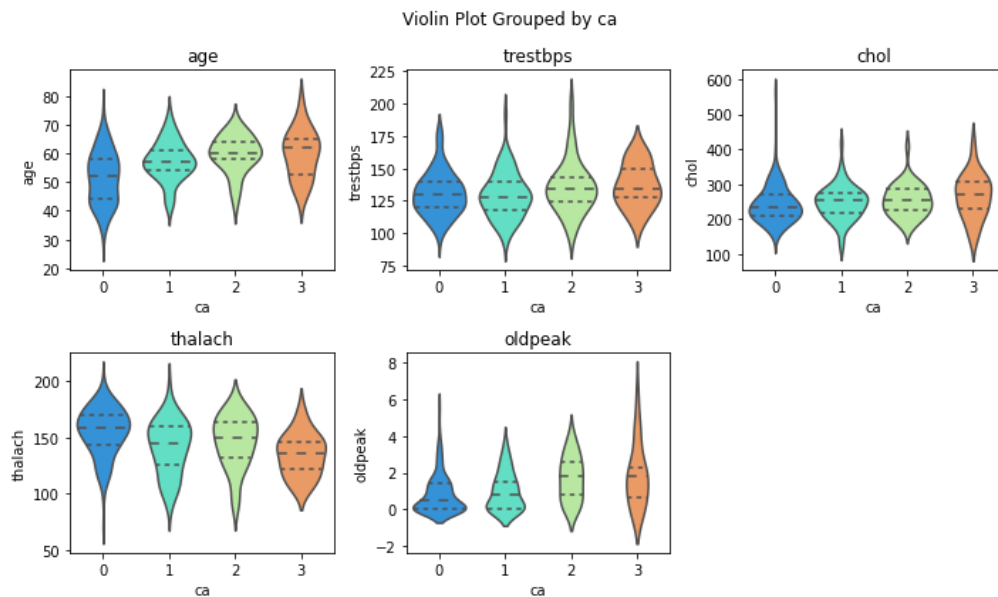
# Categorías vs Numéricas: Pendiente segmento ST

La pendiente del segmento ST en el ejercicio si afecta o se ve afectada por las variables numéricas **thalach** y **oldpeak**. Esto dado que hay una diferencia en la distribución de estas variables numéricas si la pendiente es ascendente, plana o descendente.



# Categóricas vs Numéricas: Numero vasos coloreados

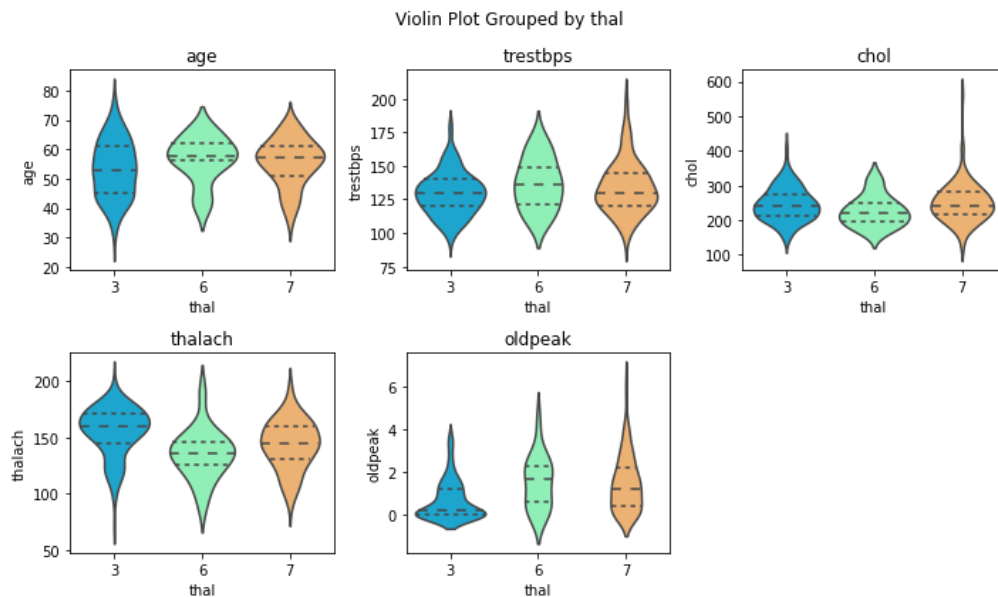
- El número de vasos coloreados por fluoroscopia si afectan o se ven afectados por las variables numéricas **age**, **thalach** y **oldpeak**. Esto dado que hay una diferencia en la distribución de estas variables numéricas si el número de vasos es 0, 1, 2 o 3.





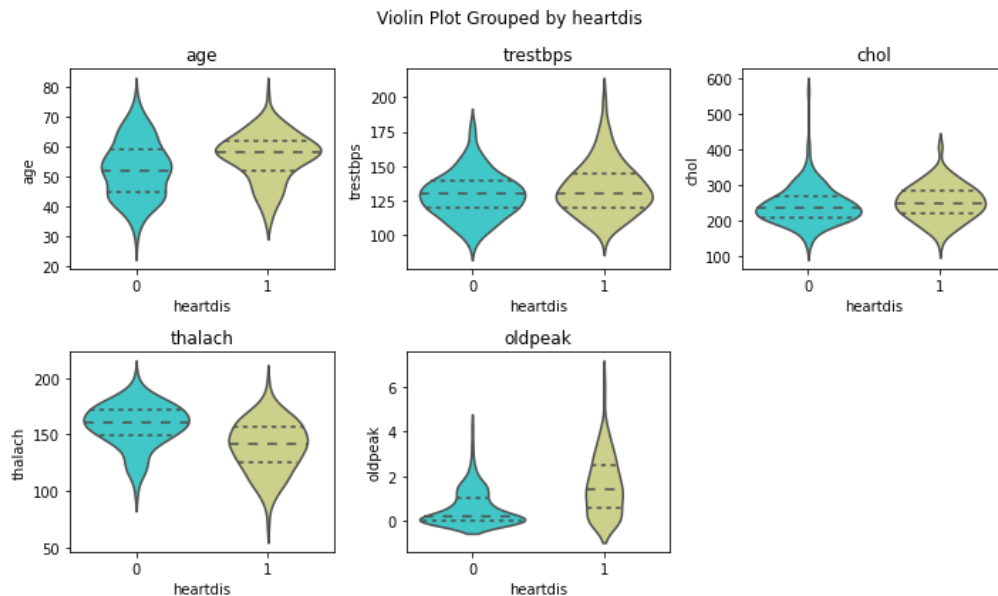
# Categóricas vs Numéricas: Talasemia

- El nivel de presencia de Talasemia si afecta o se ve afectada por la variable numérica **thalach**. Esto dado que hay una diferencia en la distribución de esta variable numérica si es normal, con defecto fijo o con defecto reversible.



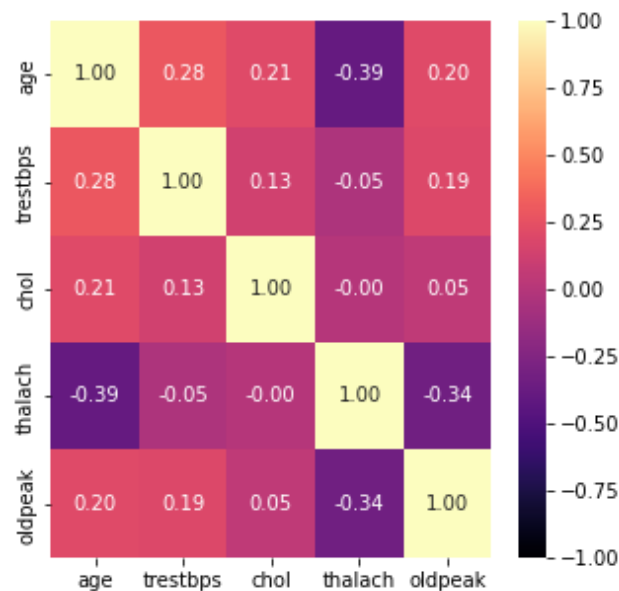
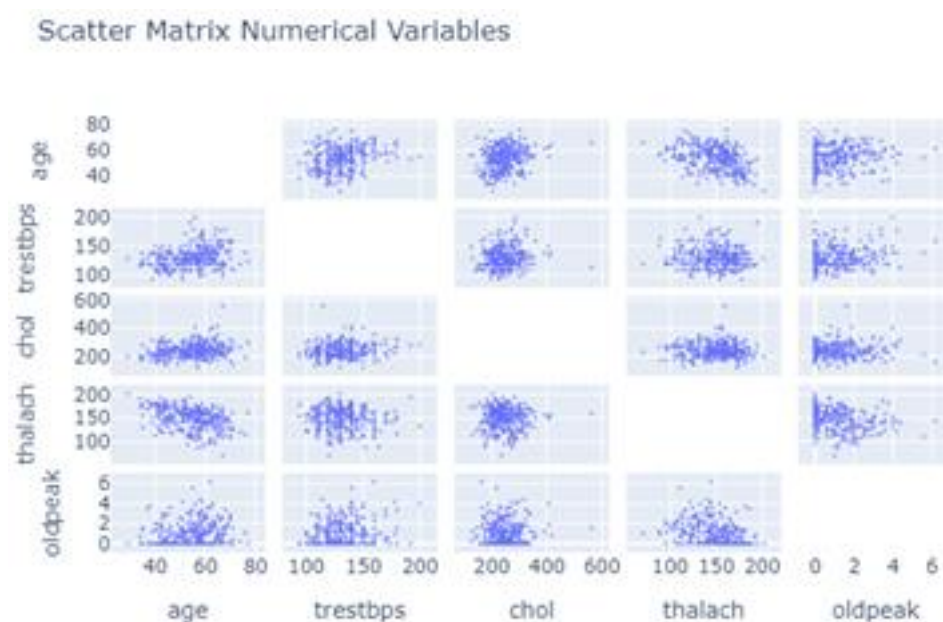
# Categorías vs Numéricas: Presencia enfermedad cardíaca

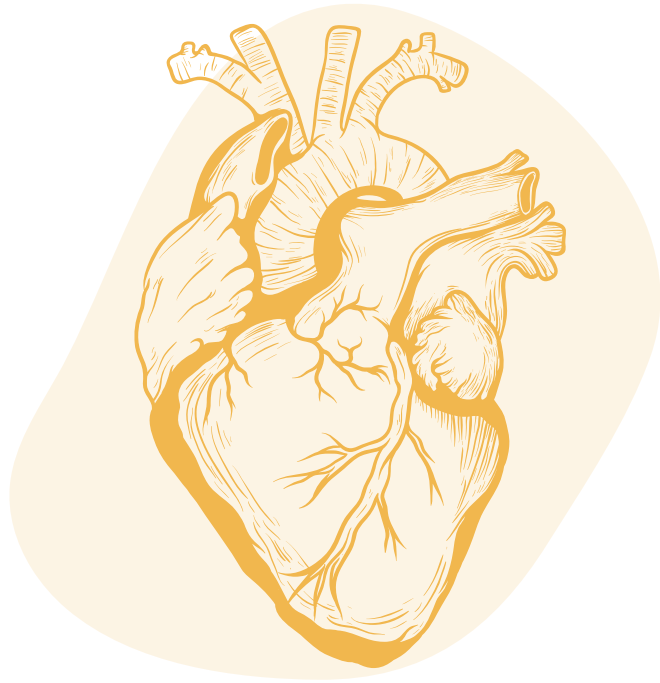
- La presencia de enfermedad cardiaca si afecta o se ve afectada por las variables numéricas **age**, **thalach** y **oldpeak**. Esto dado que hay una diferencia en la distribución de estas variables numéricas si hay presencia de enfermedad o no.



# Numéricas vs Numéricas

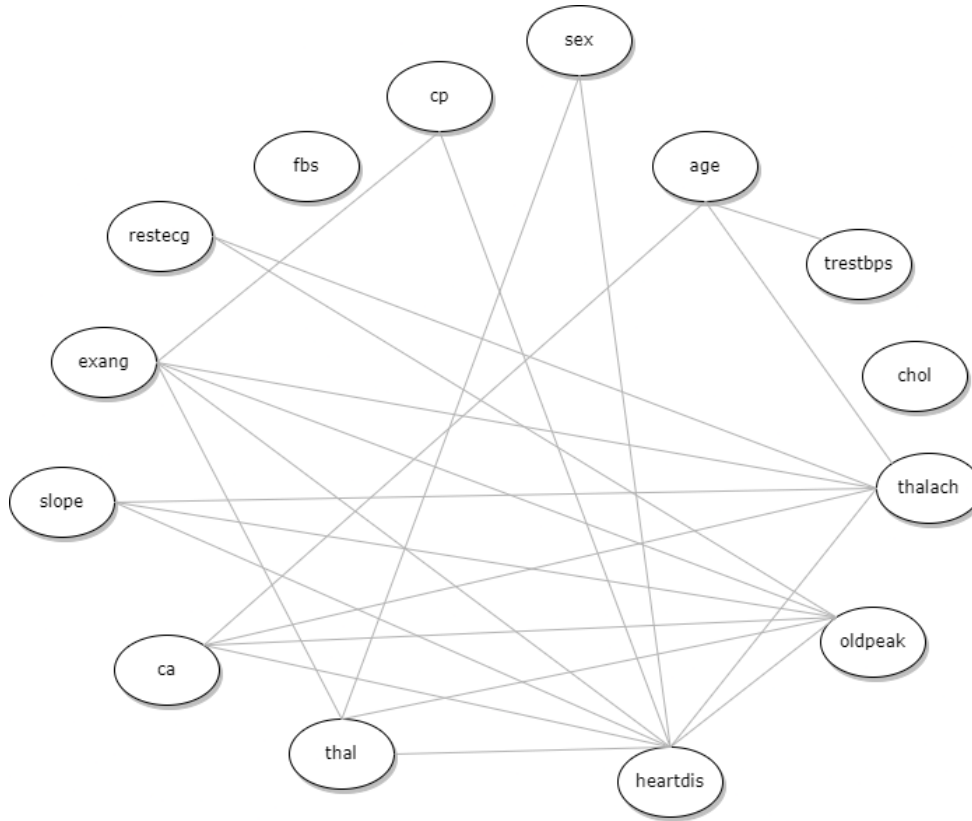
Para el análisis de variables numéricas contra variables numéricas se tomará el **coeficiente de correlación lineal** ( $\rho$ ) y **diagramas de dispersión** como base para identificar relaciones entre las variables.





# Construcción del modelo

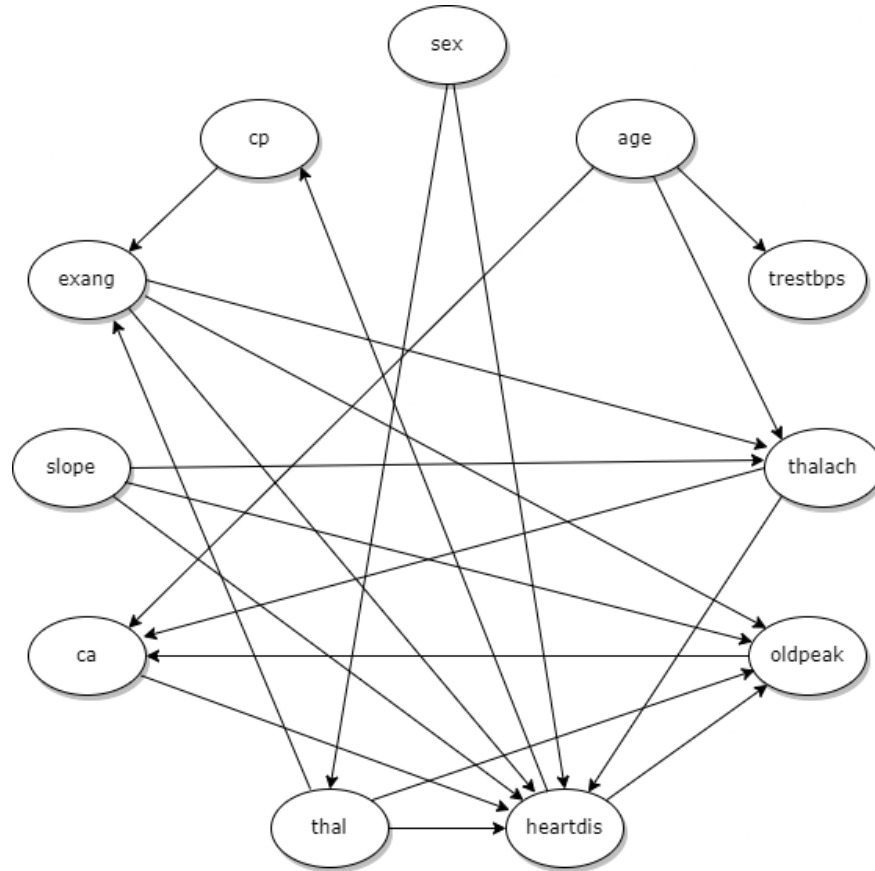
# Red Bayesiana:



Con base en el análisis de las variables del modelo se construye un grafo para representar el problema por medio de una red bayesiana.

Esta red es bidireccional por lo que necesitamos volver los arcos unidireccionales para poder construirla, nos basamos en el análisis de datos realizado anteriormente y fuentes externas.

# Red Bayesiana Unidireccional:



# Discretizar datos:

Edad:

$$(-\infty, Q_1] = 1$$

$$[Q_1, Q_2) = 2$$

$$[Q_2, Q_3) = 3$$

$$[Q_3, Q_4) = 4$$

Oldpeak:

$$(-\infty, 0.5] = 1$$

$$[0.5, 1) = 2$$

$$[1, 1.5) = 3$$

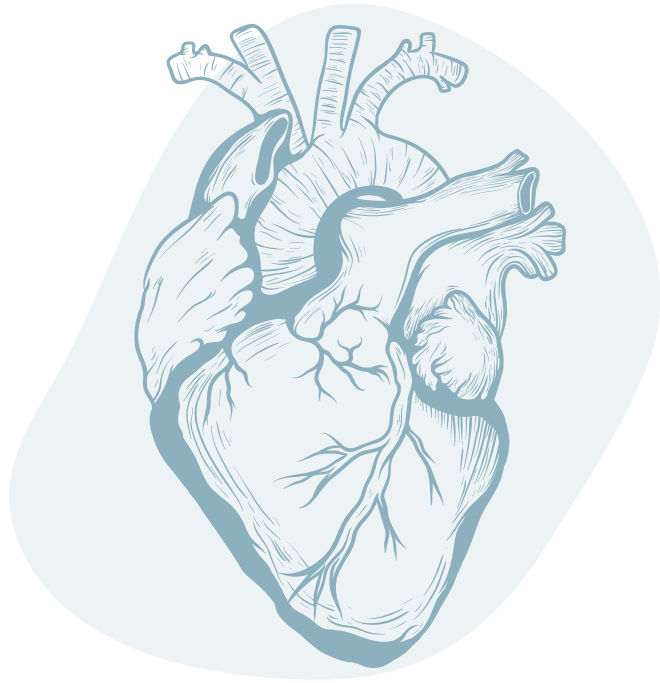
$$[1.5, 2) = 4$$

$$[2, \infty) = 5$$

Treastbps y thalach:

$$(-\infty, \bar{x}] = 1$$

$$[\bar{x}, \infty) = 2$$



# Aplicación en Dash



# Inicio

## Bienvenido al Programa de Predicción de Enfermedades Cardíacas



Las enfermedades cardíacas son una de las principales causas de muerte en todo el mundo. Estas enfermedades afectan al corazón y al sistema circulatorio, lo que puede llevar a complicaciones graves, como ataques cardíacos y accidentes cerebrovasculares.

El siguiente programa ha sido desarrollado como apoyo a médicos en el proceso de evaluación de pacientes para detectar si una persona tiene o no una enfermedad cardíaca y la toma de decisiones asociada como solicitud de exámenes, chequeos y otros procedimientos.

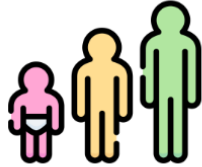
Este programa se ha desarrollado mediante la exploración de datos y el modelado con redes bayesianas, con el fin de descubrir la relación entre 14 atributos relevantes en la base de datos de enfermedades cardíacas de Cleveland. Los datos se recopilaron del repositorio de la Universidad de California en Irvine.

[Instrucciones](#)

[Programa](#)

# Instrucciones

## Instrucciones para usar el Programa



Edad



Sexo



Tipo de dolor en el pecho



Presión arterial en reposo



Colesterol sérico



Azúcar en sangre en ayunas



# Programa

## Programa de Predicción de Enfermedades Cardíacas

### Información del Paciente

#### Datos demográficos del paciente

Edad del paciente (en años):

Sexo:

Select...

#### Salud del paciente

Presión arterial (mmHg):

Frecuencia cardíaca máxima (bpm):

Colesterol sérico (mg/dL):

Azúcar en sangre en ayunas:

Tipo de dolor en el pecho:

Select...

Angina inducida por el ejercicio:

Select...

#### Resultados ECG

Resultados ECG en reposo:

Select...

Depresión del ST:

Pendiente del segmento ST:

Select...

#### Resultados de la prueba de esfuerzo nuclear.

Presencia de talasemia:

Select...

Vasos afectados:

Select...

### Riesgo previsto de enfermedad cardiaca

Seleccione los datos que conoce del paciente.

Una vez ha llenado los datos conocidos presione el botón -Predecir Modelo-.

-

Nota: No es necesario conocer todos los datos, con al menos uno bastará.

Predecir Modelo

Eliminar inputs



# Programa

## Programa de Predicción de Enfermedades Cardíacas

### Información del Paciente

#### Datos demográficos del paciente

Edad del paciente (en años):

Sexo:

#### Salud del paciente

Presión arterial (mmHg):

Frecuencia cardíaca máxima (bpm):

Colesterol sérico (mg/dL):

Azúcar en sangre en ayunas:

Tipo de dolor en el pecho:

Angina inducida por el ejercicio:

#### Resultados ECG

Resultados ECG en reposo:

Depresión del ST:

Pendiente del segmento ST:

#### Resultados de la prueba de esfuerzo nuclear.

Presencia de talasemia:

Vasos afectados:

### Riesgo previsto de enfermedad cardiaca

Dados los siguientes parámetros:  
oldpeak, con un valor de Mayor a 0.5

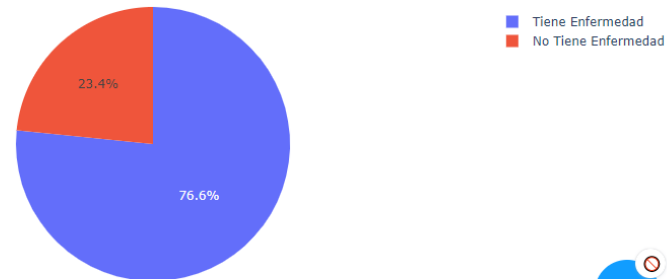
-

La probabilidad de que el paciente tenga una enfermedad cardíaca es 0.77

Predecir Modelo

Eliminar Inputs

### Probabilidad Enfermedad Cardíaca



# Conclusión:

1

La aplicación desarrollada brinda, a partir de datos ingresados por el usuario, la probabilidad de que un paciente presente una enfermedad cardíaca.

2

La aplicación sirve de apoyo para los médicos en el proceso de evaluación de pacientes para detectar si una persona presenta o no una enfermedad cardíaca y la toma de decisiones asociada como la solicitud de exámenes, chequeos y otros procedimientos.

