

Enfermedades Cardíacas: Factores de Riesgo y Variables Clínicas

CONTEXTO

Las enfermedades cardíacas son un conjunto de trastornos que afectan al corazón y los vasos sanguíneos. Según la Organización Mundial de la Salud (OMS), las enfermedades cardiovasculares son la principal causa de muerte a nivel mundial, responsables de aproximadamente el 32% de todas las muertes anuales, lo que equivale a 17.9 millones de muertes al año. De estas, se estima que más de la mitad son causadas por enfermedad arterial coronaria (EAC), que es la principal causa de infarto de miocardio, insuficiencia cardíaca y muerte súbita. La prevalencia de las enfermedades cardíacas sigue en aumento, debido a factores como el envejecimiento de la población, el sedentarismo, las dietas poco saludables y el aumento de enfermedades crónicas, como la hipertensión, la diabetes tipo y el colesterol alto.

La evaluación de la salud cardiovascular y el diagnóstico temprano de las enfermedades cardíacas requieren un análisis integral que tome en cuenta una serie de variables clínicas y de laboratorio. Algunas de estas variables, como la edad, el género, la presión sanguínea, el colesterol, la glucosa en sangre, los resultados de un electrocardiograma (ECG), las pulsaciones máximas y la presencia de angina inducida por ejercicio, son fundamentales para determinar el riesgo de desarrollar enfermedades cardiovasculares. Además, el análisis del segmento ST y su pendiente en el ECG puede proporcionar información vital sobre la presencia de isquemia miocárdica, un indicador clave de la enfermedad coronaria.

En este contexto, el objetivo de este estudio es analizar cómo diferentes variables pueden predecir la presencia o ausencia de afecciones cardíacas, así como identificar patrones que ayuden a prevenir o detectar de manera temprana enfermedades del corazón.

HIPÓTESIS

En este estudio se plantean varias hipótesis relacionadas con los factores de riesgo y los síntomas asociados a las enfermedades cardíacas, con el objetivo de entender mejor su prevalencia y los indicadores clave de estas afecciones. Las hipótesis son las siguientes:

- 1. ¿Qué género tiene más probabilidad de sufrir enfermedad cardíaca?**
En teoría, los hombres deben sufrir más enfermedades cardíacas que las mujeres. Esto se debe a factores como la mayor prevalencia de hábitos de vida poco saludables en hombres, además de diferencias hormonales.
- 2. ¿Hay relación entre edad y enfermedad cardíaca? ¿En qué edad hay mayor probabilidad de sufrir enfermedad cardíaca?**
En teoría, las personas más adultas tienen mayor probabilidad de sufrir afecciones cardíacas. A medida que avanza la edad, aumentan los factores de riesgo como la hipertensión, el colesterol elevado y la diabetes, lo que incrementa el riesgo de

desarrollar enfermedades cardíacas. Se espera que el rango de mayor probabilidad de afección esté en personas mayores de 60 años.

3. **¿Cuál es el síntoma predominante en caso de afección cardíaca?**
El síntoma predominante en caso de afección cardíaca debería ser algún tipo de dolor en el pecho, que es uno de los signos más comunes de la enfermedad arterial coronaria.
4. **¿Hay relación entre el valor ST y enfermedad cardíaca?**
El valor del segmento ST en un electrocardiograma (ECG) podría estar relacionado con la presencia de isquemia miocárdica. Aunque no se sabe con certeza la magnitud de esta relación, se espera que una alteración en el segmento ST, como un descenso o elevación, indique una mayor probabilidad de afección cardíaca, en especial en casos de enfermedad arterial coronaria.
5. **¿Hay relación entre presión sanguínea en reposo y afección cardíaca?**
En teoría, a mayor presión sanguínea en reposo, mayor será la posibilidad de sufrir una afección cardíaca. La hipertensión es un factor de riesgo conocido para enfermedades cardiovasculares, ya que aumenta la carga de trabajo del corazón y puede dañar las arterias con el tiempo.
6. **¿Hay relación entre el colesterol y afección cardíaca?**
Se espera que, a mayores niveles de colesterol, haya una mayor probabilidad de sufrir enfermedades cardíacas. El colesterol elevado contribuye a la formación de placas en las arterias, lo que puede provocar aterosclerosis y aumentar el riesgo de infarto de miocardio.
7. **¿Hay relación entre ritmo cardíaco máximo y afección cardíaca?**
En teoría, a mayor ritmo cardíaco máximo durante el ejercicio, mayor podría ser la posibilidad de sufrir una afección cardíaca. Sin embargo, este factor también depende de la capacidad de recuperación del corazón y de otros factores como la presencia de arritmias o el daño en el músculo cardíaco, por lo que la relación no es completamente lineal.

DATOS

RECOLECCIÓN

Los datos utilizados en este estudio han sido extraídos de un conjunto de datos disponible en la plataforma Kaggle.com, proporcionado por el usuario FedeSoriano. El conjunto de datos titulado "Heart Failure Prediction" (Predicción de Insuficiencia Cardíaca) está diseñado para predecir la probabilidad de que un paciente sufra insuficiencia cardíaca, basándose en diversas variables clínicas.

Este conjunto de datos contiene información de pacientes con afecciones cardíacas y abarca varias variables importantes como edad, género, presión sanguínea, colesterol, azúcar en sangre, resultados de electrocardiograma (ECG), ritmo cardíaco máximo, tipo de dolor en el pecho, entre

otras. La fuente de los datos se encuentra en el siguiente enlace: [Heart Failure Prediction Dataset - Kaggle](#).

La base de datos ha sido utilizada para analizar la relación entre los diferentes factores de riesgo y la probabilidad de que un paciente sufra enfermedades cardíacas. Se trata de un conjunto de datos tabular que contiene registros de características clínicas de pacientes, las cuales son fundamentales para la predicción y diagnóstico de enfermedades cardiovasculares.

El proceso de recolección de estos datos implicó la descarga directa desde Kaggle, donde el archivo fue limpiado y preprocesado para garantizar que los valores faltantes se manejaran correctamente y que los datos fueran adecuados para el análisis posterior.

LIMPIEZA Y PREPARACIÓN DE LOS DATOS

Durante el proceso de limpieza y preparación de los datos, se realizaron varios pasos para garantizar la calidad y la precisión de la información antes de su análisis.

1. **Análisis de valores atípicos:** En un primer análisis exploratorio de los datos, se identificaron valores inusuales en las variables colesterol y presión sanguínea en reposo, donde algunos registros tenían un valor de 0. Dado que estas mediciones no pueden ser 0 en la práctica, se considera reemplazarlos con la mediana.
2. **Comprobación de valores nulos:** Se verificó que no hubiera valores nulos en el conjunto de datos. Esta revisión es esencial para garantizar la integridad de la información, y en este caso, no se encontraron valores faltantes en ninguna de las variables.
3. **Verificación de duplicados:** Se realizó una revisión exhaustiva para detectar posibles registros duplicados. Afortunadamente, no se encontraron duplicados en los datos, lo que asegura que cada paciente esté representado solo una vez en el conjunto de datos.
4. **Conversión de variables:** Para mejorar el análisis posterior, las variables azúcar en sangre en reposo y afección cardíaca fueron convertidas a tipo categórico en lugar de numérico. Esta conversión facilita el análisis y la interpretación de estos datos, ya que la afección cardíaca tiene valores binarios (presencia o ausencia), y el azúcar en sangre en reposo se maneja de forma categórica para clasificar los niveles en categorías significativas (inferior a 120 o superior a 120).

Con estas etapas de limpieza y preparación, los datos quedaron listos para ser analizados de manera efectiva, asegurando su calidad y fiabilidad en los análisis posteriores.

En la siguiente tabla se resumen las principales características de las variables del dataframe final que se usa como base del análisis:

Columna/Variable	Descripción	Tipo variable	de Importancia inicial
Age	Age of the patient (years)	Numérica discreta	0
Sex	Sex of the patient (M: Male, F: Female)	Binaria	0
ChestPainType	Chest pain type (TA: Typical Angina, ATA: Atypical Angina, NAP: Non-Anginal Pain, ASY: Asymptomatic)	Categórica nominal	0
RestingBP	Resting blood pressure (mm Hg)	Numérica discreta	0
Cholesterol	Serum cholesterol (mm/dl)	Numérica discreta	0
FastingBS	Fasting blood sugar (1: if FastingBS > 120 mg/dl, 0: otherwise)	Binaria	0
RestingECG	Resting electrocardiogram results (Normal: Normal, ST: having ST-T wave abnormality, LVH: showing left ventricular hypertrophy)	Categórica nominal	1
MaxHR	Maximum heart rate achieved (Numeric value between 60 and 202)	Numérica discreta	0
ExerciseAngina	Exercise-induced angina (Y: Yes, N: No)	Binaria	0
Oldpeak	Oldpeak = ST (Numeric value measured in depression)	Numérica discreta	2
ST_Slope	The slope of the peak exercise ST segment (Up: upsloping, Flat: flat, Down: downsloping)	Categórica nominal	2
HeartDisease	Output class (1: heart disease, 0: Normal)	Binaria Directora	0

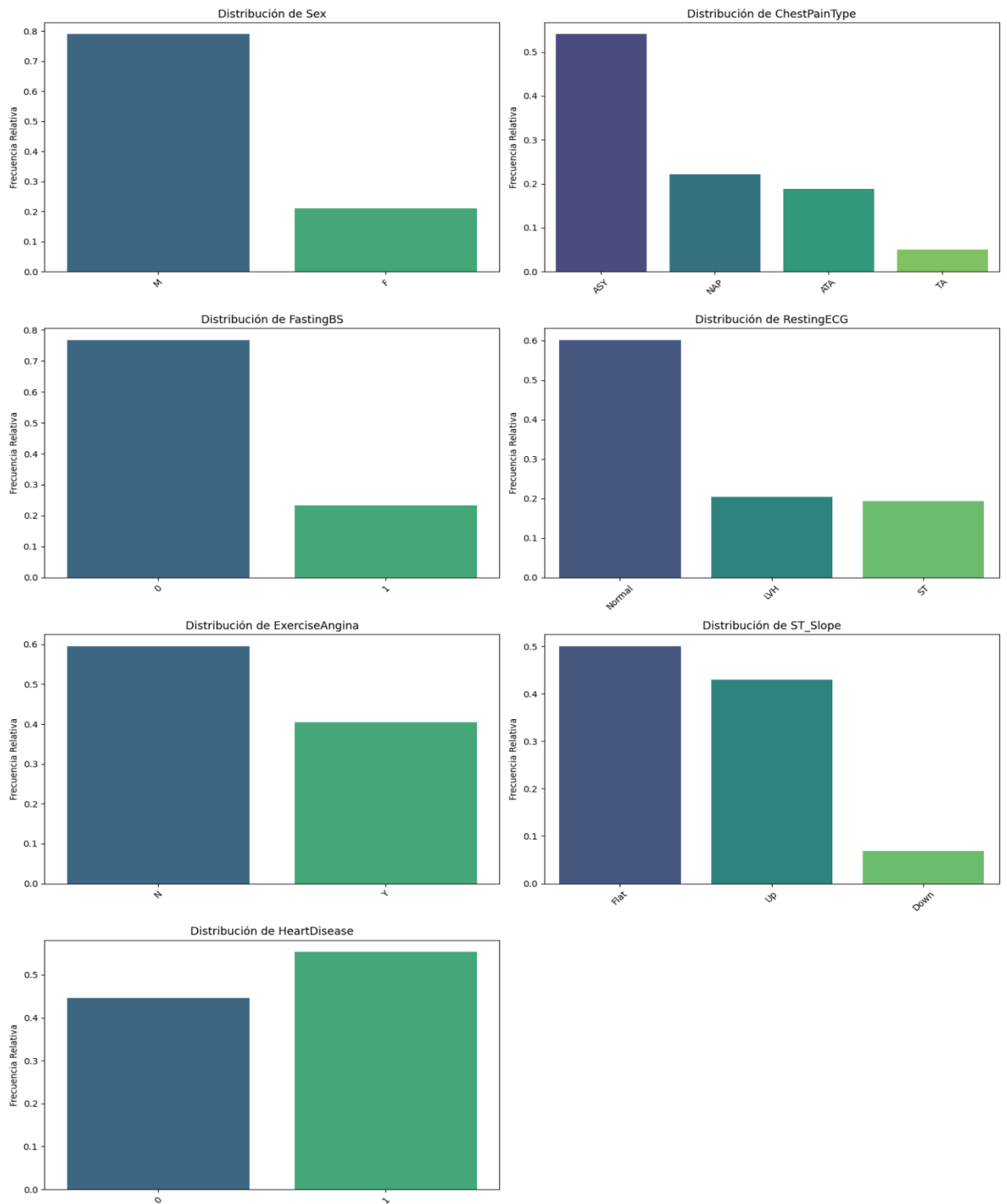
ANÁLISIS

UNIVARIANTE

Una vez realizado el análisis de tendencia central, se puede observar que las medias y las medianas de las variables "Age", "RestingBP", "MaxHR" y "Oldpeak" son muy similares, por lo que se podría indicar que los datos son bastante simétricos. Sin embargo, en la variable "Cholesterol" la media y mediana difieren más, por lo que los datos no son simétricos.

Variable	Media	Mediana
Age	53	54
RestingBP	132	130
Cholesterol	198	223
MaxHR	136	138
Oldpeak	0,88	0,6

A continuación, se realiza el análisis de frecuencias relativas para las variables categóricas:

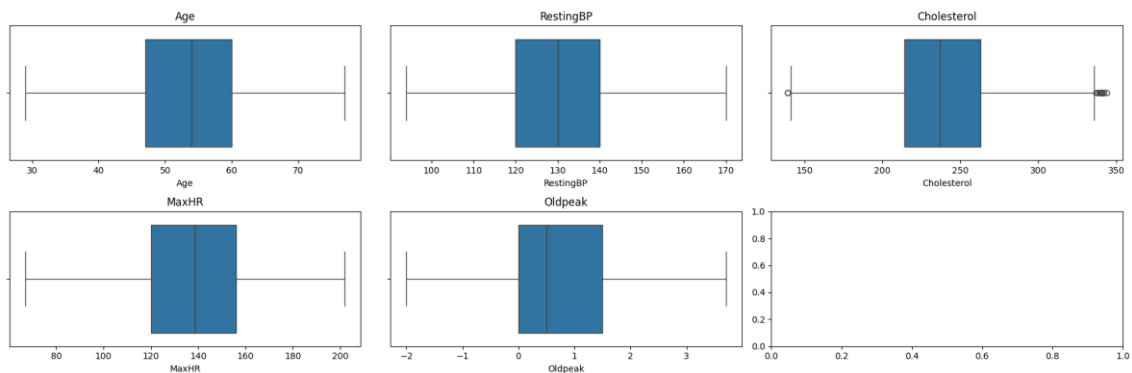


De este análisis de frecuencias relativas se puede extraer lo siguiente:

- La proporción de hombres es mayor que de mujeres (casi 80% vs 20%)
- La proporción de asintomáticos es bastante importante (55%)
- El azúcar en sangre en ayunas por debajo de 120 prevalece por mucho (casi 80%)
- ECG en reposo es normal en la gran mayoría de los casos (60%)

- Prevalece que no haya angina inducida por el ejercicio (60%)
- La pendiente ST está repartida principalmente entre plano (50%) y alto (45%)
- Las afecciones están repartidas (55% sí que ha tenido problema cardíaco vs 45% que no)

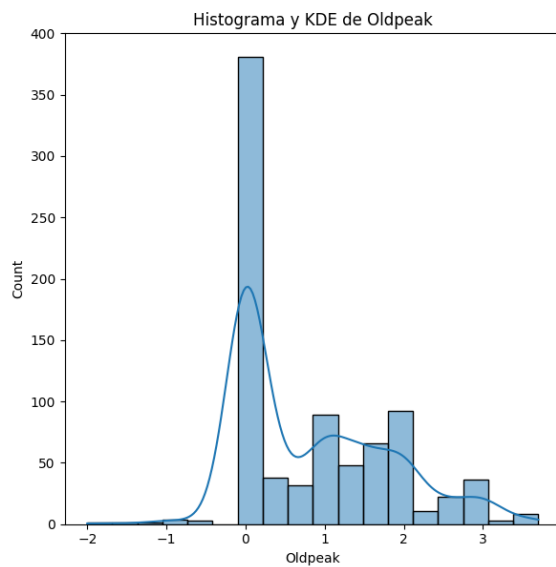
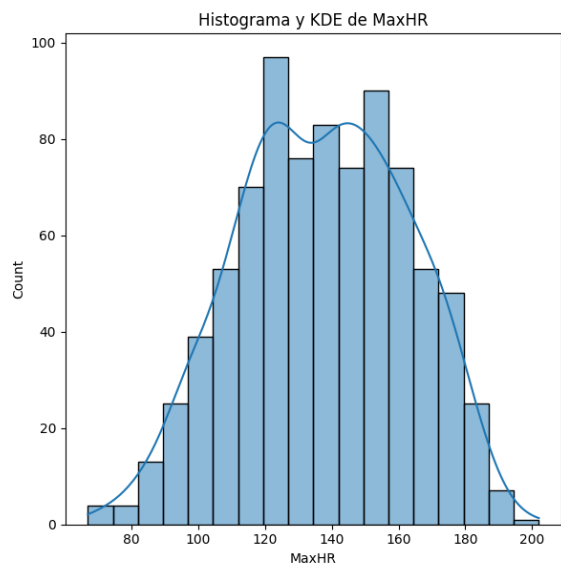
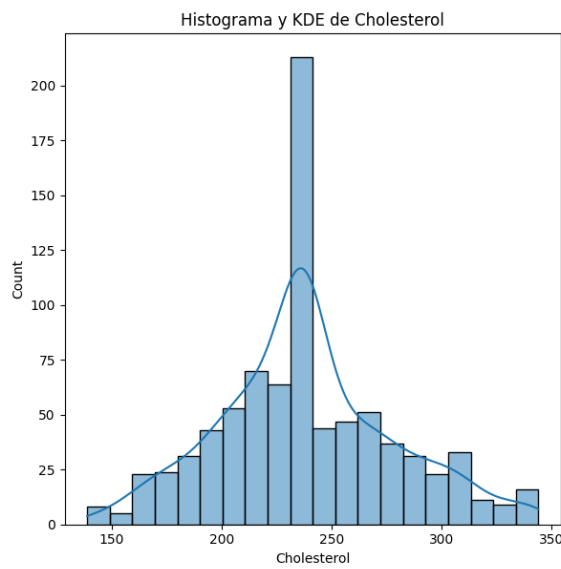
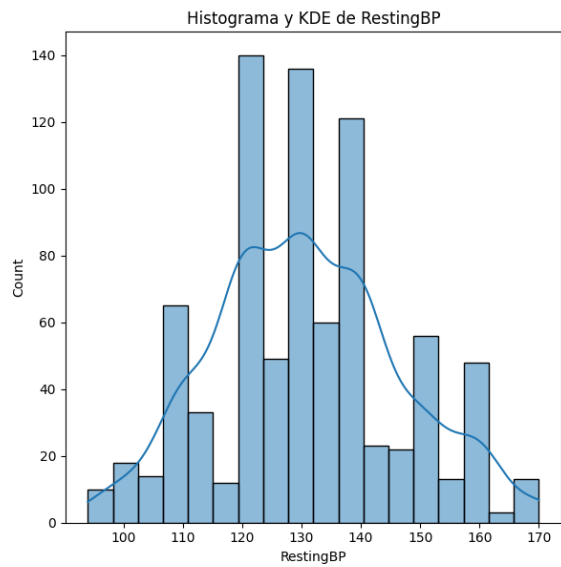
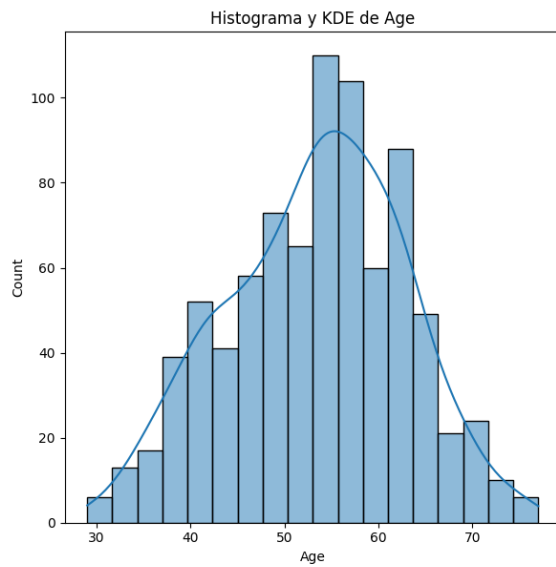
La obtención de los *boxplot* nos permite visualizar rápidamente la existencia de valores anómales o “*outliers*” en las variables “*RestingBP*”, “*Cholesterol*” y “*MaxHR*”. En el ámbito de la salud se suele indicar como valor 0 a aquellas medidas de las que no se dispone de dato. En este ámbito se suele modificar el valor de 0 por la mediana. Esto es lo que vamos a hacer para la variable colesterol, mientras que el resto de *outliers* los vamos a eliminar. Los *boxplot* una vez realizado todo lo anterior quedarían de tal forma:



Y podemos comentar lo siguiente:

- Se siguen apreciando outliers en la variable “*Cholesterol*”, pero en este caso por el momento los mantenemos
- En la mayoría de casos se ve cierta simetría de los datos dentro del rango intercuartílico (las cajas a cada lado de la mediana son bastante similares). Donde menos se aprecia sería en “*Oldpeak*”

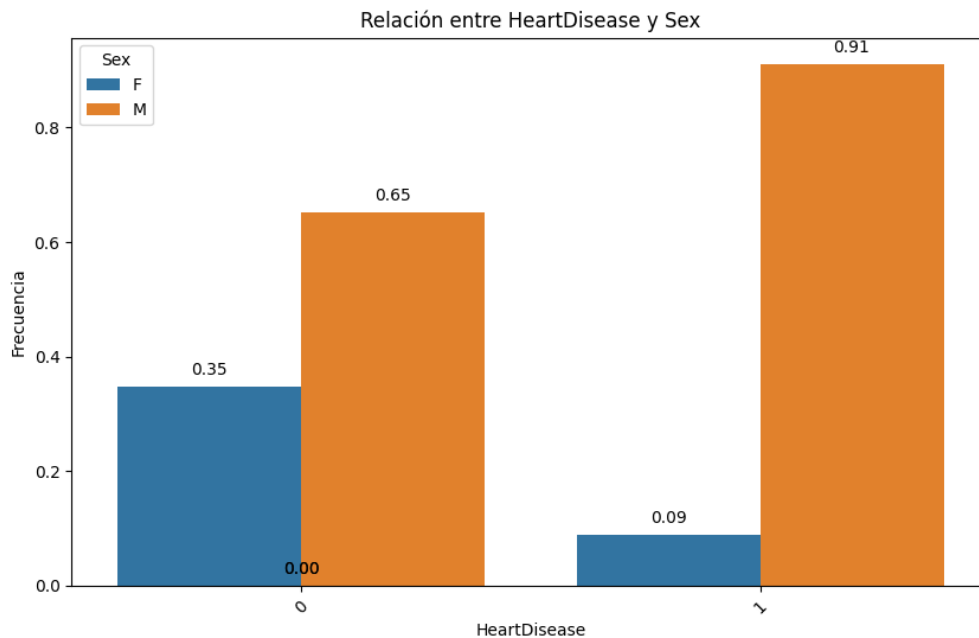
El siguiente paso ha sido comprobar las distribuciones de las variables numéricas y su función de densidad de probabilidad:



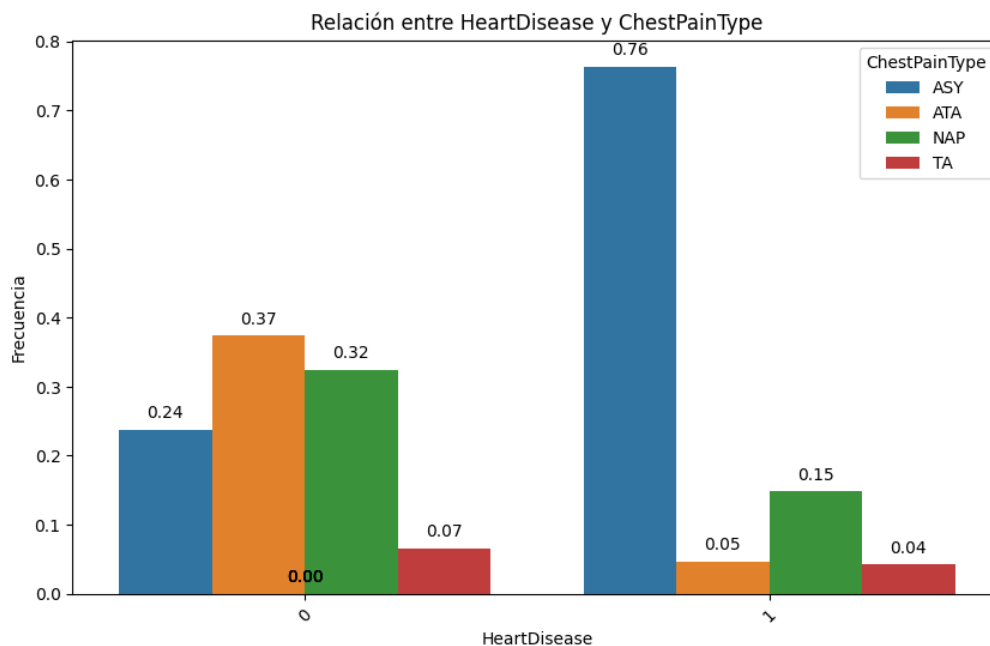
Como se puede observar en la imagen superior, las distribuciones no se asemejan a una normal. Sin embargo, esto se ratifica con los tests de Saphiro y de Anderson, donde ambos indican la no asemejanza de las variables a una normal.

BIVARIANTE

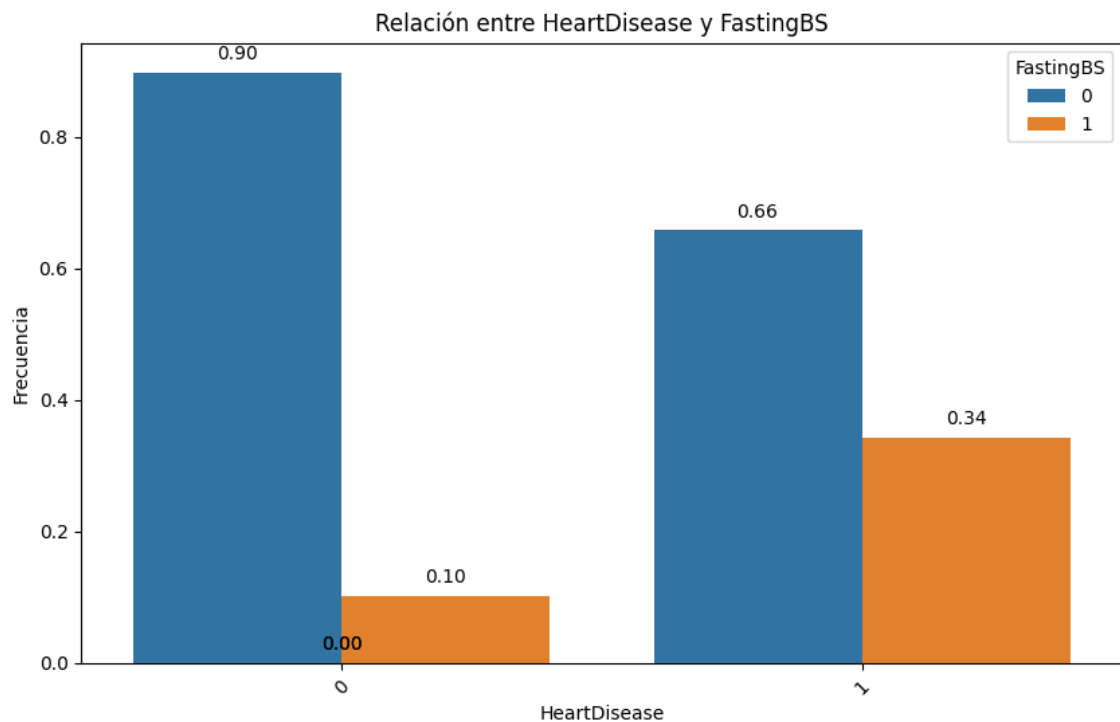
En primer lugar, se realiza el análisis bivalente de la columna categórica directora “HeartDisease” con el resto de variables categóricas. Se muestran los resultados obtenidos más abajo:



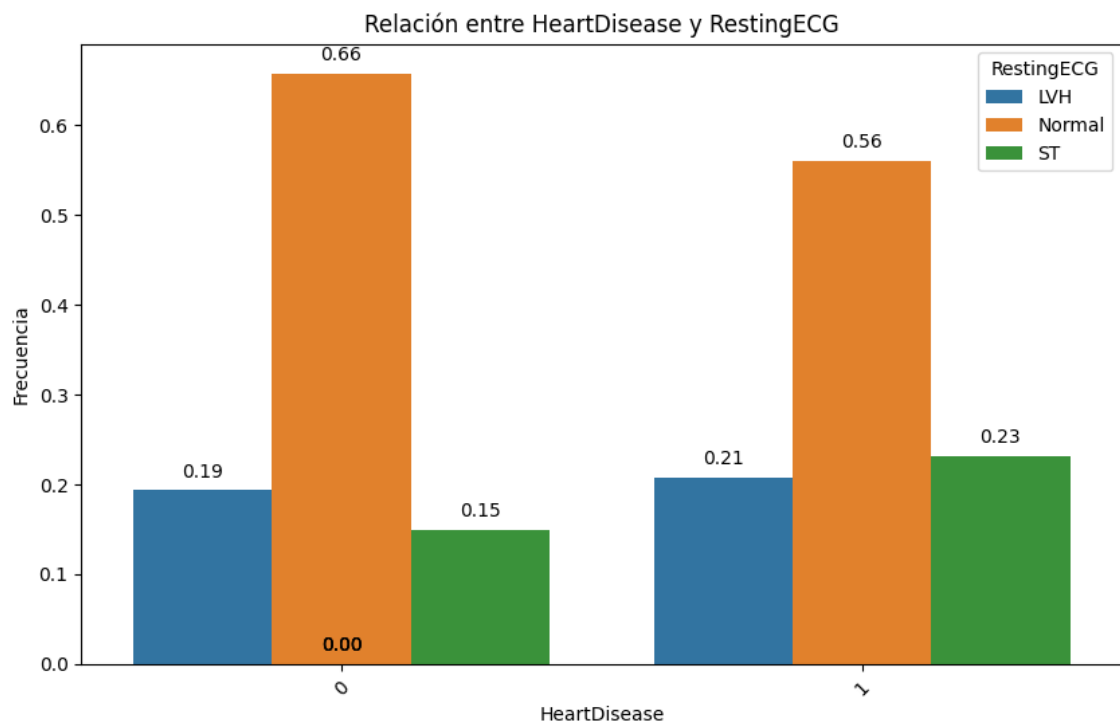
- Si bien la proporción de hombres vs mujeres de la muestra es de un 80-20%, se puede comprobar que el 91% de los hombres han sufrido una enfermedad cardíaca vs 9% de las mujeres analizadas. De todas formas, esto no se puede generalizar sin una muestra de mujeres mayor



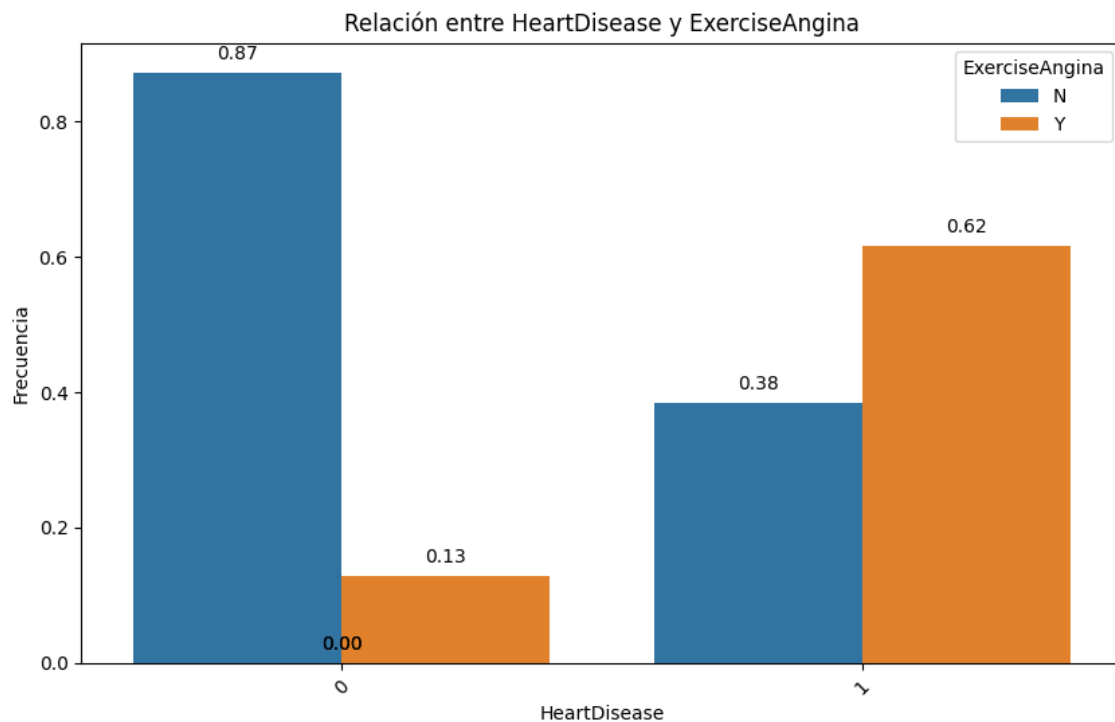
- El 76% de la muestra que ha sufrido una enfermedad cardíaca no ha tenido síntomas



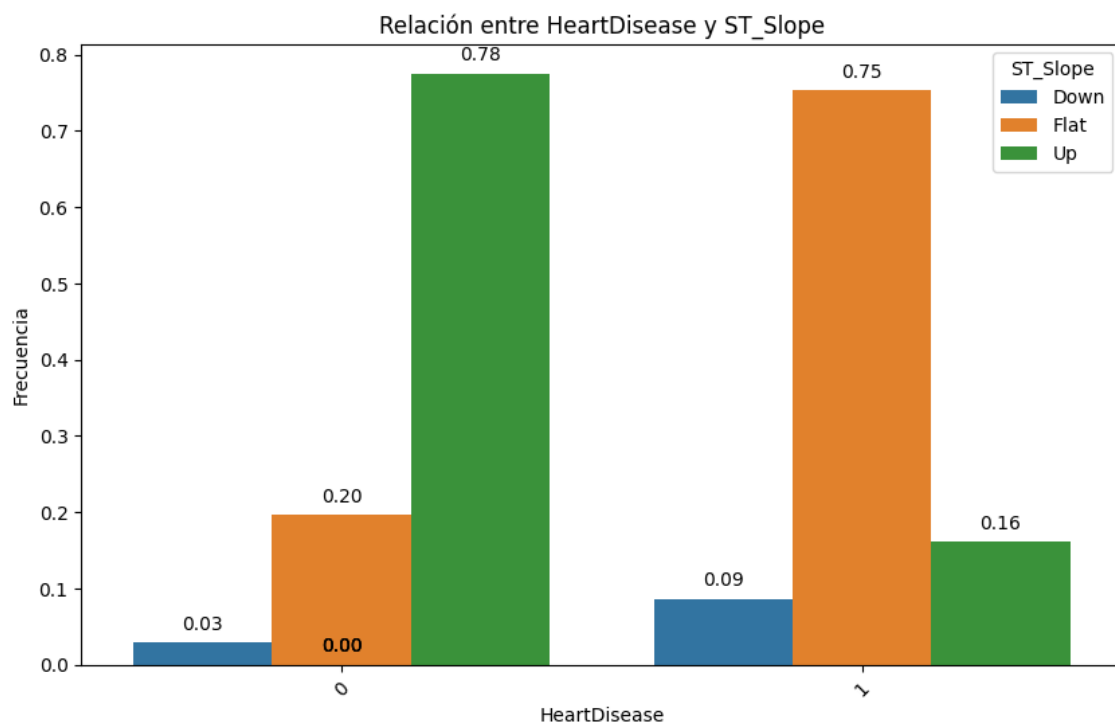
- El azúcar en sangre en ayunas suele estar por debajo de 120 (salida variable = 0) tanto en los casos en que no hay enfermedad cardíaca como en los que sí que la hay. Sí que cabe destacar que en los casos que hay enfermedad cardíaca los casos de azúcar en sangre aumentan (34%)



- ECG en reposo no veo que haya mucha relación. En ambos casos predomina el "Normal"

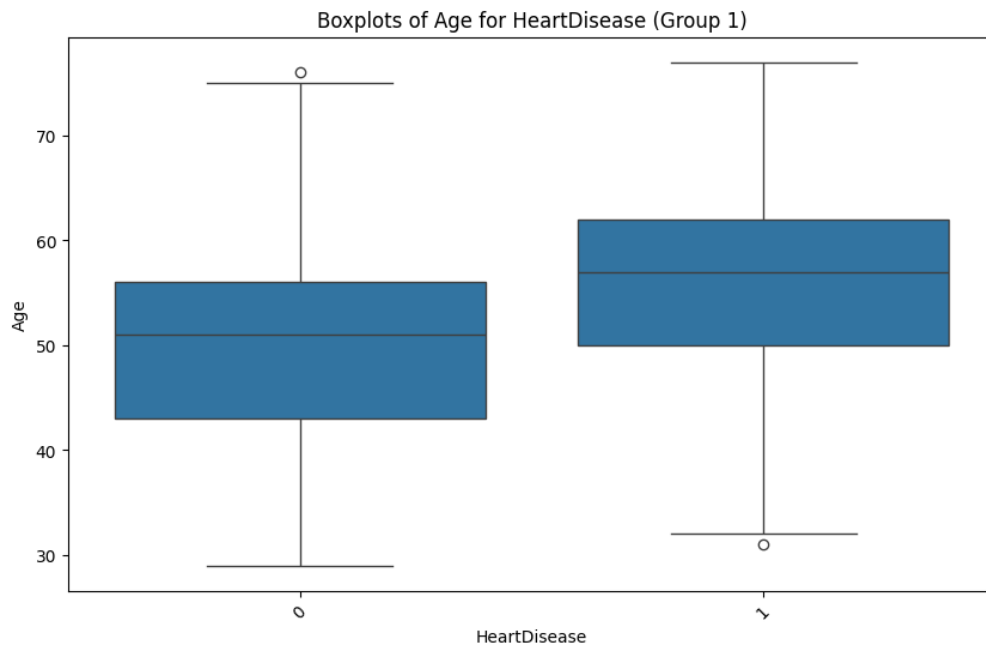


- En los casos de enfermedad cardíaca, sí que predomina la angina inducida por el ejercicio. Esto podría explicar el segundo punto, en el que la muestra que ha sufrido enfermedad cardíaca no ha tenido síntomas. Haremos un multivariante posteriormente

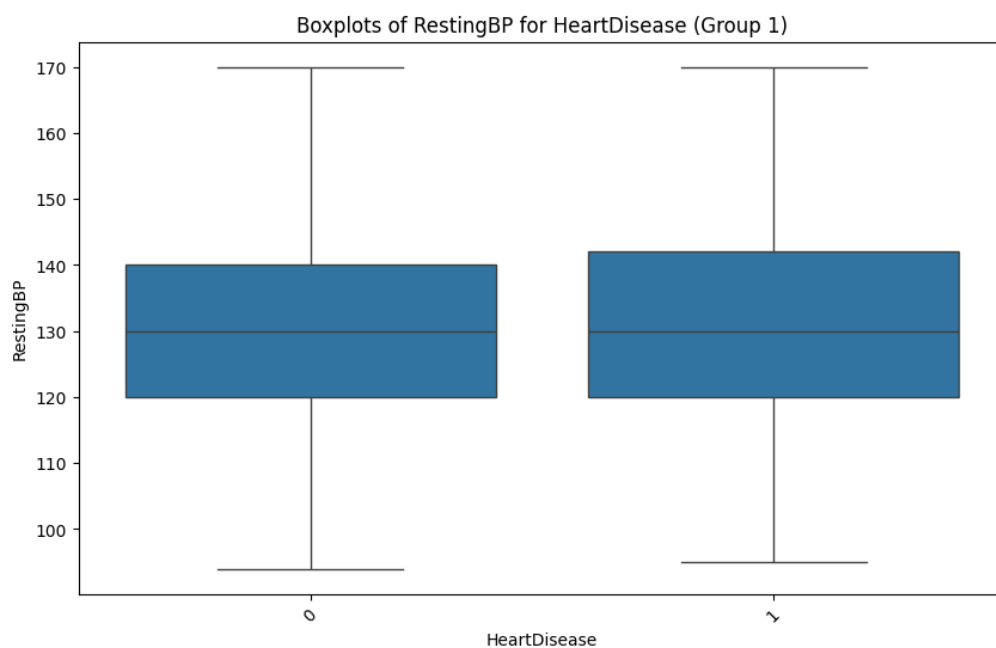


- Interesante también que pasamos de una pendiente "alta" en los casos en que no hay enfermedad cardíaca a "plana" en los casos en que sí que la hay

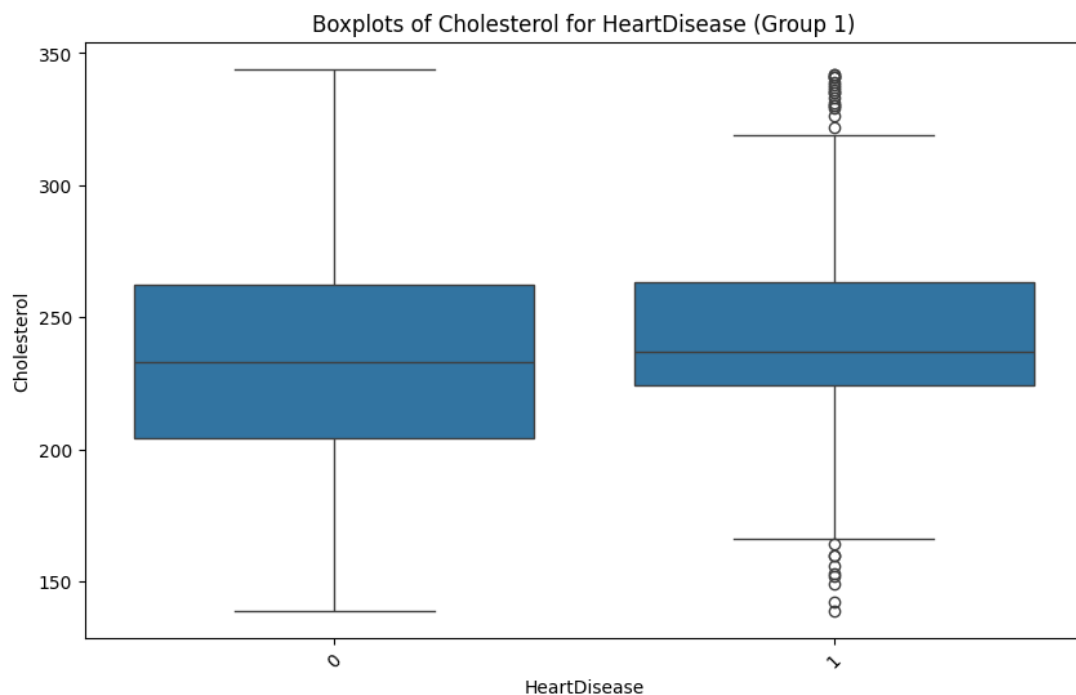
Tras lo obtenido en el análisis de frecuencias anterior, realizamos el análisis categórico-numérico de la variable directora “HeartDisease” con las diferentes variables numéricas. Se realiza el análisis de frecuencias relativas y distribución/función de densidad de probabilidad, pero lo que vamos a mostrar en las siguientes líneas son los *boxplot*:



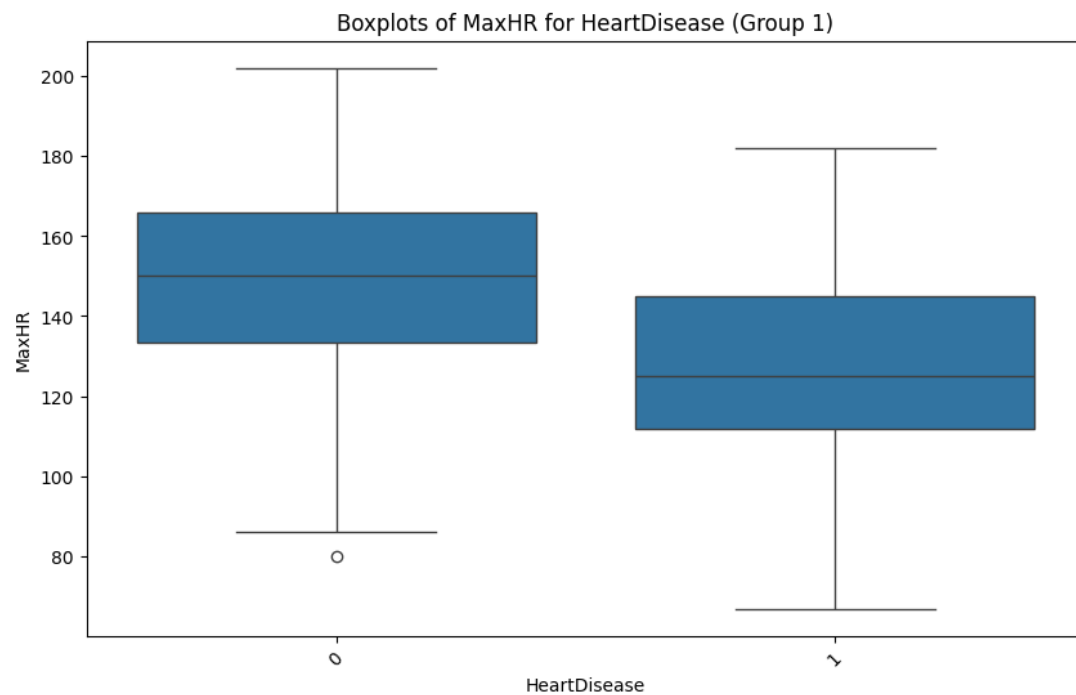
- Las enfermedades cardíacas ocurren en mayor medida a una edad de unos 58 años. Además, el IQR está en una posición mayor en este caso frente a NO enfermedad cardíaca, lo que podría indicar que a una edad mayor hay más riesgo de sufrir enfermedad cardíaca



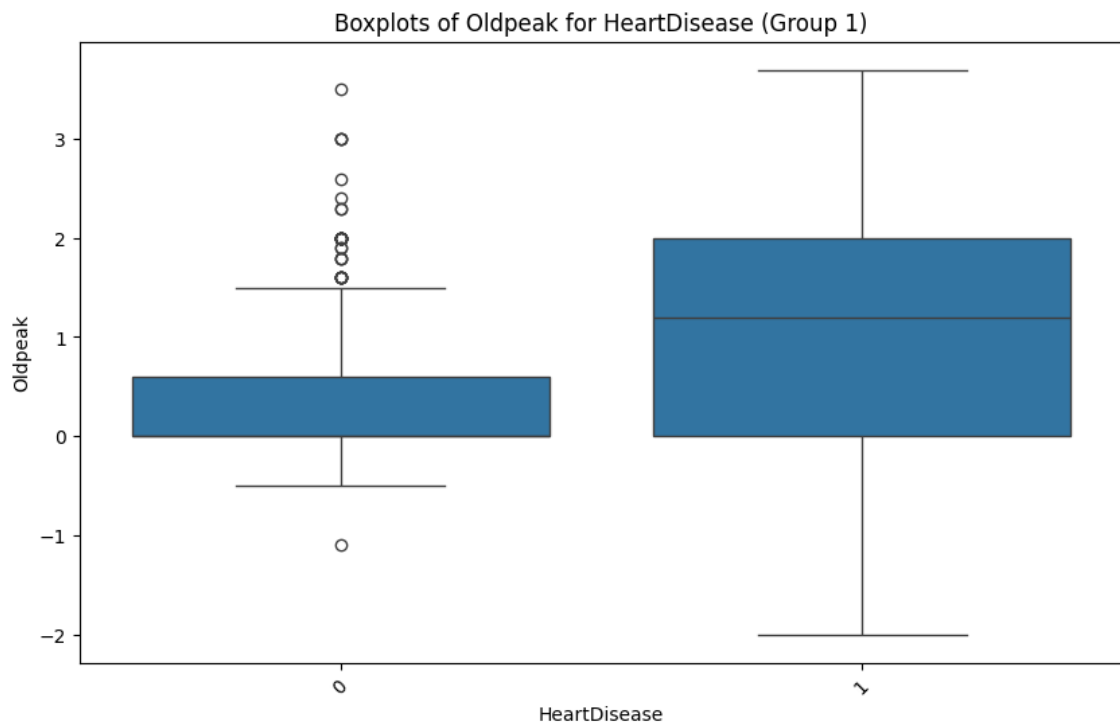
- La presión arterial en reposo es la misma tanto si se ha tenido o no enfermedad cardíaca (en torno a 130)



- La mediana del colesterol es la misma tanto si se ha tenido o no enfermedad cardíaca (225). Eso también nos podría indicar que a un colesterol mayor (225) hay más riesgo de que haya enfermedad cardíaca. Además, el IQR es menor en el caso de afección cardíaca lo que podría reafirmar esta posición



- Las máximas pulsaciones sean inferiores si se ha tenido enfermedad cardíaca que si no se ha tenido

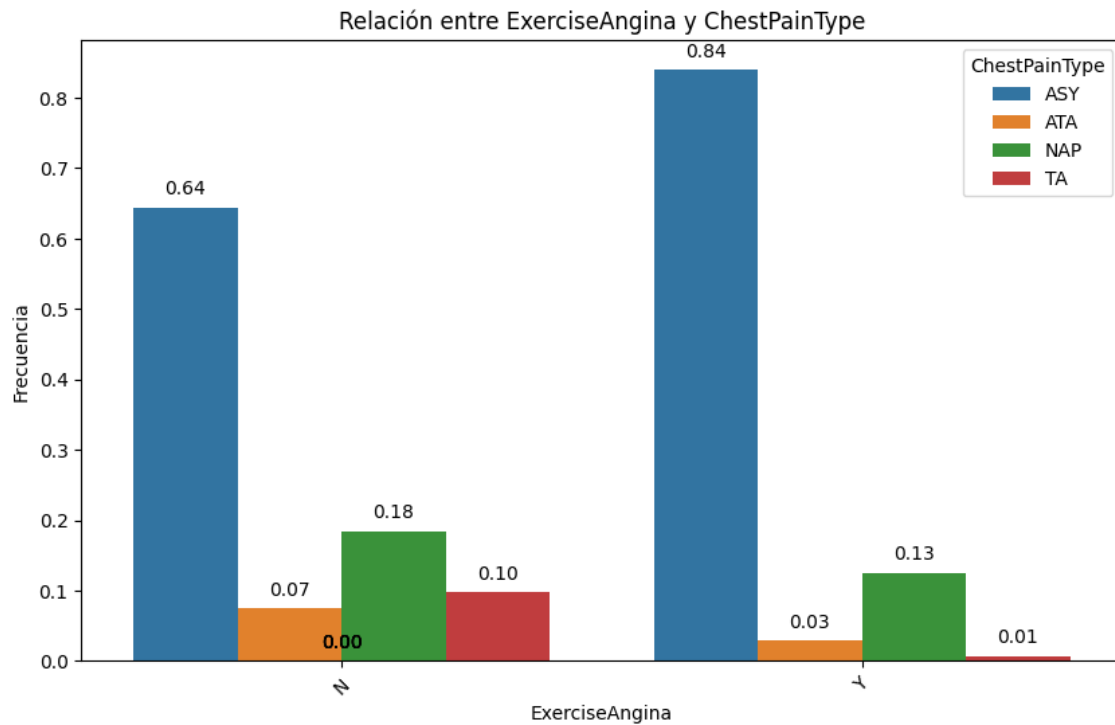


- El IQR de "*Oldpeak*" en los casos en que ha habido enfermedad cardíaca es mucho más amplio y con un percentil 75% de casi 2. Esto podría indicar que el tener depresión (y dependiendo del grado) podría afectar de forma considerable a sufrir una afección cardíaca

MULTIVARIANTE

Finalmente se ha realizado un análisis multivariante entre la variable categórica directora “HeartDisease”, “ChestPainType” y “ExerciseAngina” para comprobar lo comentado en apartados previos y ver si hay algún tipo de relación entre ellas:

Respuesta HeartDisease = 1:



- Se observa la prevalencia de los casos asintomáticos tanto si la angina se induce mediante el ejercicio como si no es así. En el caso de que sí que lo induzca, el 84% de la muestra no presenta síntomas

CONCLUSIONES Y RECOMENDACIONES BASADAS EN LOS HALLAZGOS

- Género y enfermedad cardíaca:** Aunque los datos disponibles presentan una desproporcionada representación de hombres frente a mujeres (80% vs. 20%), se observa que los hombres tienen una probabilidad significativamente mayor de sufrir enfermedades cardíacas (91% frente a 9%). Este hallazgo podría ser una base para futuras investigaciones, donde se debería equilibrar la recopilación de datos entre ambos géneros para obtener conclusiones más concluyentes y explorar mejor las diferencias entre hombres y mujeres en cuanto a las afecciones cardíacas.
- Edad y enfermedad cardíaca:** Existe una clara relación entre la edad y el riesgo de enfermedad cardíaca. A medida que aumenta la edad, también lo hace la probabilidad de sufrir afecciones cardíacas. El análisis revela que el percentil 25% se encuentra en los

50 años, la mediana en 58 años y el 75% en 62 años, lo que sugiere que las enfermedades cardíacas afectan principalmente a personas mayores de 50 años. Este patrón subraya la importancia de realizar chequeos periódicos a medida que se envejece para una detección temprana.

3. **Síntoma predominante en afección cardíaca:** Sorprendentemente, el síntoma predominante en los casos de afección cardíaca es la **asintomática** (76%), lo que subraya la complejidad del diagnóstico de enfermedades cardíacas, ya que muchas personas pueden estar en riesgo sin experimentar síntomas evidentes. Además, se observa que el ejercicio físico actúa como un importante desencadenante de la afección cardíaca, ya que el **62% de los pacientes** con enfermedad cardíaca indican el ejercicio como factor desencadenante. Curiosamente, el **84% de estos pacientes** no presentan síntomas, lo que refuerza la importancia de las pruebas de esfuerzo para detectar afecciones cardiovasculares en etapas tempranas.
4. **Relación entre el segmento ST y enfermedad cardíaca:** Se ha identificado una relación significativa entre el valor del segmento ST en el electrocardiograma (ECG) y la presencia de enfermedad cardíaca. Los pacientes sin afección cardíaca presentan un rango intercuartílico (IQR) del segmento ST entre 0 y 0.5, mientras que los pacientes con afección cardíaca tienen un IQR mucho mayor, entre 0 y 2. Además, la pendiente del segmento ST es **creciente** en los pacientes sin afección cardíaca, mientras que en los pacientes con afección, esta pendiente se vuelve **plana**, lo que indica posibles alteraciones en la circulación coronaria y la presencia de isquemia.
5. **Presión sanguínea en reposo y enfermedad cardíaca:** Aunque en un primer análisis no se observa una relación clara entre la presión sanguínea en reposo y la afección cardíaca, es necesario profundizar más en esta variable en futuros estudios. La presión alta es un conocido factor de riesgo cardiovascular, y por lo tanto, este tema debería ser explorado en más detalle para obtener conclusiones definitivas.
6. **Colesterol y enfermedad cardíaca:** Existe una relación positiva entre los niveles de colesterol y la probabilidad de sufrir afecciones cardíacas. Los rangos de colesterol entre **225 y 250 mg/dl** parecen concentrar una mayor probabilidad de afección, lo que refuerza la necesidad de monitorear los niveles de colesterol en los pacientes, especialmente aquellos con antecedentes familiares o factores de riesgo preexistentes.
7. **Ritmo cardíaco máximo y enfermedad cardíaca:** El análisis muestra una relación entre el ritmo cardíaco máximo y la presencia de enfermedad cardíaca. Los pacientes sin afección cardíaca presentan un ritmo cardíaco máximo superior (entre **138 y 162 bpm**), mientras que aquellos con enfermedad cardíaca tienen un ritmo máximo inferior (entre **120 y 145 bpm**). Este hallazgo sugiere que una menor capacidad de alcanzar altas pulsaciones podría ser un indicador temprano de disfunción cardiovascular, lo que resalta la importancia de realizar pruebas de esfuerzo en la evaluación del riesgo cardiovascular.