

ANÁLISIS DE DATOS

Análisis de datos sobre ataques al corazón
Judith Barrón Rodríguez

Información general del dataset

Tamaño del dataset

- 246, 022 filas
- 40 columnas

Cantidad de variables

40

Descripción de variables

- **State:** Estado de procedencia.
- **Sex:** Sexo de la persona, masculino o femenino.
- **GeneralHealth:** Estado de salud general.
- **LastCheckupTime:** Última vez que se realizó una revisión médica.
- **PhysicalActivities:** Realización de actividades físicas.
- **SleepHours:** Horas de sueño
- **RemovedTeeth:** Dientes removidos.
- **HadHeartAttack:** Ataques al corazón en el pasado.
- **HadAngina:** Haber tenido anginas en el pasado.
- **HadStroke:** Derrame cerebral en el pasado.
- **HadAsthma:** Padecer asma
- **HadSkinCance:** Haber sufrido o sufrir cancer de piel.
- **HadCOPD:** Padecer una enfermedad pulmonar obstructiva crónica.
- **HadDepressiveDisorder:** Desorden depresivo
- **HadKidneyDisease:** Padecer problemas de riñon
- **HadArthritis:** Padecer artritis.
- **HadDiabetes:** Padecer diabetes.
- **DeafOrHardOfHearing:** Sordera o dificultad para escuchar.
- **BlindOrVisionDifficulty:** Ceguera o dificultad para ver.
- **DifficultyConcentrating:** Dificultad de concentración

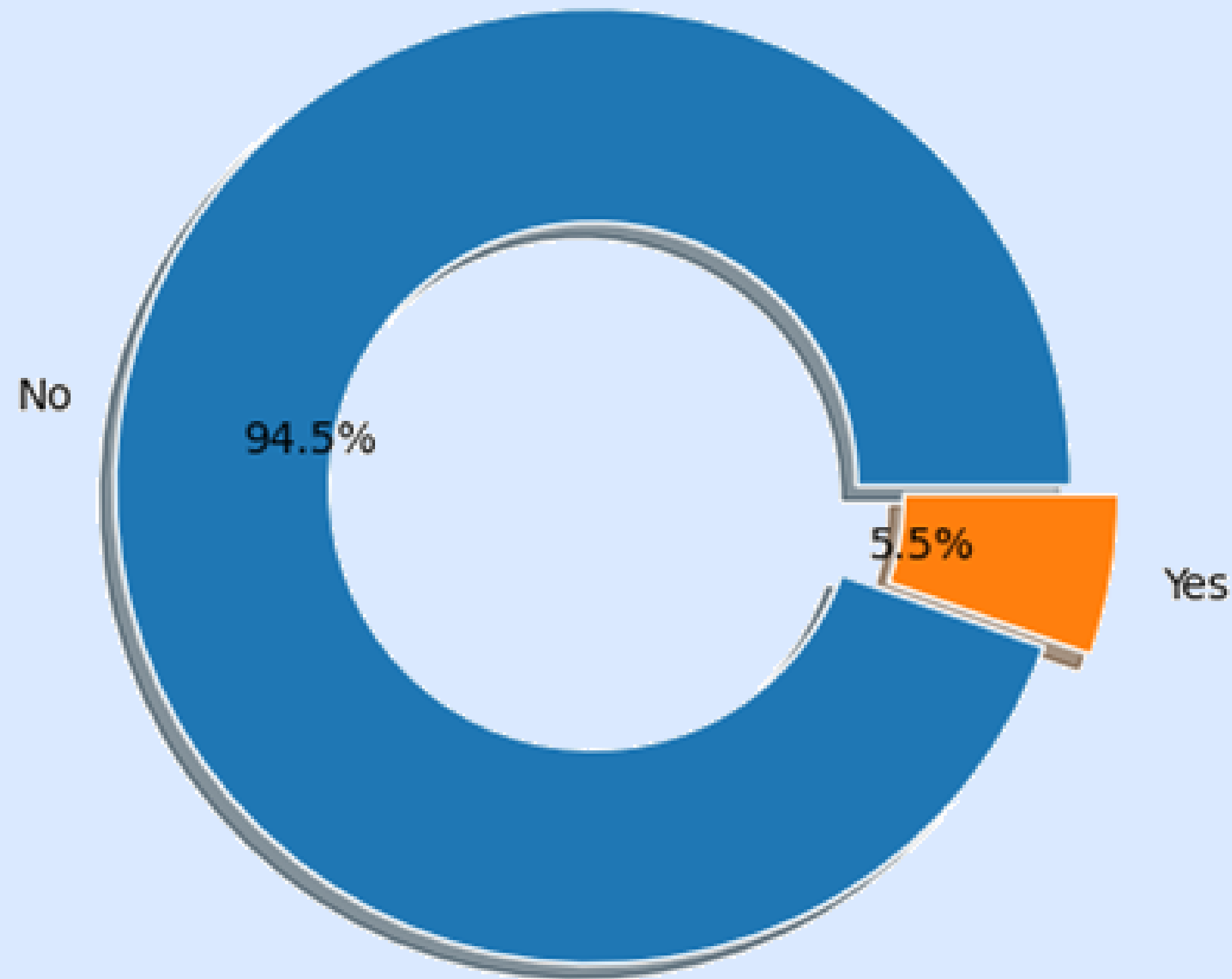
Descripción de variables

- **HeightInMeters:** Estatura en metros.
- **WeightInKilograms:** Peso en kilogramos.
- **BMI:** índice de masa corporal.
- **AlcoholDrinkers:** Consumo de alcohol.
- **HIVTesting:** Haberse realizado un examen de (VIH)
- **FluVaxLast12:** Haber recibido una vacuna contra la gripe en los últimos 12 meses.
- **PneumoVaxEver:** Alguna vez haber recibido una vacuna contra la neumonía.
- **TetanusLast10Tdap:** Haber recibido vacuna contra el tetanos en los últimos 10 años (o refuerzo)
- **CovidPos:** Ser una persona post Covid
- **DifficultyWalking:** Dificultad para caminar.
- **DifficultyDressingBathing:** Dificultad para vestirse o bañarse.
- **SmokerStatus:** Atividad como persona fumadora.
- **ECigaretteUsage:** Uso de cigarros electrónicos.
- **ChestScan:** Haberse realizado un tomografía de tórax.
- **RaceEthnicityCategory:** Raza o etnia.
- **AgeCategory:** Edad por intervalos.
- **HighRiskLastYear:** Haber estado en alto riesgo en el último año.



Gráficas generales de los datos

PERSONAS QUE HAN TENIDO UN ATAQUE AL CORAZÓN

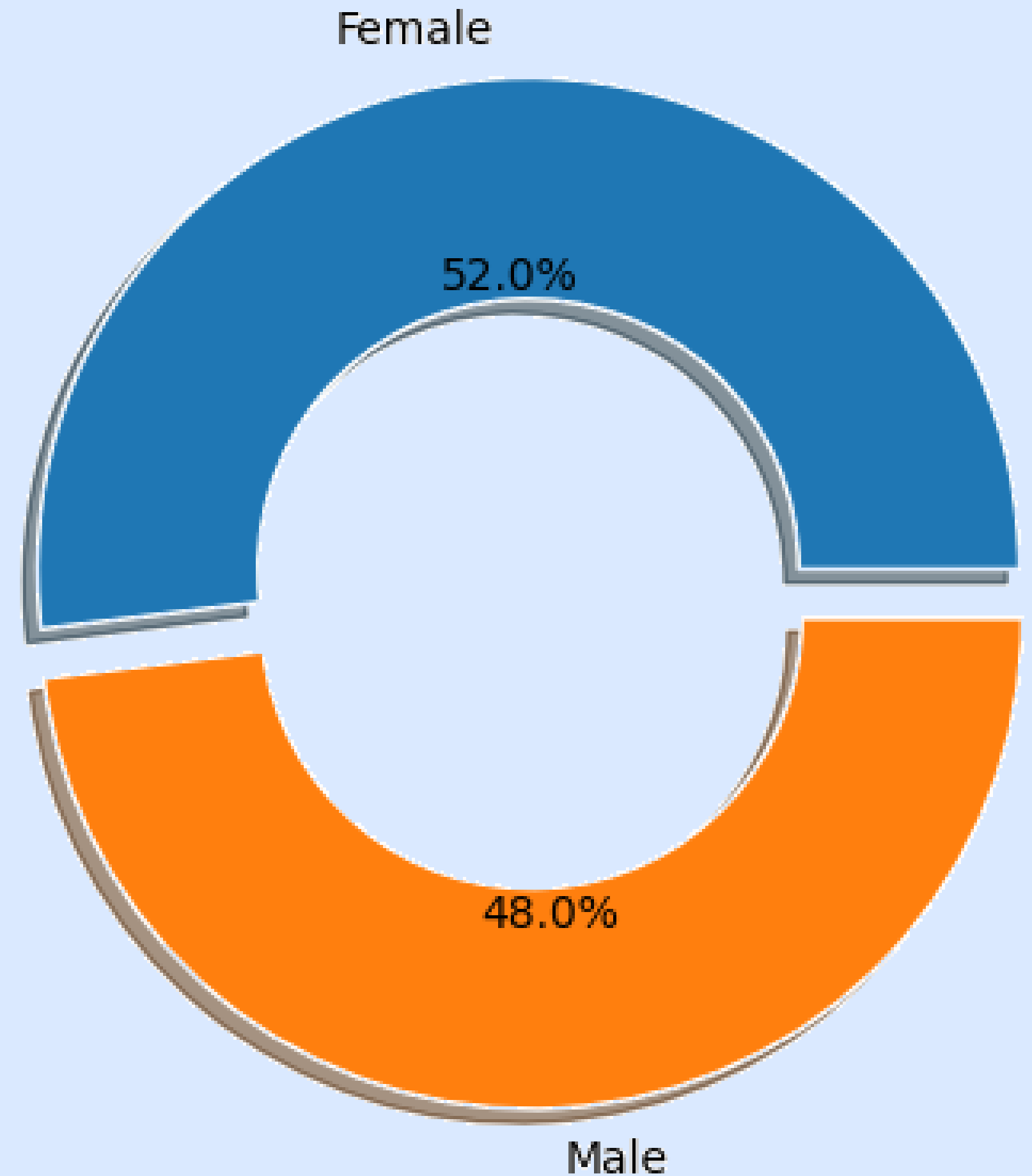


La gráfica representa el porcentaje de personas que han sufrido o no un ataque al corazón

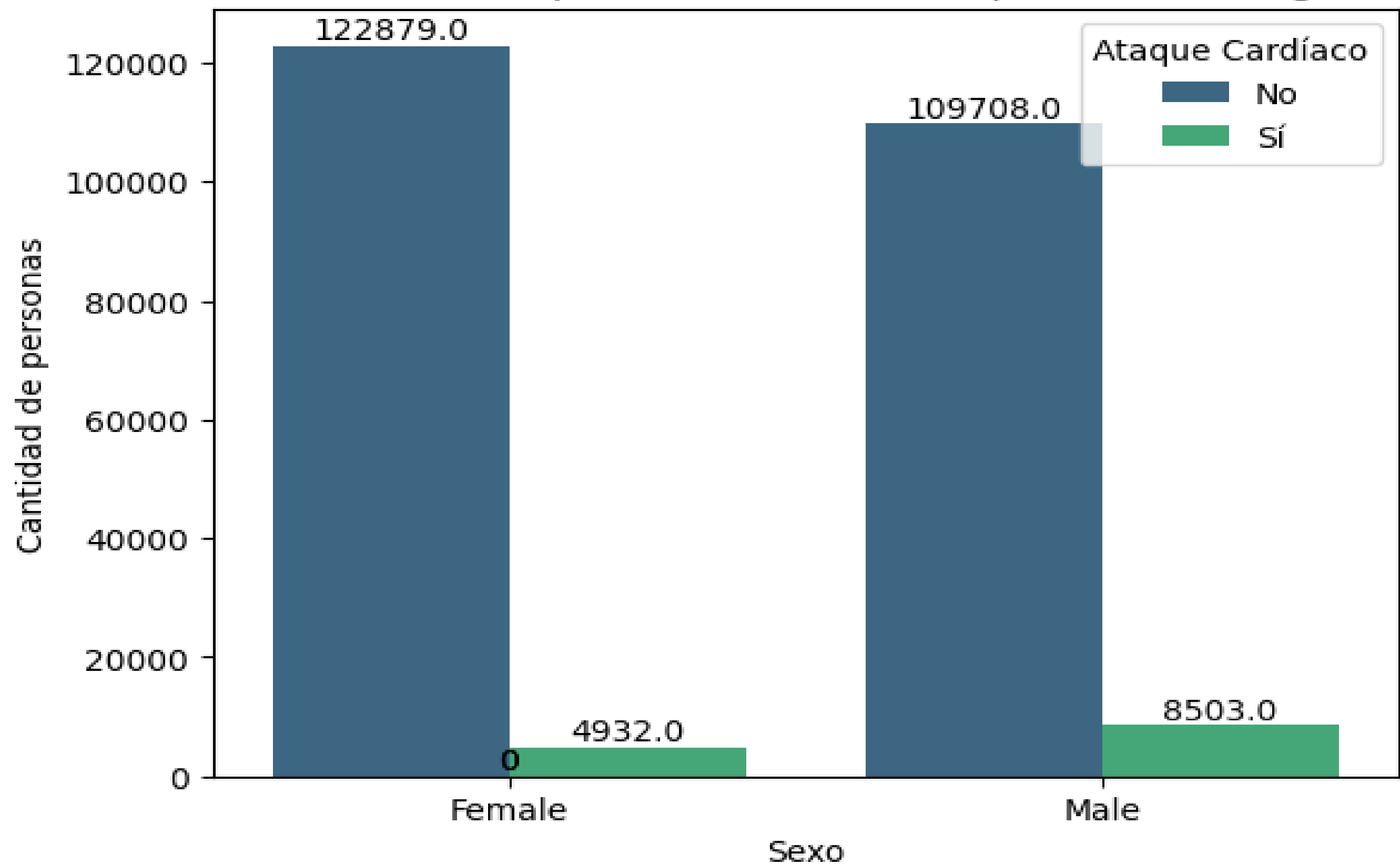
DISTRIBUCIÓN DE PERSONAS SEGÚN SU SEXO

La gráfica representa el porcentaje de hombres y mujeres presentes en la muestra.

Distribución por sexo



La gráfica representa la cantidad de personas que ha sufrido o no un ataque al corazón según su sexo.



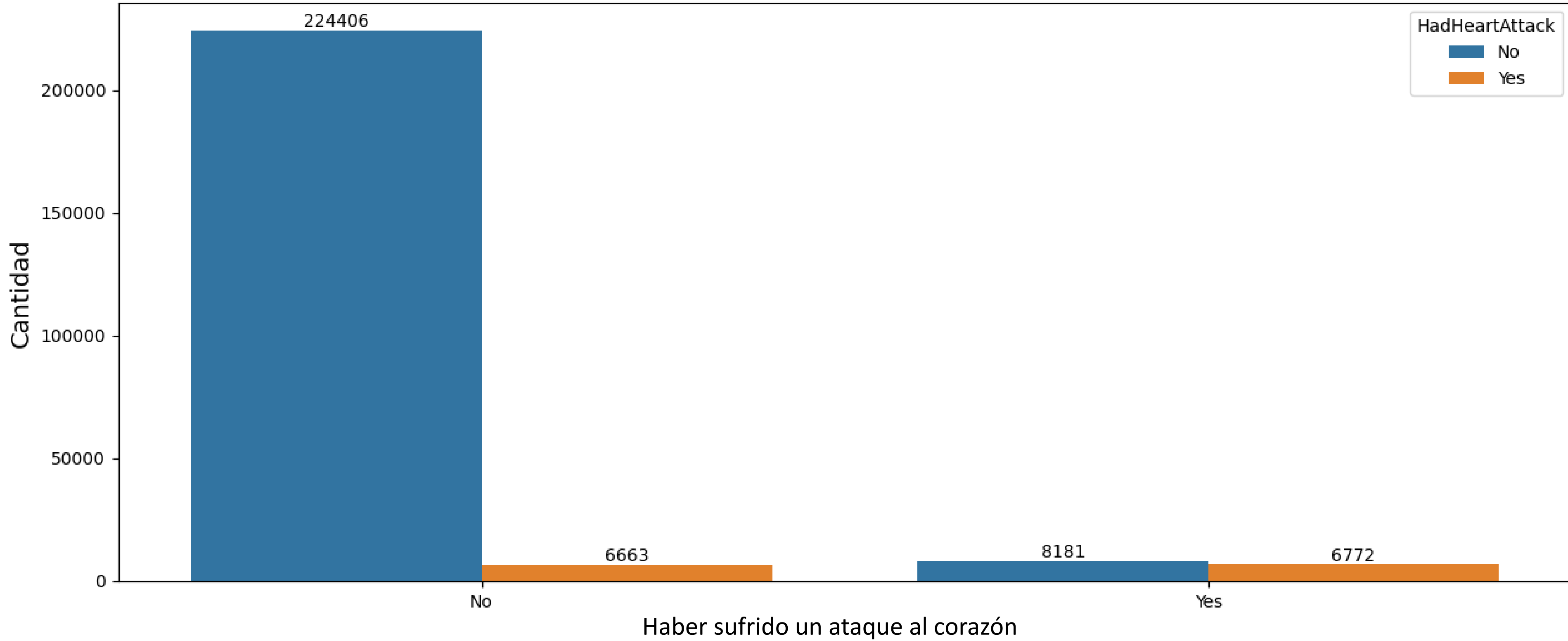
Matriz de correlación

La tabla representa la relación entre las variables con haber tenido un ataque al corazón

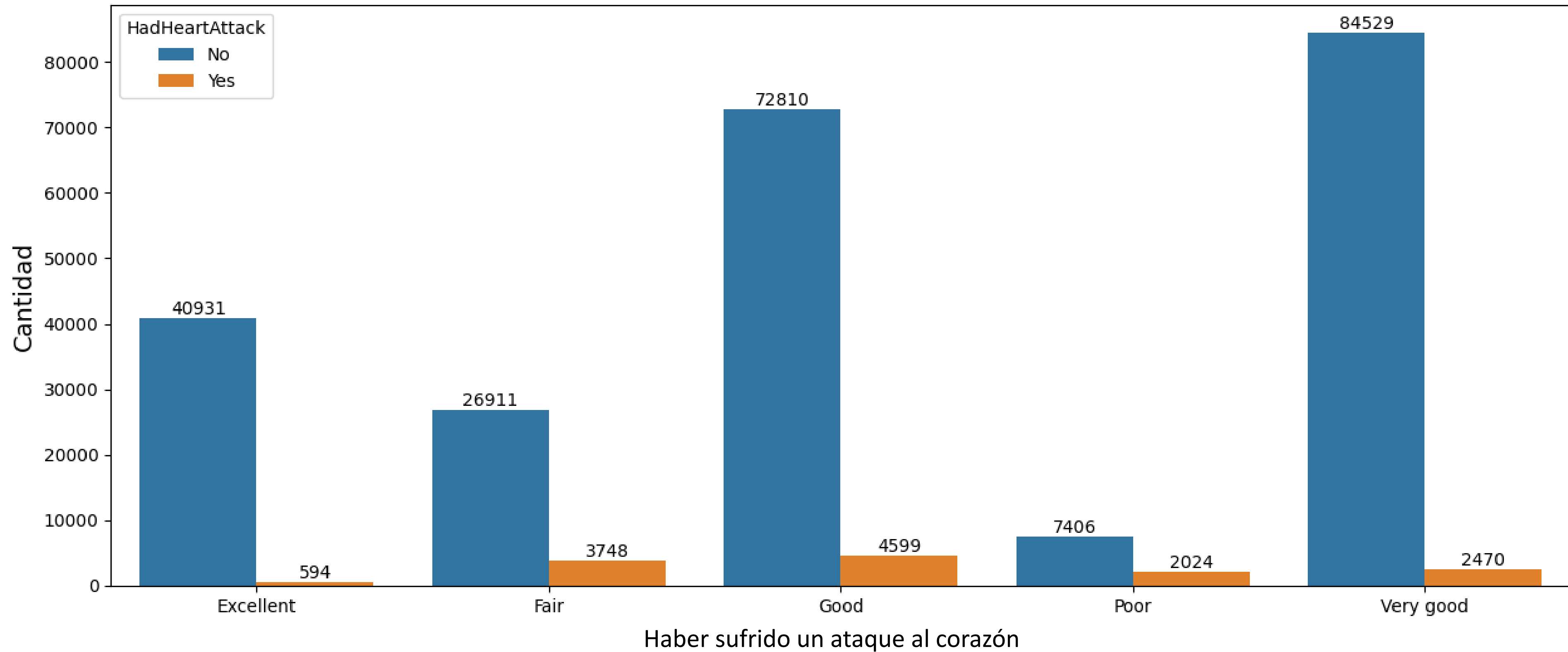
La variable de “HadAnginas” (Tener anginas), tiene mayor relación con haber tenido un ataque al corazón; mientras que la variable “GeneralHealth” es la variable con menor relación.

Variable	Correlación
GeneralHealth	-0.1856
HadAngina	0.4459
HadStroke	0.1771
DifficultyWalking	0.1598
AgeCategory	0.1721

Ataques al corazón entre personas que padecen/padecieron anginas



Ataques al corazón en relación a la salud en general

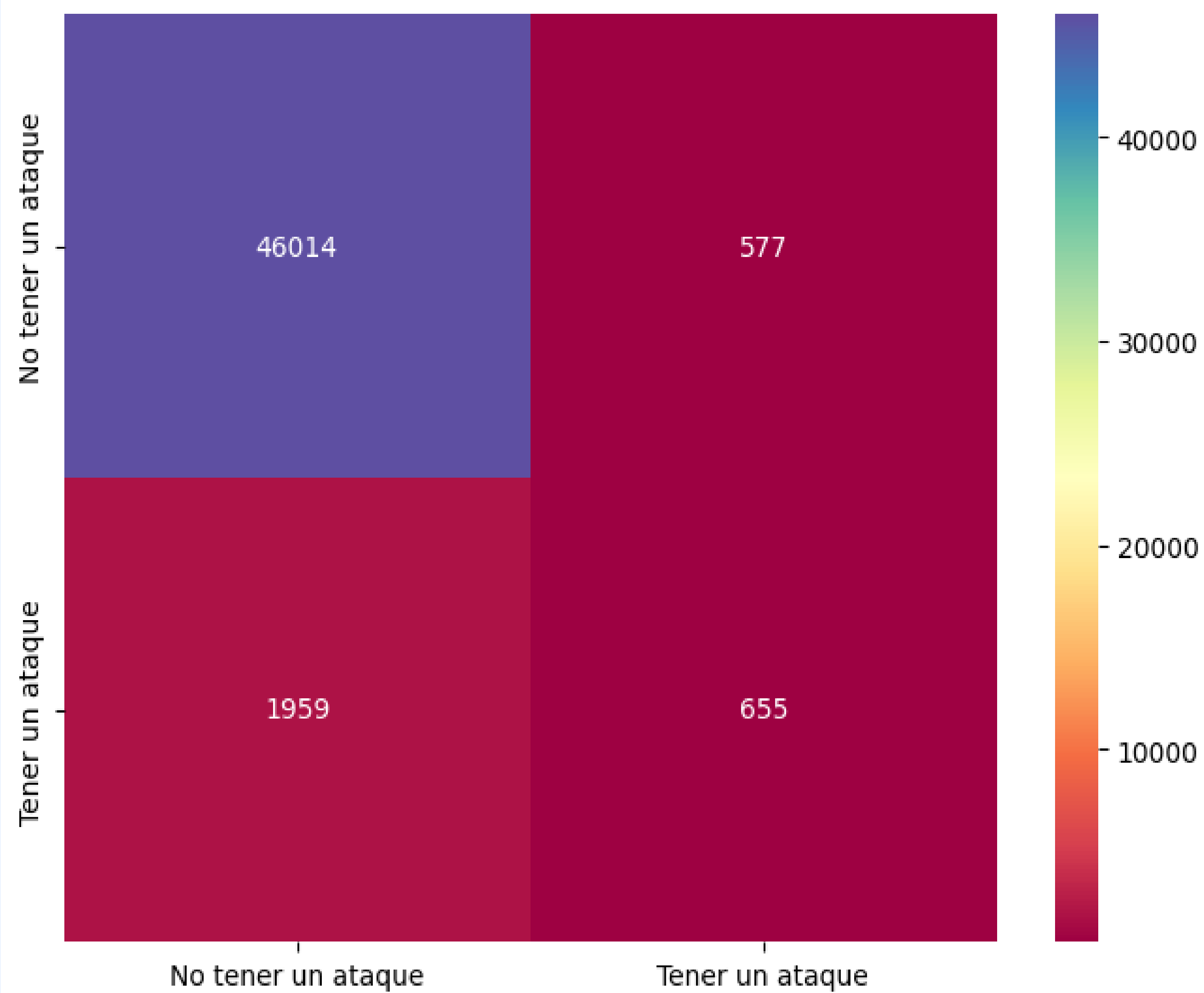




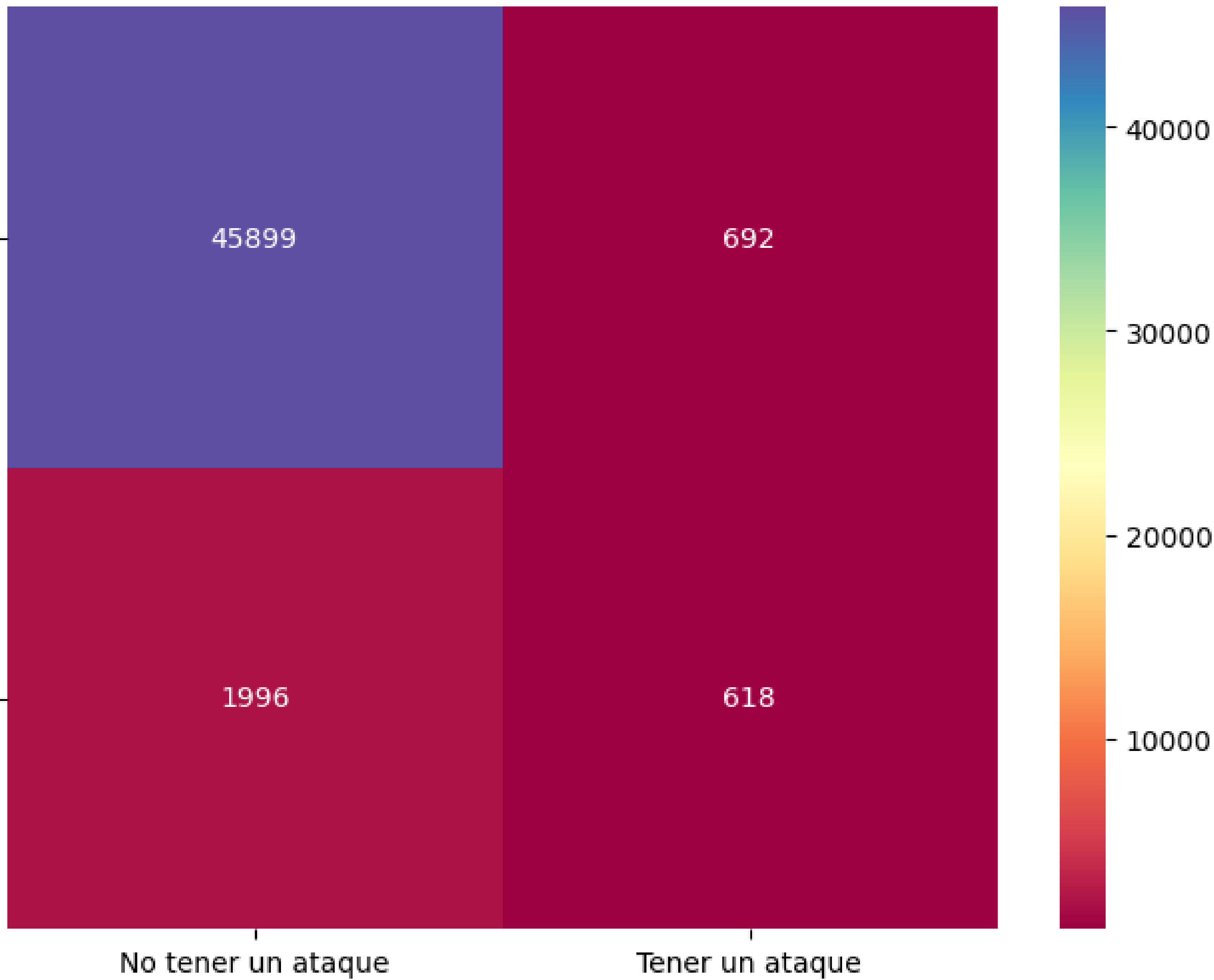
Evaluación de modelos

COMPARACIÓN DE MODELOS

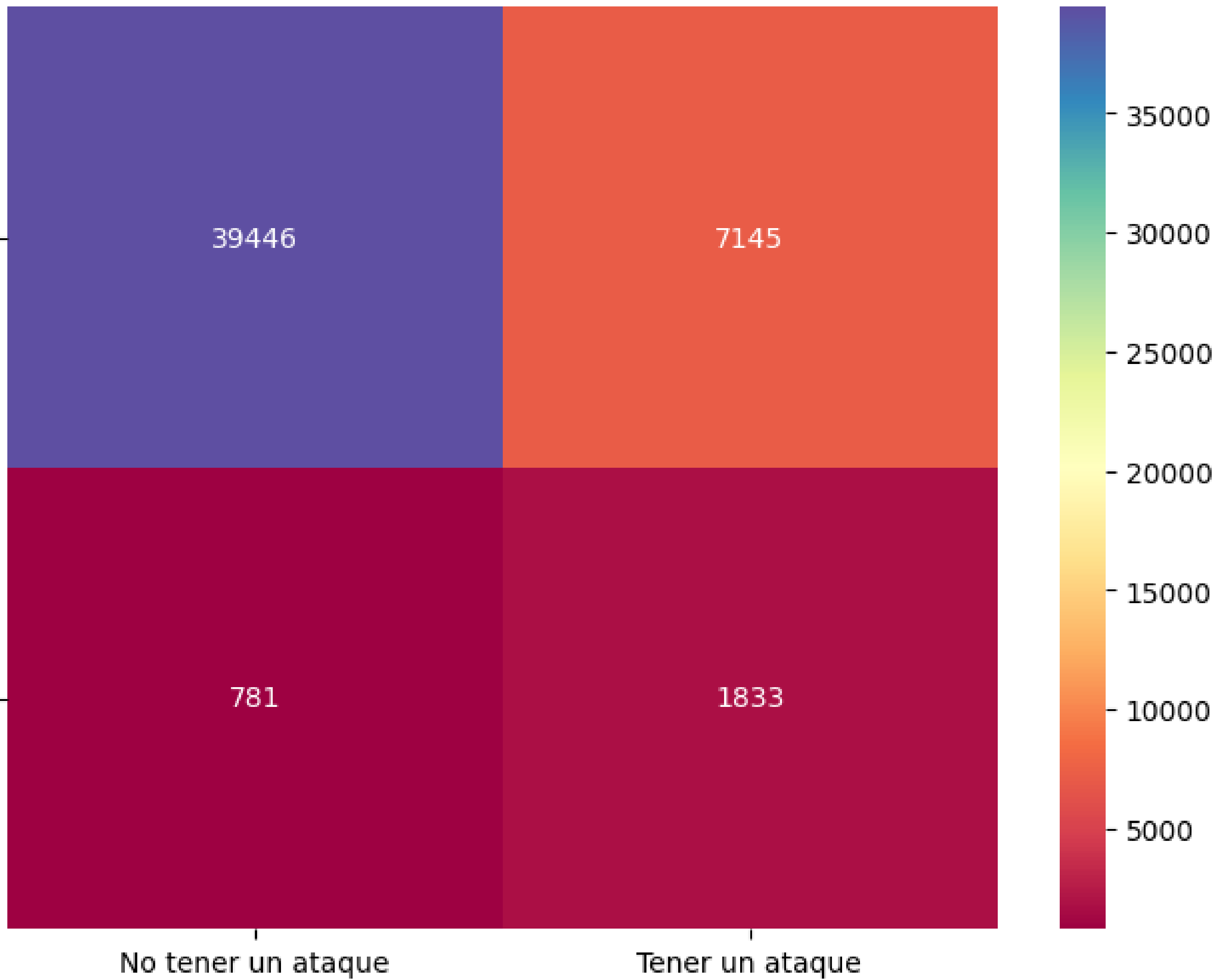
MODELO	ACCURACY	PRECISION	RECALL	F1-SCORE
Logistic Regression	0.948461	0.531656	0.250574	0.340614
K-Nearest Neighbors	0.945371	0.471756	0.236419	0.314985
Naive Bayes	0.838919	0.204166	0.701224	0.316253



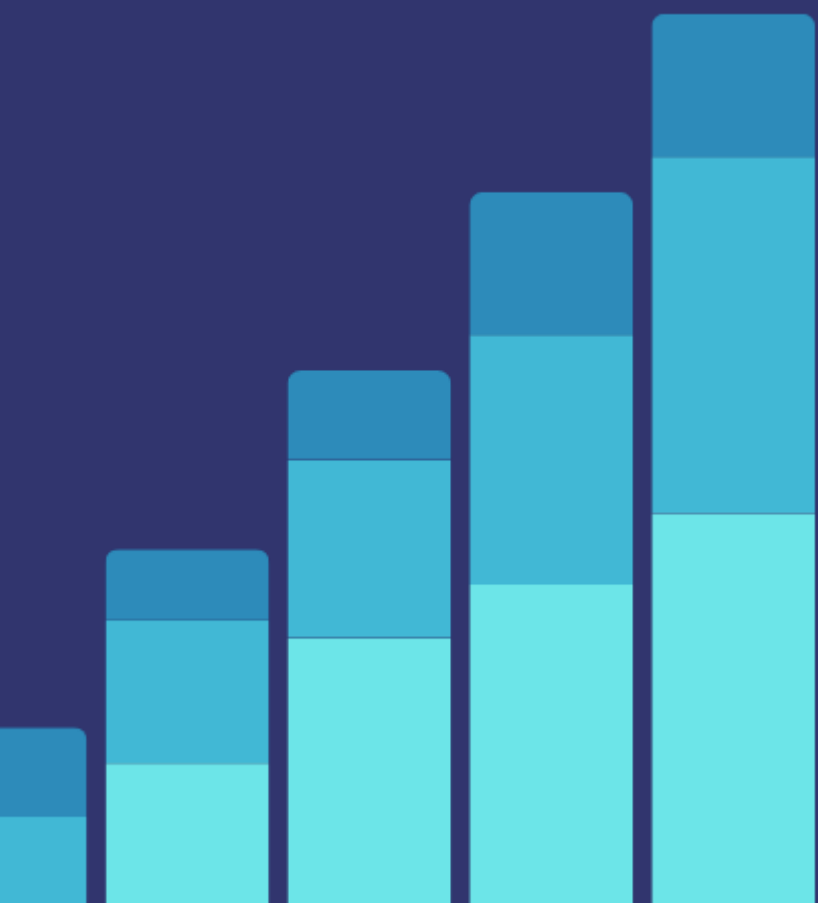
Matriz de
confusion
correspondiente
al modelo
Logistic
Regression



Matriz de
confusion
correspondiente
al modelo K-
Nearest
Neighbors
(KNN)



Matriz de
confusion
correspondiente
al modelo Naïve
Bayes



GRACIAS 😊