



EDA- Predicción del riesgo de ataque cardíaco

**"DESCIFRANDO LOS FACTORES QUE SALVAN
VIDAS"**

**MARIO ALCALDE ALVARADO
THE BRIDGE**

09-MAYO-2025

CONTEXTO DEL EDA

El conjunto de datos para la predicción del riesgo de ataque cardíaco es una herramienta clave para analizar las complejas relaciones entre la salud cardiovascular y sus factores determinantes. Los ataques cardíacos siguen siendo un desafío crítico de salud a nivel mundial, lo que exige una comprensión más profunda de sus causas y factores de prevención. Este conjunto recopila información detallada sobre edad, niveles de colesterol, presión arterial, tabaquismo, actividad física, hábitos alimenticios y otros aspectos relevantes, con el objetivo de identificar cómo interactúan estos factores en el riesgo de sufrir un infarto. Gracias al uso de analítica predictiva y técnicas de aprendizaje automático, este recurso permite a investigadores y profesionales de la salud diseñar estrategias más efectivas para la prevención y el manejo de enfermedades cardíacas. Así, este conjunto de datos representa un esfuerzo global por mejorar la comprensión de la salud cardiovascular y contribuir a la construcción de un futuro más saludable.

Contenido

Este conjunto de datos sintético proporciona un abanico completo de características relevantes para la salud del corazón y las elecciones de estilo de vida, abarcando detalles específicos de los pacientes como:

- Edad
- Género
- Niveles de colesterol
- Presión arterial
- Frecuencia cardíaca

Así como indicadores como:

- Diabetes
- Historial familiar
- Hábitos de tabaquismo
- Obesidad
- Consumo de alcohol

Adicionalmente, incluye factores de estilo de vida como:

- Horas de ejercicio
- Hábitos alimenticios
- Niveles de estrés
- Horas de sedentarismo

También se consideran aspectos médicos como:

- Problemas cardíacos previos
- Uso de medicación
- Niveles de triglicéridos

Y aspectos socioeconómicos como:

- Ingreso económico
- Atributos geográficos (país, continente y hemisferio)

El conjunto de datos, compuesto por **8763 registros** de pacientes de todo el mundo, culmina en una **variable binaria crucial** que indica la **presencia o ausencia del riesgo de ataque cardíaco**, ofreciendo un recurso completo para el análisis predictivo y la investigación en salud cardiovascular.

Descripción

Este conjunto de datos contiene **8763 registros** de pacientes de todo el mundo, con información sobre salud cardiovascular y estilo de vida, como edad, colesterol, presión arterial, diabetes, tabaquismo, obesidad y ejercicio. Incluye también aspectos médicos y socioeconómicos. Su objetivo es predecir la **presencia o ausencia de riesgo de ataque cardíaco** mediante análisis y modelos predictivos.

Hipótesis sobre los ataques cardíacos según hábitos y ritmo de vida

“Los hábitos y condiciones de vida tienen mayor impacto en la probabilidad de sufrir un ataque al corazón”

Con base en los datos del **Dataset de Predicción de Riesgo de Ataque al Corazón**, se pueden generar diversas **hipótesis** relacionadas con los hábitos y el estilo de vida de los individuos que podrían influir en el riesgo de sufrir un ataque al corazón. Estas hipótesis se enfocan en

factores **comportamentales, fisiológicos, psicológicos y socioeconómicos** que afectan la salud cardiovascular.

Además, el análisis de estas variables puede ayudar a diseñar estrategias de prevención personalizadas y políticas de salud pública más efectivas.

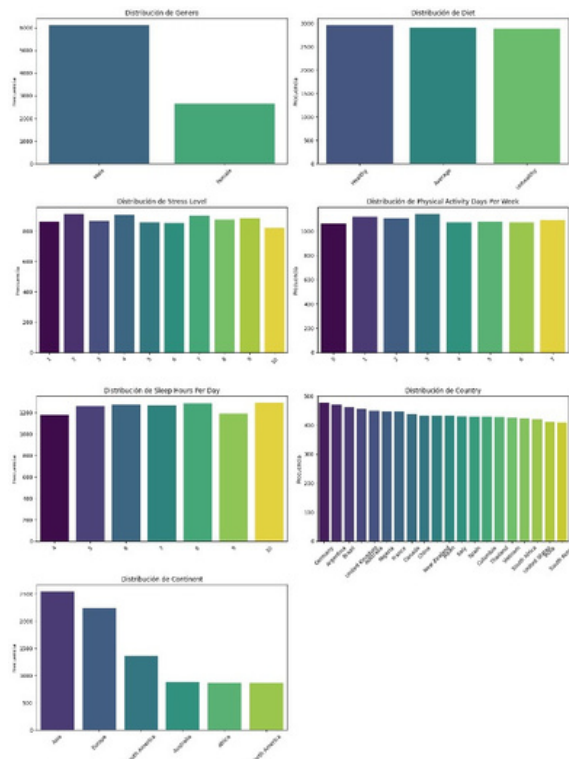
Análisis Univariante de Variables Categóricas

En general, las distribuciones de las variables categóricas en el conjunto de datos son bastante equilibradas, sin grandes diferencias entre categorías. Sin embargo, destacan dos excepciones claras:

- Género: Hay una mayor proporción de hombres en comparación con mujeres, lo que puede influir en los resultados relacionados con el riesgo de ataque cardíaco.
- Continente: La mayoría de los registros provienen de Asia, lo que indica un sesgo geográfico importante en la muestra y debe considerarse al generalizar las conclusiones.

Esta composición sugiere que cualquier análisis posterior debe tener en cuenta estas diferencias para evitar sesgos en la interpretación de los resultados.

Graficas de barras variables categóricas



Análisis Claro del Resultado

- Smoking (Tabaquismo): Hay 7853 fumadores frente a solo 902 no fumadores. Esto indica que el tabaquismo está altamente presente en la muestra, lo que ya es un fuerte factor de riesgo cardiovascular.
- Diabetes: Hay más personas con diabetes (5711 casos) que sin diabetes (3044). Esto es relevante porque, en la vida real, la diabetes no suele ser tan común en la población general. Aquí parece estar sobre-representada, probablemente porque el dataset está enfocado en riesgos cardíacos.
- Obesity (Obesidad): La distribución está más equilibrada: 4389 sí vs 4366 no. Esto sugiere que la obesidad afecta a casi la mitad de la muestra.
- Alcohol Consumption: Hay más personas que consumen alcohol (5236) frente a quienes no (3519). No tan desbalanceado como el tabaquismo, pero sigue siendo significativo.
- Previous Heart Problems (Antecedentes): Hay 4340 personas que ya tuvieron problemas cardíacos, frente a 4415 que no. Esto está bastante equilibrado.
- Heart Attack Risk (Riesgo de ataque cardíaco):
 - o 5189 personas están clasificadas como en riesgo alto.
 - o 3566 están como sin riesgo.
 - o Es decir, aproximadamente un 59% están en riesgo, lo que muestra que el dataset está bastante enfocado en población de riesgo.

Conclusión Fuerte:

Este resumen revela que en tu conjunto de datos:

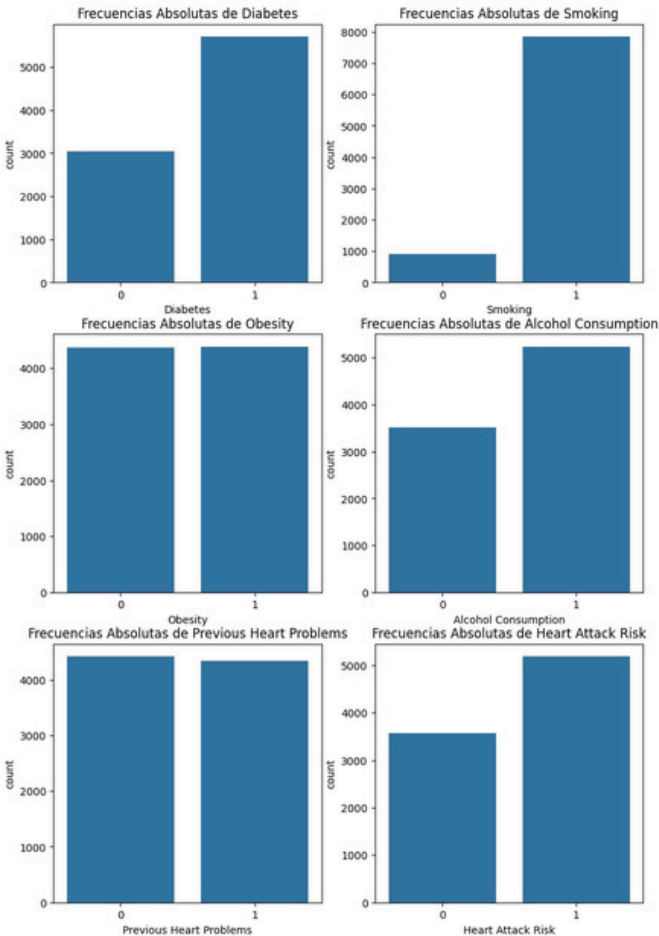
- El tabaquismo y la diabetes están altamente presentes, lo que indica que son focos críticos para intervenir.
- Hay una prevalencia alta de personas en riesgo de ataque cardíaco (casi 6 de cada 10).
- Además, la obesidad y el consumo de alcohol son factores frecuentes, lo que refuerza la relación entre estos hábitos y el riesgo cardiovascular.

- En pocas palabras: fumar, tener diabetes y obesidad son las combinaciones más peligrosas que se repiten en la mayoría de los casos en riesgo en tu dataset.

	Diabetes	Smoking	Obesity	Alcohol Consumption	Previous Heart Problems	Heart Attack Risk
Diabetes						
1	5711	7853	4389	5236	4340	5189
0	3044	902	4366	3519	4415	3566

	Diabetes	Smoking	Obesity	Alcohol Consumption	Previous Heart Problems	Heart Attack Risk
Diabetes						
1	65.23	89.7	50.13	59.81	49.57	59.27
0	34.77	10.3	49.87	40.19	50.43	40.73

Graficas de barras variables binarias

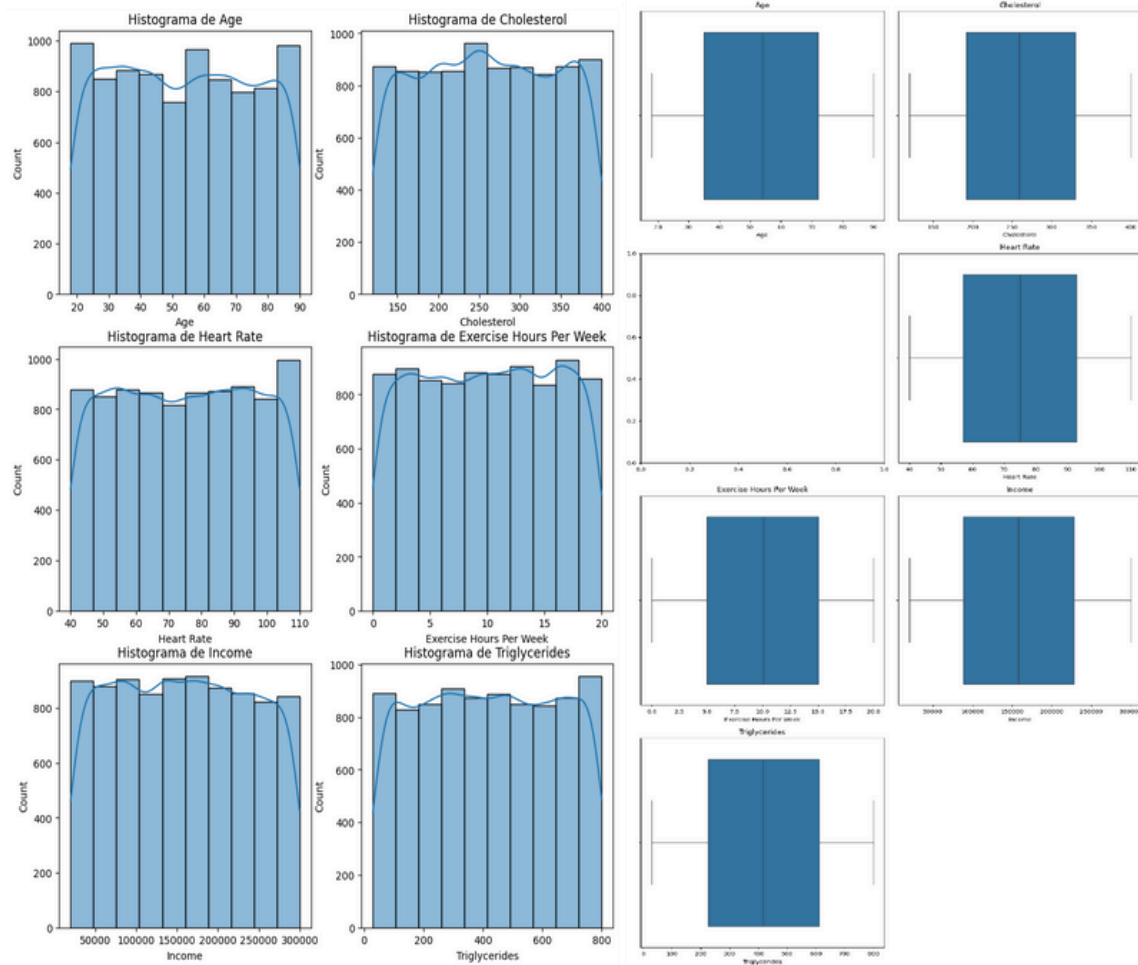


Análisis Univariante Variables Numéricas

Del análisis descriptivo de las variables numéricas, se concluye que la población estudiada presenta un perfil metabólico elevado y con riesgo. La edad promedio es de 53 años, abarcando principalmente a adultos de mediana edad y mayores, el grupo más propenso a enfermedades cardíacas. Los niveles de colesterol (260 mg/dL) y triglicéridos (418 mg/dL) están muy por encima de los valores recomendados, indicando un riesgo cardiovascular latente en la mayoría de los casos. Aunque el promedio de ejercicio semanal es de 10 horas, existe un 25% de la población que realiza menos de 5 horas, evidenciando un segmento con baja actividad física. Además, los datos económicos muestran que, si bien la muestra tiene ingresos relativamente altos, los problemas cardiovasculares no distinguen nivel socioeconómico, y parecen estar más relacionados con los hábitos y el descontrol metabólico. En conjunto, estos indicadores refuerzan la necesidad de intervenciones en estilo de vida y control biométrico en esta población.

	Age	Cholesterol	Heart Rate	Exercise Hours Per Week	Income	Triglycerides
count	8755.000000	8755.000000	8755.000000	8755.000000	8755.000000	8755.000000
mean	53.710109	259.907938	75.019760	10.013977	158260.555797	417.719475
std	21.249780	80.827869	20.552066	5.783809	80574.380529	223.704091
min	18.000000	120.000000	40.000000	0.002442	20062.000000	30.000000
25%	35.000000	192.000000	57.000000	4.979900	88310.000000	226.000000
50%	54.000000	259.000000	75.000000	10.069559	157866.000000	417.000000
75%	72.000000	330.000000	93.000000	15.050018	227786.500000	612.000000
max	90.000000	400.000000	110.000000	19.998709	299954.000000	800.000000

Histogramas y boxplot de variables numéricas



Análisis de Correlación entre Triglicéridos y Colesterol:

En esta muestra, analizamos dos variables muy significativas de nuestro dataset, en el cual no se observa una correlación lineal significativa entre los niveles de **triglicéridos** y **colesterol**. Aunque ambas variables, por separado, tienen una relación significativa con el **riesgo de ataque cardíaco** (según el análisis previo), no están directamente correlacionadas entre sí.

Por ejemplo:

- Una persona puede presentar **triglicéridos elevados** y **colesterol normal**.
- Otra persona podría tener **colesterol elevado** pero **triglicéridos bajos**.

Esto ilustra cómo las dos variables pueden variar de manera independiente sin una relación lineal clara entre ellas.

Conclusión:

- El valor de correlación de **-0.0051** es prácticamente nulo, lo que sugiere que **no existe una relación lineal** entre los niveles de triglicéridos y colesterol en esta muestra.
- El **p-valor de 0.6329**, que es mucho mayor que el umbral de 0.05, refuerza que la correlación entre estas dos variables no es **estadísticamente significativa**.

Por lo tanto, aunque tanto **triglicéridos** como **colesterol** son importantes en la predicción del riesgo cardiovascular, no hay una dependencia lineal directa entre ellos en este caso.

	Triglycerides	Cholesterol
Triglycerides	1.000000	-0.005106
Cholesterol	-0.005106	1.000000

Test de relación entre variables y Heart Attack Risk

Según los resultados obtenidos de este conjunto de datos, las variables de **triglicéridos** y **colesterol** son las dos principales que muestran diferencias significativas entre las personas con **riesgo de ataque cardíaco** y aquellas sin este riesgo.

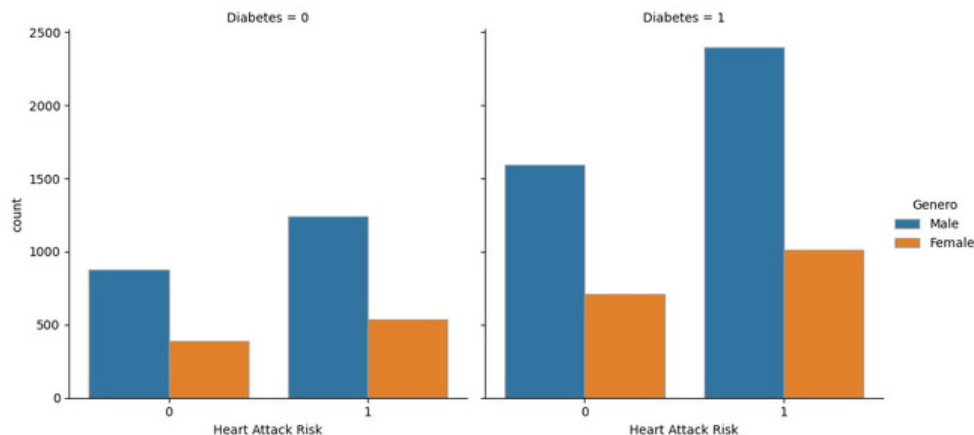
En cambio, otras variables como **edad**, **nivel de ejercicio**, **ingreso** y **frecuencia cardíaca** no muestran diferencias destacables en relación con el riesgo de ataque cardíaco en este análisis.

Conclusión: El **colesterol** y los **triglicéridos** están fuertemente relacionados con el riesgo de ataque cardíaco, siendo las variables más relevantes en la predicción del riesgo cardiovascular según esta muestra. Esto sugiere que un control adecuado de estos factores podría ser clave para la prevención.

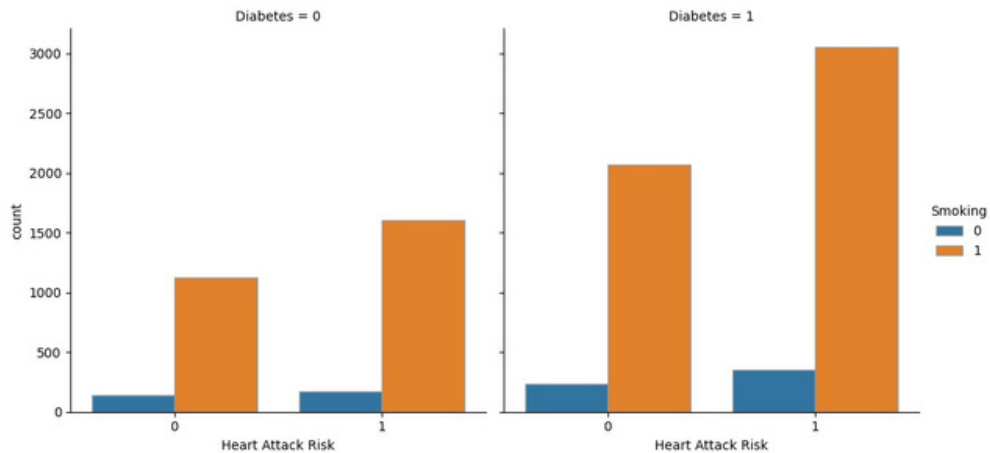
	Variable	U-Estadístico	P-valor	Significativo (<0.05)
2	Triglycerides	8253511.5	8.475531e-18	Sí
1	Cholesterol	8636049.0	1.152694e-07	Sí
0	Age	9134633.5	3.124728e-01	No
4	Exercise Hours Per Week	9167453.0	4.669186e-01	No
5	Income	9224733.5	8.145647e-01	No
3	Heart Rate	9261391.5	9.354898e-01	No

Análisis bivalente y multivariante

Análisis variables (Heart Attack Risk, Genero y Diabetes) El análisis revela que el género masculino y la presencia de diabetes son factores determinantes que elevan significativamente el riesgo de ataque cardíaco en esta población. Estos hallazgos refuerzan la importancia de considerar ambos factores como indicadores prioritarios en las estrategias de prevención cardiovascular.

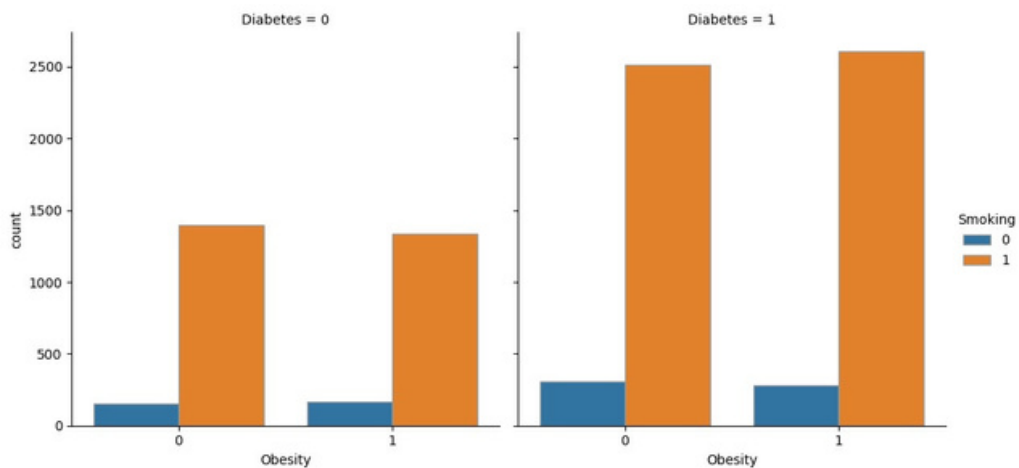


Análisis variables (Heart Attack Risk , Smoking y Diabetes) Los resultados del análisis demuestran de manera contundente que la combinación de diabetes y tabaquismo incrementa significativamente las probabilidades de sufrir un ataque cardíaco. Estos dos factores juntos representan un riesgo cardiovascular elevado, lo que resalta la importancia de abordar tanto la diabetes como el hábito de fumar para la prevención de eventos cardíacos graves.



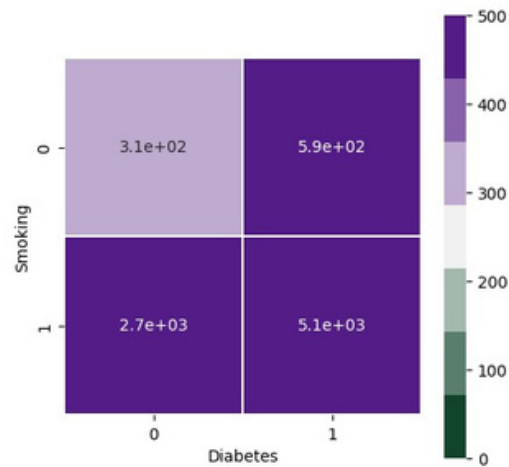
Análisis variables (Heart Attack Risk , Smoking y Diabetes)

La obesidad muestra una fuerte asociación con la presencia de diabetes y el hábito de fumar, conformando un perfil de riesgo elevado que potencia la probabilidad de desarrollar enfermedades cardiovasculares.



Analisis Heatmap (Smoking y Diabetes)

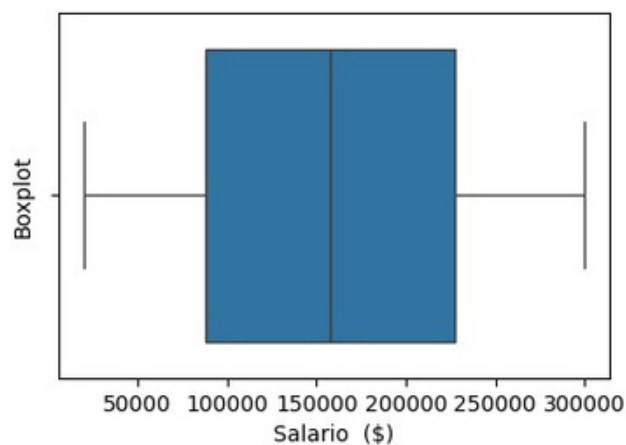
El hábito de fumar emerge como un factor claramente asociado a una mayor incidencia de diabetes en esta población.



Diabetes	0	1
Smoking		
0	314	588
1	2730	5123

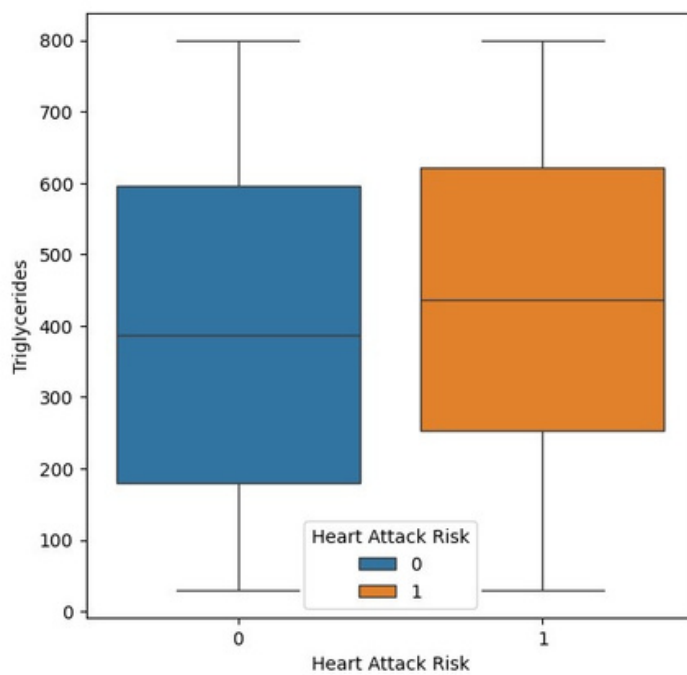
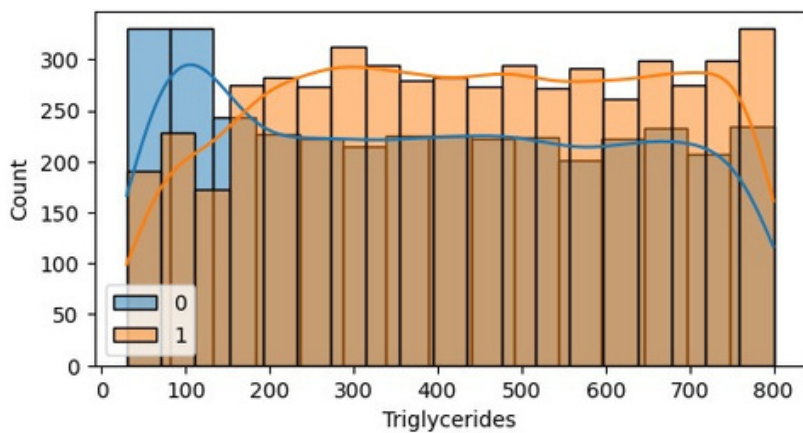
Analisis Boxplot Ingresos

El análisis del boxplot de ingresos muestra una distribución bastante homogénea, sin presencia de outliers significativos. La mayoría de los valores se agrupan en un rango estrecho, lo que indica que los ingresos dentro de esta muestra son relativamente similares, sin grandes diferencias entre los participantes.



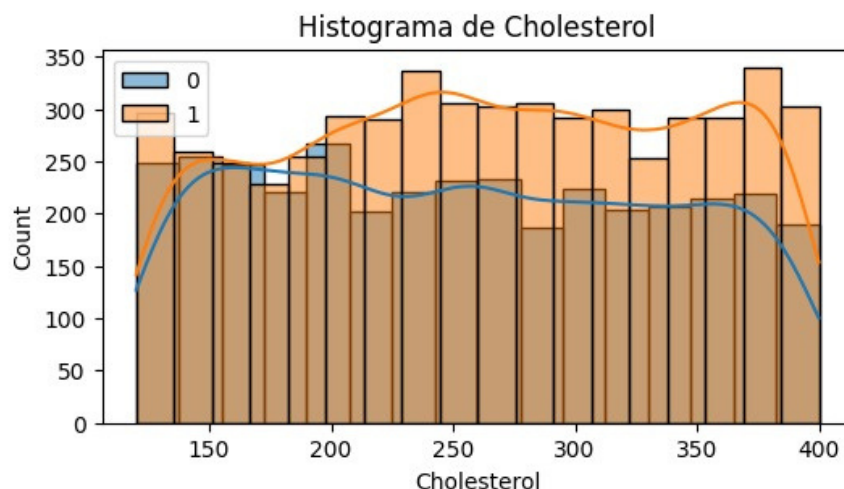
Analisis Histograma-Boxplot (Heart Attack Risk - Triglycerides)

Se observa que niveles bajos de triglicéridos se asocian con un menor riesgo de ataque cardíaco, mientras que niveles elevados aumentan significativamente dicha probabilidad. Esto destaca la importancia de controlar los triglicéridos como medida preventiva clave en salud cardiovascular.



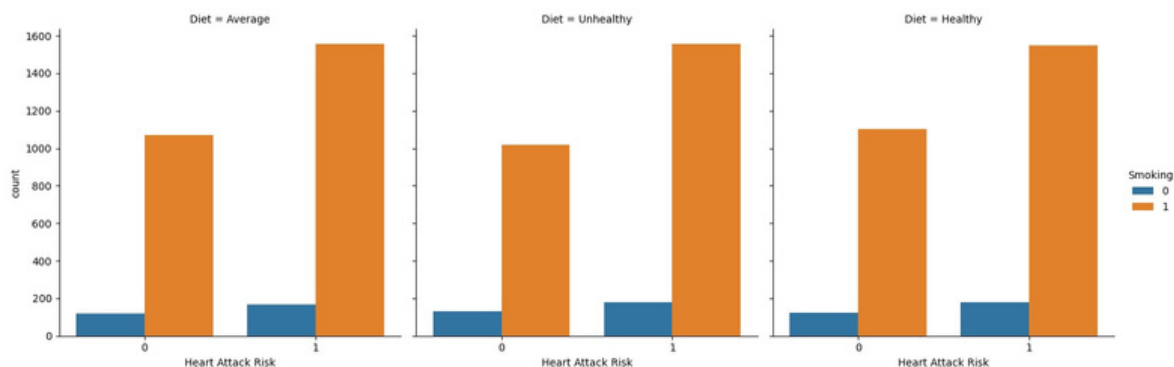
Analisis Histograma (Heart Attack Risk - Colesterol)

Este análisis confirma de forma clara que niveles elevados de colesterol se relacionan directamente con un mayor riesgo de ataque cardíaco. Las personas con colesterol alto predominan en el grupo de alto riesgo, lo que subraya la importancia de mantener estos valores controlados para prevenir enfermedades cardiovasculares.



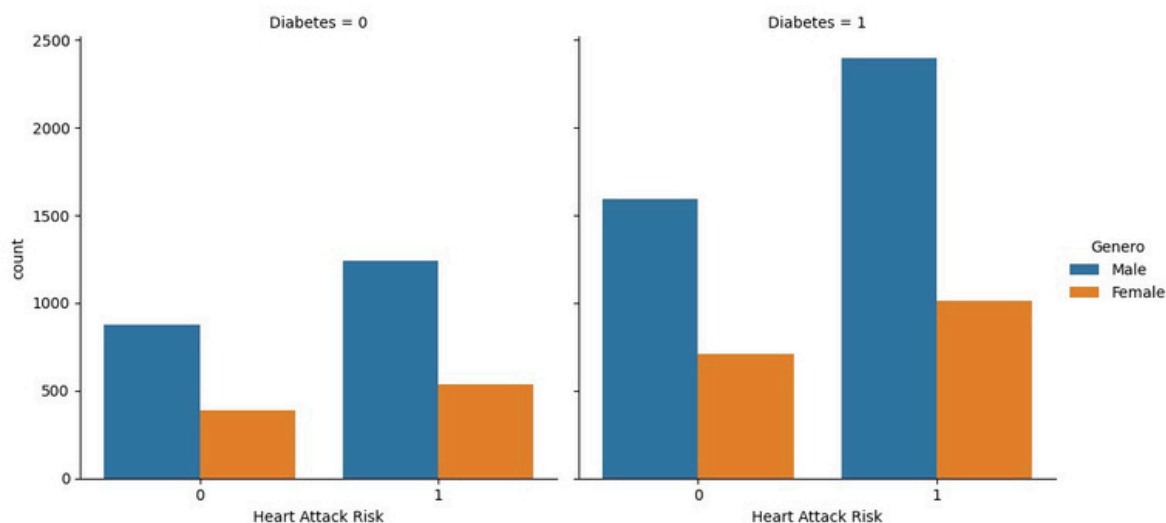
Analisis (Heart Attack Risk- Smoking- Diet)

El tabaquismo emerge como un factor clave que incrementa significativamente el riesgo de ataque cardíaco, independientemente de la dieta seguida. Incluso entre quienes mantienen una dieta saludable, los fumadores tienen un mayor número de casos en riesgo en comparación con los no fumadores.



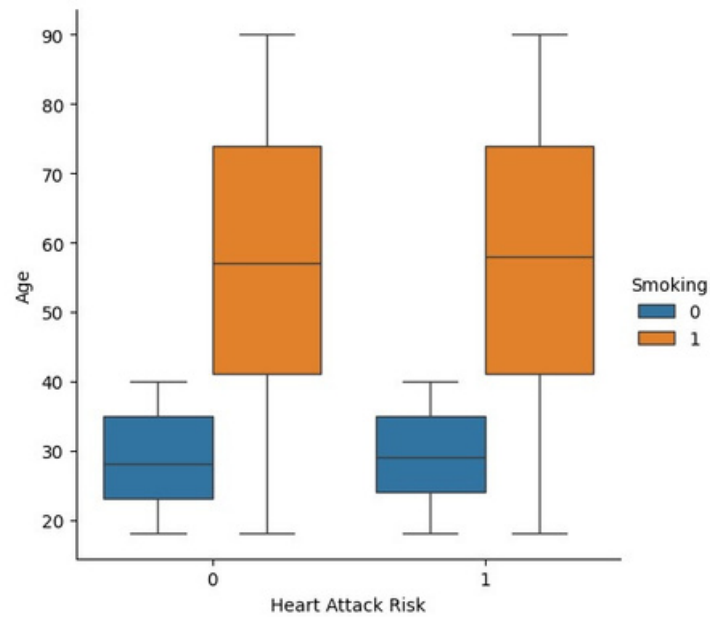
Analisis (Heart Attack Risk- Diabetes- Genero)

Este análisis pone de manifiesto que los hombres, en particular aquellos con diabetes, enfrentan un riesgo considerablemente mayor de sufrir un ataque cardíaco en comparación con las mujeres.



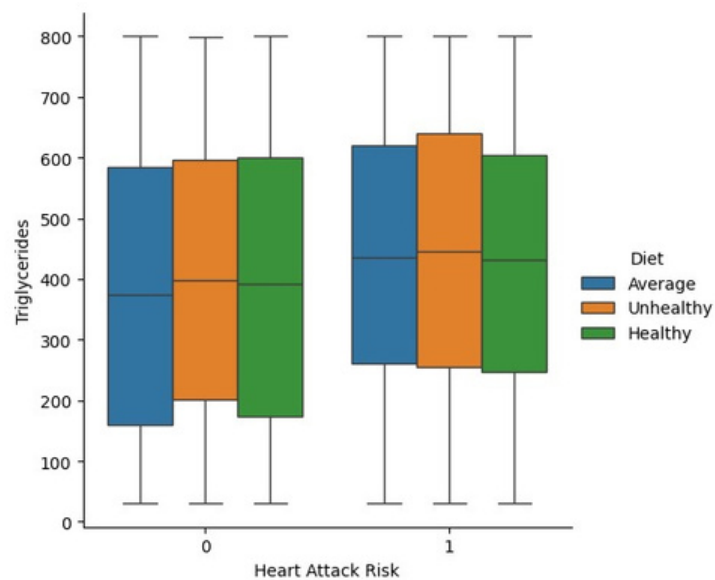
Analisis (Heart Attack Risk- Smoking- Edad)

El análisis revela que las personas mayores de 40 años que fuman presentan un riesgo significativamente mayor de sufrir un ataque cardíaco, lo que subraya el impacto negativo del tabaquismo en la salud cardiovascular, especialmente en edades avanzadas. No obstante, también se observa que incluso en personas no fumadoras, existe un riesgo de ataque cardíaco, lo que indica que otros factores además del tabaquismo contribuyen al riesgo cardiovascular.



Analisis (Heart Attack Risk- Triglycerides- Diet)

El análisis muestra que las personas con dietas poco saludables tienen niveles de triglicéridos significativamente más altos, lo que incrementa notablemente el riesgo de sufrir un ataque cardíaco.



Conclusión del Análisis Exploratorio de Datos (EDA)

El análisis realizado sobre el conjunto de datos de riesgo de ataque cardíaco revela diversas relaciones significativas entre varios factores de salud y el riesgo cardiovascular. A continuación, se destacan los hallazgos más importantes:

1. Triglicéridos y colesterol: Aunque no se observa una correlación directa y lineal entre los niveles de triglicéridos y colesterol, ambos factores por separado muestran una clara asociación con el riesgo de ataque cardíaco. Los niveles elevados de ambos se relacionan con un riesgo incrementado de sufrir un evento cardiovascular, lo que refuerza la importancia de controlar estos indicadores como medida preventiva.
2. Factores de riesgo combinados: Se identificó que la combinación de factores como la diabetes y el hábito de fumar aumenta considerablemente las probabilidades de sufrir un ataque cardíaco. Estas personas presentan un riesgo cardiovascular mucho más alto, subrayando la necesidad de abordar estos factores de manera integral para una prevención eficaz.
3. El tabaquismo como factor clave: Fumar se confirma como un factor decisivo para el aumento del riesgo de ataque cardíaco, independientemente de la dieta seguida. Incluso aquellos que mantienen una dieta saludable tienen un mayor riesgo de ataque cardíaco si son fumadores. Esto pone de manifiesto la importancia de eliminar el tabaquismo como uno de los principales objetivos en la prevención cardiovascular.
4. El género y la diabetes: Se observó una clara diferencia de género en el riesgo cardiovascular, donde los hombres, especialmente los que padecen diabetes, presentan un riesgo significativamente mayor de sufrir un ataque cardíaco. Esto resalta la necesidad de dirigir los esfuerzos de prevención cardiovascular, especialmente hacia los hombres diabéticos.
5. Obesidad y sus factores asociados: La obesidad se encuentra estrechamente relacionada con la presencia de diabetes y el hábito de fumar, lo que agrava aún más el riesgo cardiovascular. Este hallazgo refuerza la importancia de un enfoque integral que considere múltiples factores de riesgo simultáneamente.
6. Impacto de la dieta: Se observó que las personas con dietas no saludables tienen niveles más altos de triglicéridos, lo que incrementa notablemente su riesgo de ataque cardíaco. Este resultado subraya la importancia de mantener una alimentación equilibrada como un factor clave en la prevención de enfermedades cardiovasculares.

7. El perfil de los fumadores: Además, se confirmó que la mayoría de los fumadores en este dataset son mayores de 40 años, y dentro de este grupo, el riesgo de ataque cardíaco es notablemente mayor. Sin embargo, también se destacó que incluso aquellas personas que no fuman aún tienen un riesgo considerable de sufrir un ataque, lo que subraya la importancia de otros factores como la dieta y la actividad física en la salud cardiovascular.

Recomendaciones para la prevención: A partir de estos hallazgos, se recomienda un enfoque preventivo centrado en reducir el tabaquismo, controlar los niveles de triglicéridos y colesterol, promover dietas saludables, y gestionar la diabetes y la obesidad. Además, es crucial implementar estrategias de prevención dirigidas específicamente a los hombres y las personas mayores de 40 años, que parecen ser los más vulnerables a los riesgos cardiovasculares.