

# ANSIEDAD

## SEVERIDAD DE LOS ATAQUES EN PERSONAS ADULTAS



Nohelia Espinoza



## **TRASTORNO DE ANSIEDAD EN DATOS**

La ansiedad es uno de los trastornos mentales más prevalentes a nivel mundial. Según la Organización Mundial de la Salud (OMS), más de 300 millones de personas en el mundo padecen de un trastorno de ansiedad, lo que representa aproximadamente el 4% de la población mundial. Este trastorno tiene un impacto significativo en la vida diaria, no solo a nivel emocional y psicológico, sino también en la salud física, ya que los ataques de ansiedad pueden generar síntomas como taquicardia, dificultad para respirar, sudoración excesiva y mareos.

# ALCANCE Y MOTIVACIÓN

Entender los factores que influyen en la severidad de los ataques de ansiedad y desarrollar un modelo predictivo que permita prever qué tan grave puede ser un ataque en función de ciertas variables. A través de la ciencia de datos, se espera encontrar patrones en un conjunto de datos que incluye información sobre hábitos de vida, antecedentes familiares, y respuestas fisiológicas que podrían estar relacionados con la intensidad de los episodios de ansiedad.



# HIPOTESIS

1



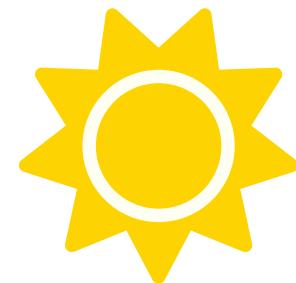
¿La severidad de los ataques de ansiedad varía según la edad y el género de los individuos?

2



¿A mayor nivel de estrés y menor calidad del sueño están relacionados con una mayor severidad de los ataques de ansiedad?

3



¿Puede que la intervención médica (medicación y terapia) reduzca la severidad de los ataques de ansiedad?

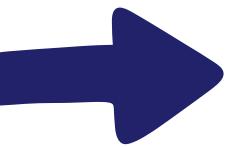
4



¿La frecuencia cardíaca y la frecuencia respiratoria durante un ataque de ansiedad ser pueden predictores significativos de la severidad de dicho ataque?

# DATA WRANGLING

CARGA DE DATASET Y PRIMERA VISTA



REVISIÓN DE NULOS,  
COLUMNAS DUPLICADAS Y  
ESTANDARIZACIÓN DE DATOS

```
RangeIndex: 12000 entries, 0 to 11999
Data columns (total 20 columns):
 #   Column           Non-Null Count Dtype  
 --- 
 0   ID               12000 non-null  int64   
 1   Age              12000 non-null  int64   
 2   Gender            12000 non-null  object  
 3   Occupation        12000 non-null  object  
 4   Sleep Hours       12000 non-null  float64 
 5   Physical Activity (hrs/week) 12000 non-null  float64 
 6   Caffeine Intake (mg/day)    12000 non-null  int64   
 7   Alcohol Consumption (drinks/week) 12000 non-null  int64   
 8   Smoking            12000 non-null  object  
 9   Family History of Anxiety      12000 non-null  object  
 10  Stress Level (1-10)          12000 non-null  int64   
 11  Heart Rate (bpm during attack) 12000 non-null  int64   
 12  Breathing Rate (breaths/min)   12000 non-null  int64   
 13  Sweating Level (1-5)         12000 non-null  int64   
 14  Dizziness            12000 non-null  object  
 15  Medication           12000 non-null  object  
 16  Therapy Sessions (per month) 12000 non-null  int64   
 17  Recent Major Life Event     12000 non-null  object  
 18  Diet Quality (1-10)         12000 non-null  int64   
 19  Severity of Anxiety Attack (1-10) 12000 non-null  int64  
 dtypes: float64(2), int64(11), object(7)
 memory usage: 1.8+ MB
```

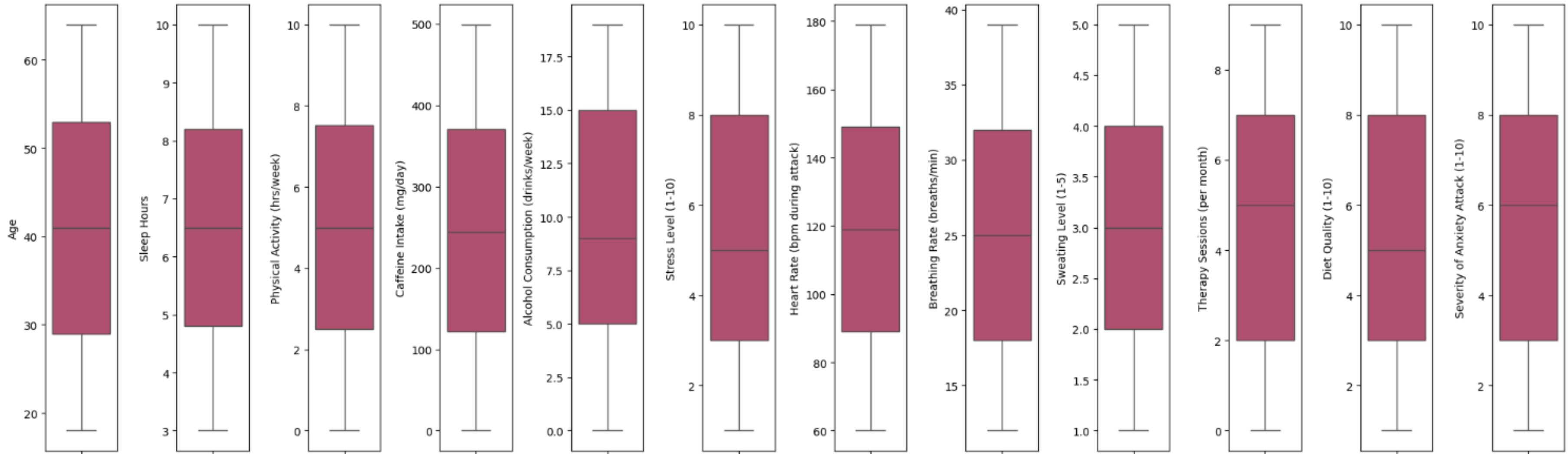
SE TIENEN 20 VARIABLES:

- 11 VARIABLES CATEGÓRICAS (5 INT Y 6 OBJECT)
- 9 VARIABLES NUMÉRICAS
- 1 VARIABLE DE ID (POSTERIORMENTE ELIMINADA)

LA VARIABLE OBJETIVO DE ESTE ESTUDIO ES "SEVERITY OF ANXIETY ATTACK (1-10)"

EN GENERAL ES UN DATASET LIMPIO, NO HAY LINEAS DUPLICADAS, VALORES NULOS NI ES NECESARIA LA ESTANDARIZACIÓN EN NINGUNA CATEGORÍA TIPO OBJECT

# REVISIÓN DE OUTLINERS

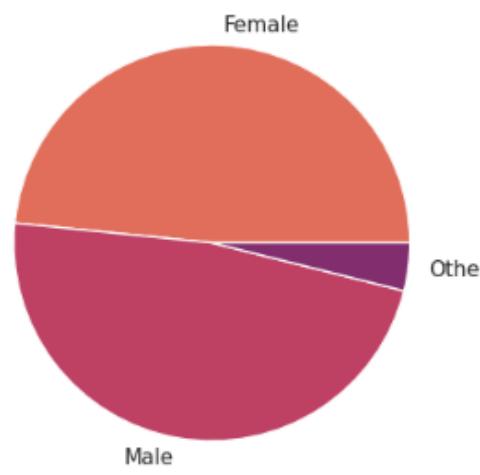


Para el uso correcto del dataset en el modelado se debe verificar si alguna variable tiene outliers que puedan afectar las métricas al momento del entrenamiento. En este caso, todo el dataset esta muy balanceado y no se tiene la necesidad de borrar líneas de datos.

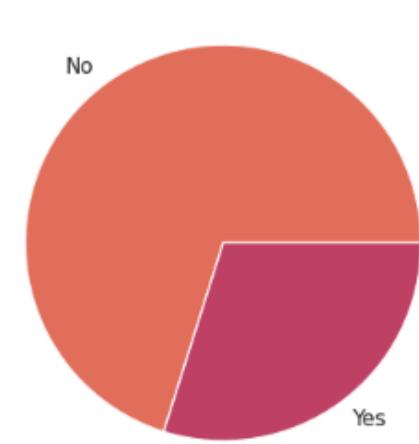
# EXPLORACIÓN DE LOS DATOS Y DISTRIBUCIÓN

## Distribución de las variables categóricas tipo object

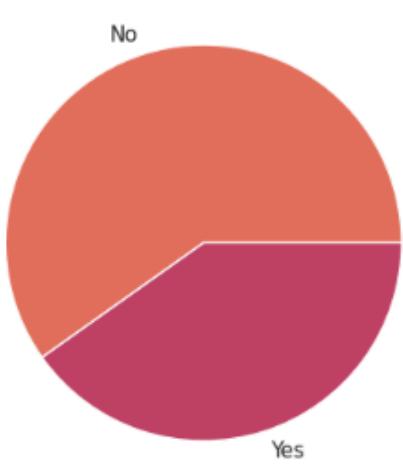
Distribución de 'Gender'



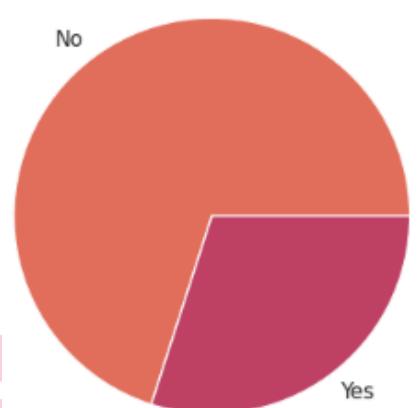
Distribución de 'Smoking'



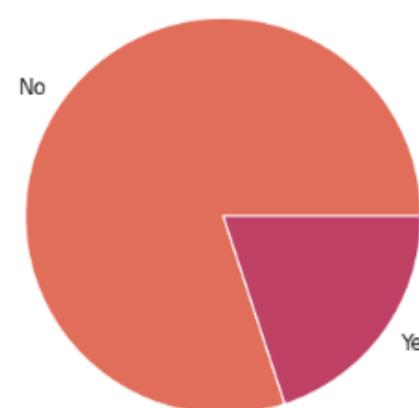
Distribución de 'Family History of Anxiety'



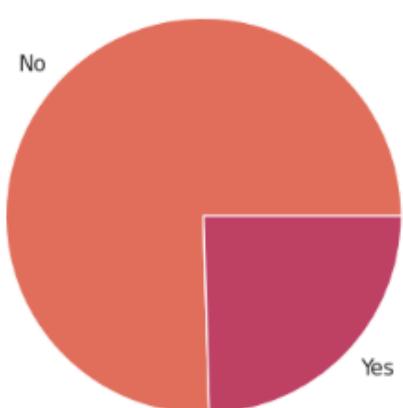
Distribución de 'Dizziness'



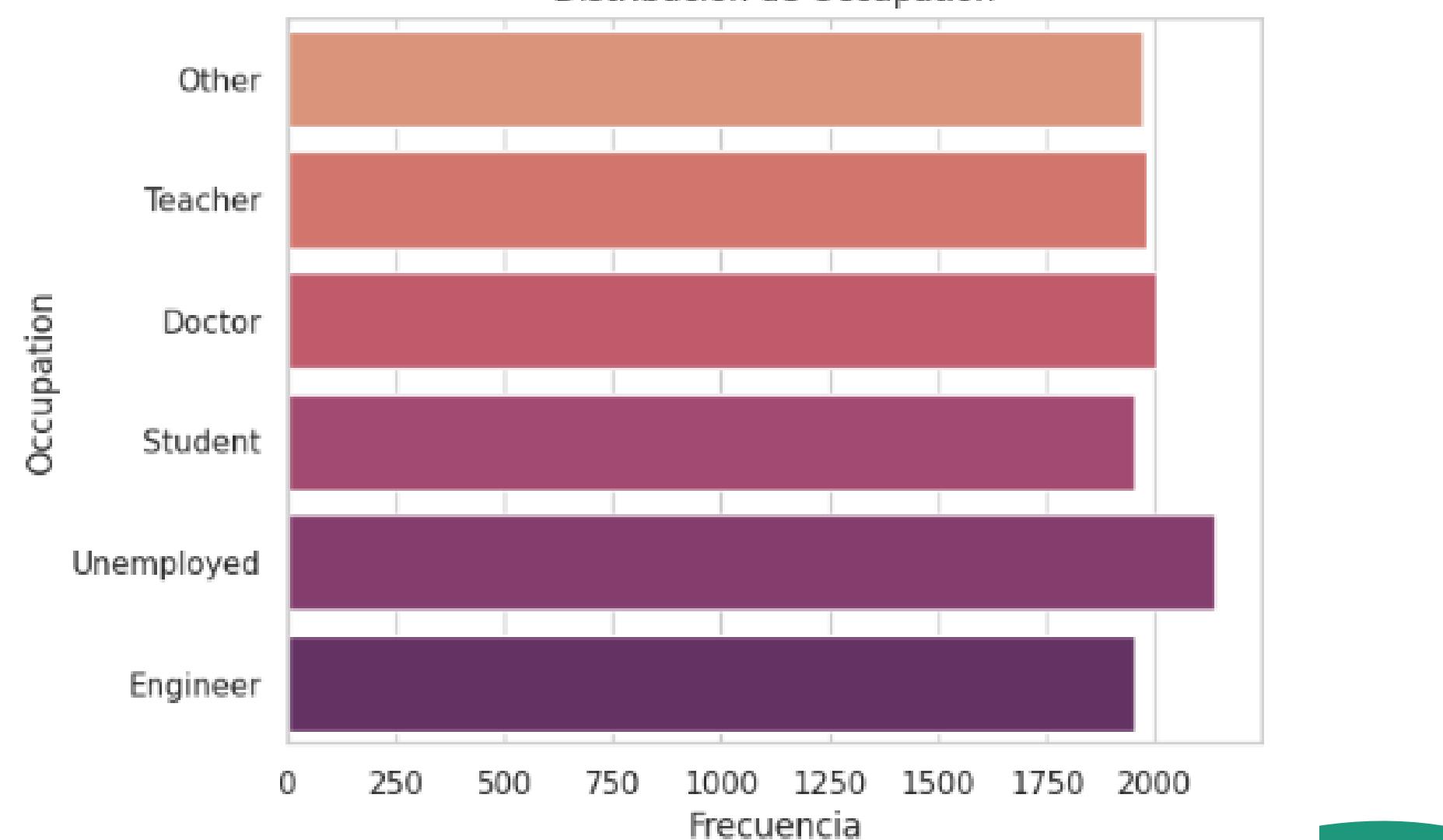
Distribución de 'Medication'



Distribución de 'Recent Major Life Event'

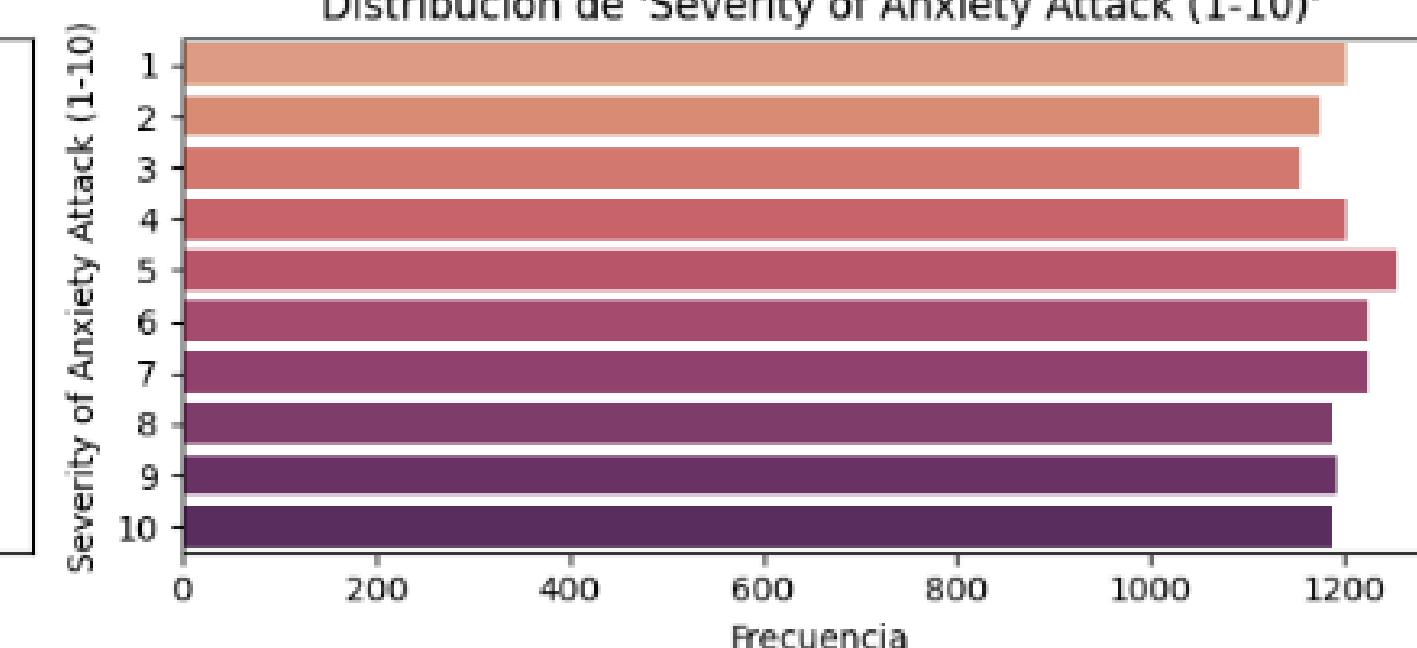
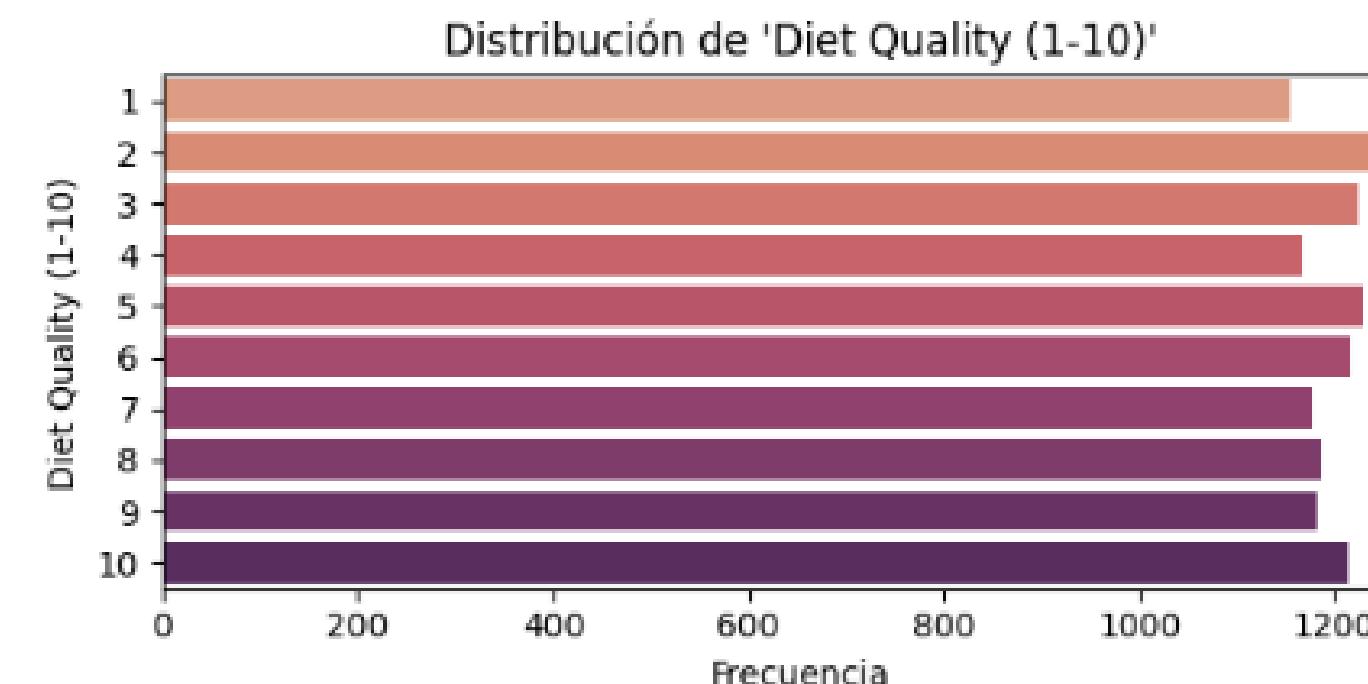
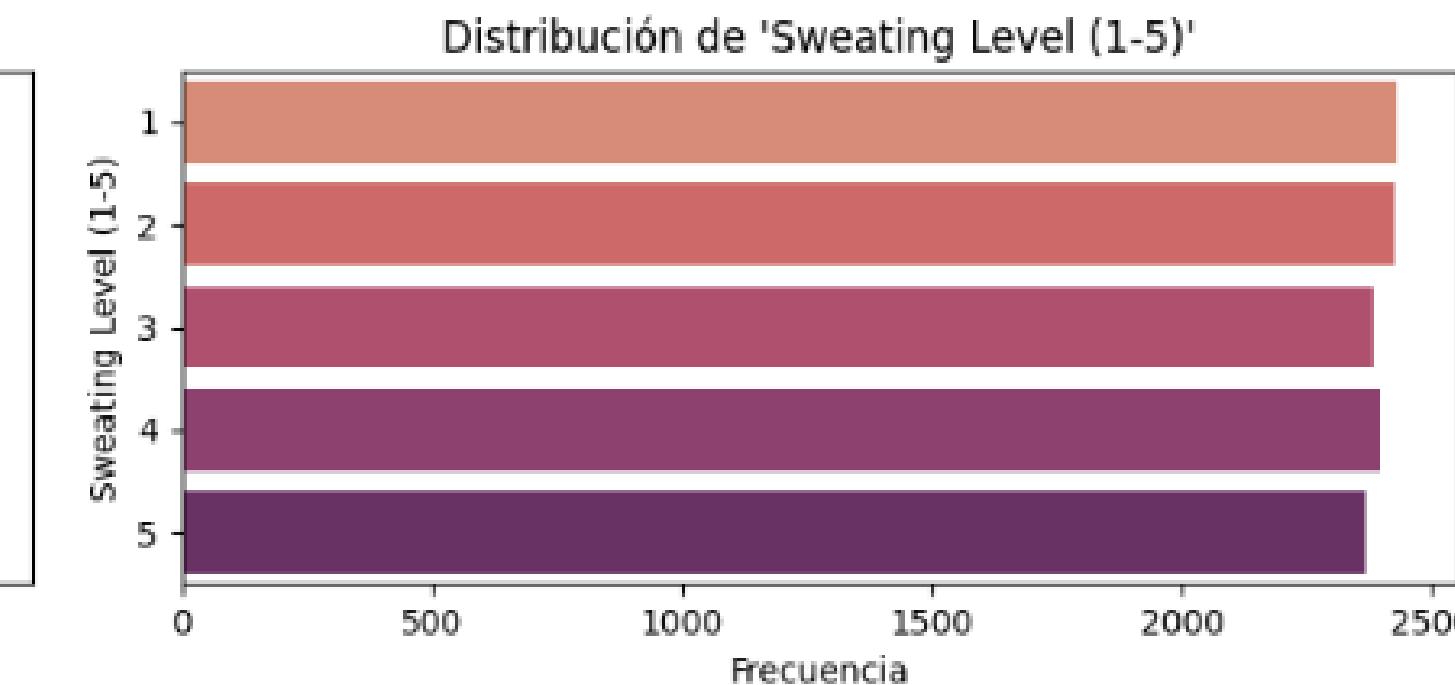
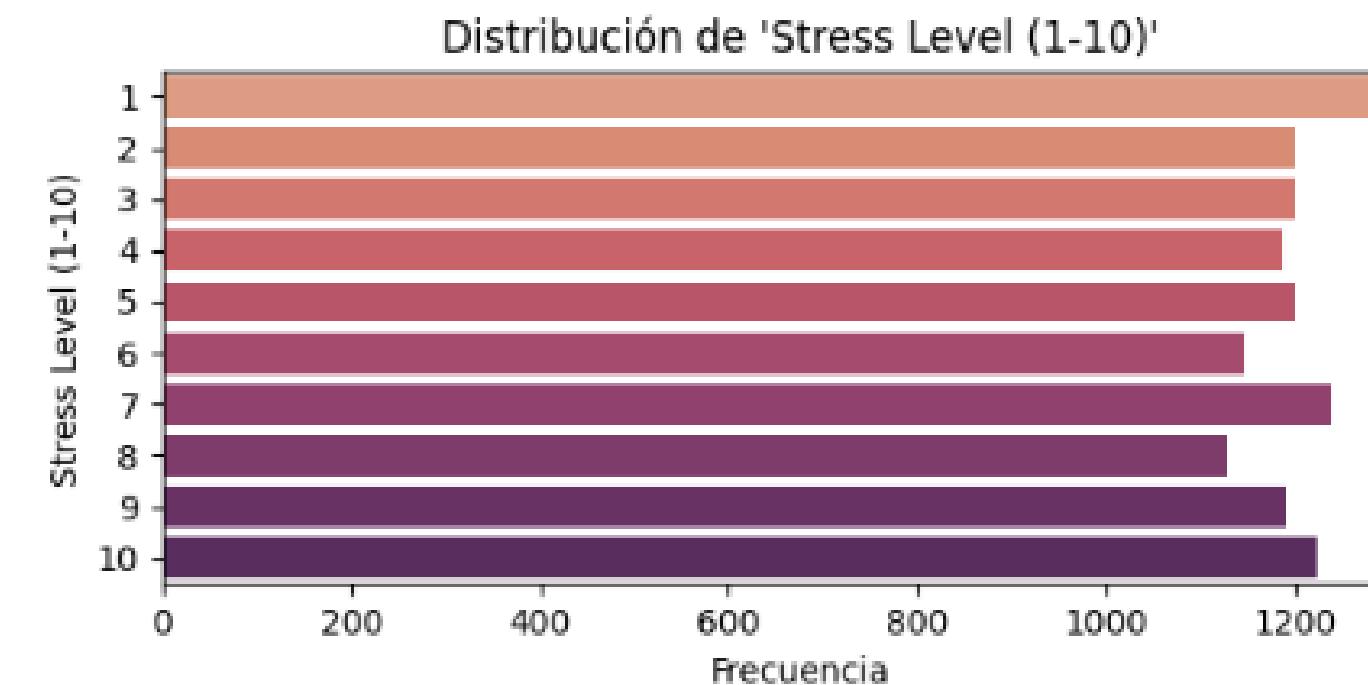


Distribución de Occupation



# EXPLORACIÓN DE LOS DATOS Y DISTRIBUCIÓN

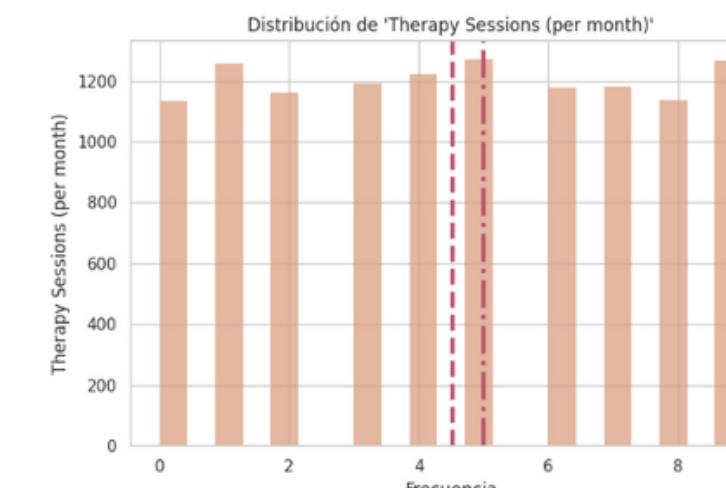
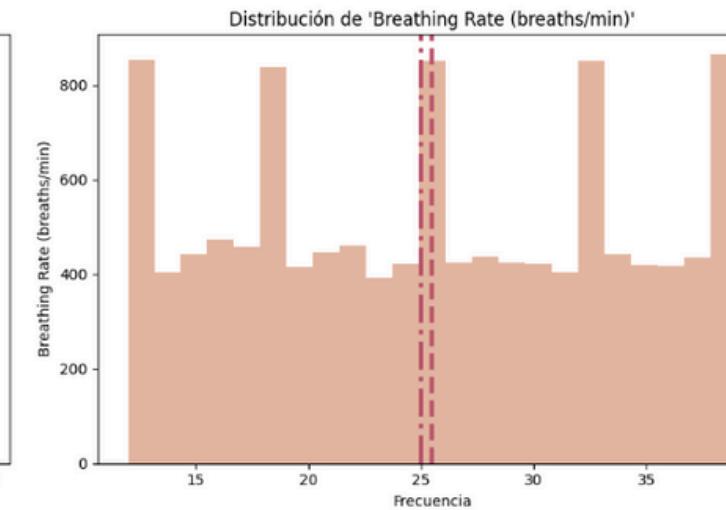
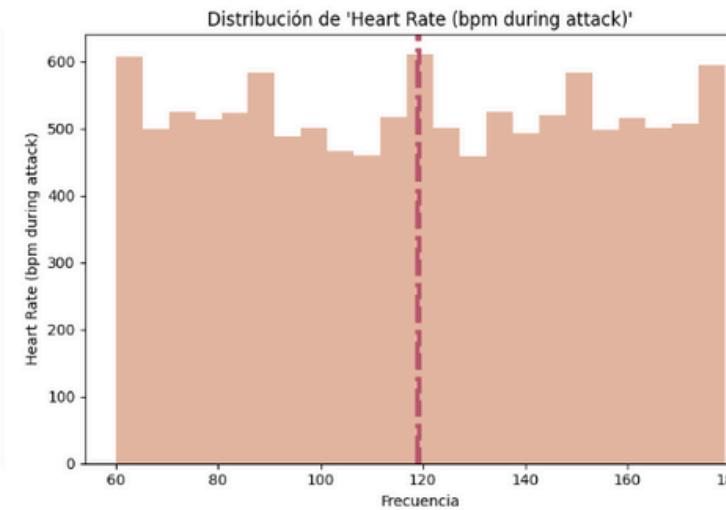
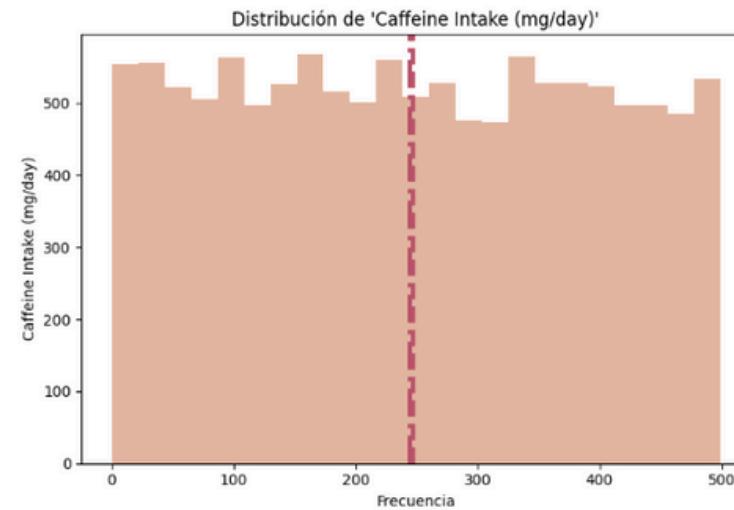
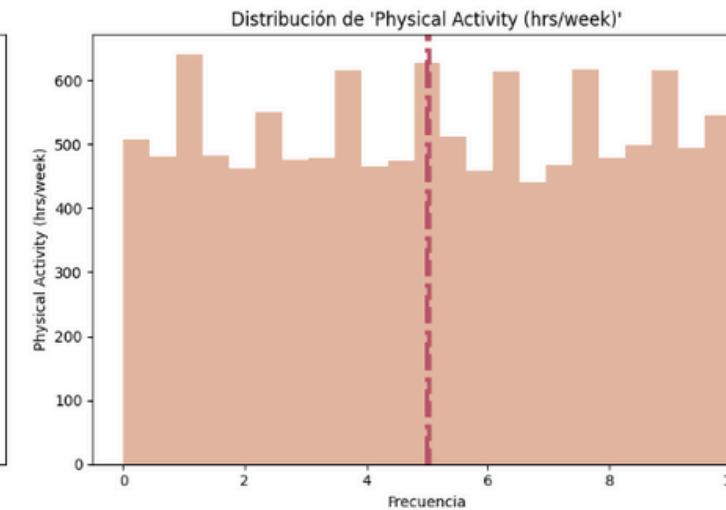
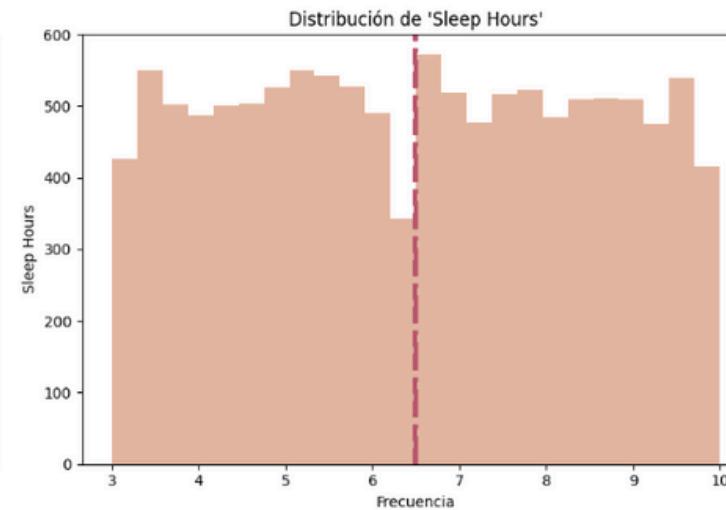
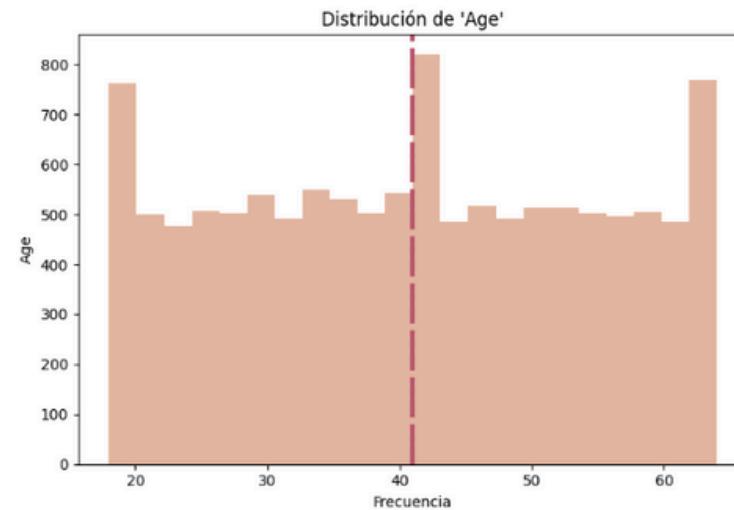
## Distribución de las variables categóricas tipo numérica



Se observa muy buena distribución entre las variables categóricas numéricas, incluida la variable objetivo.

# EXPLORACIÓN DE LOS DATOS Y DISTRIBUCIÓN

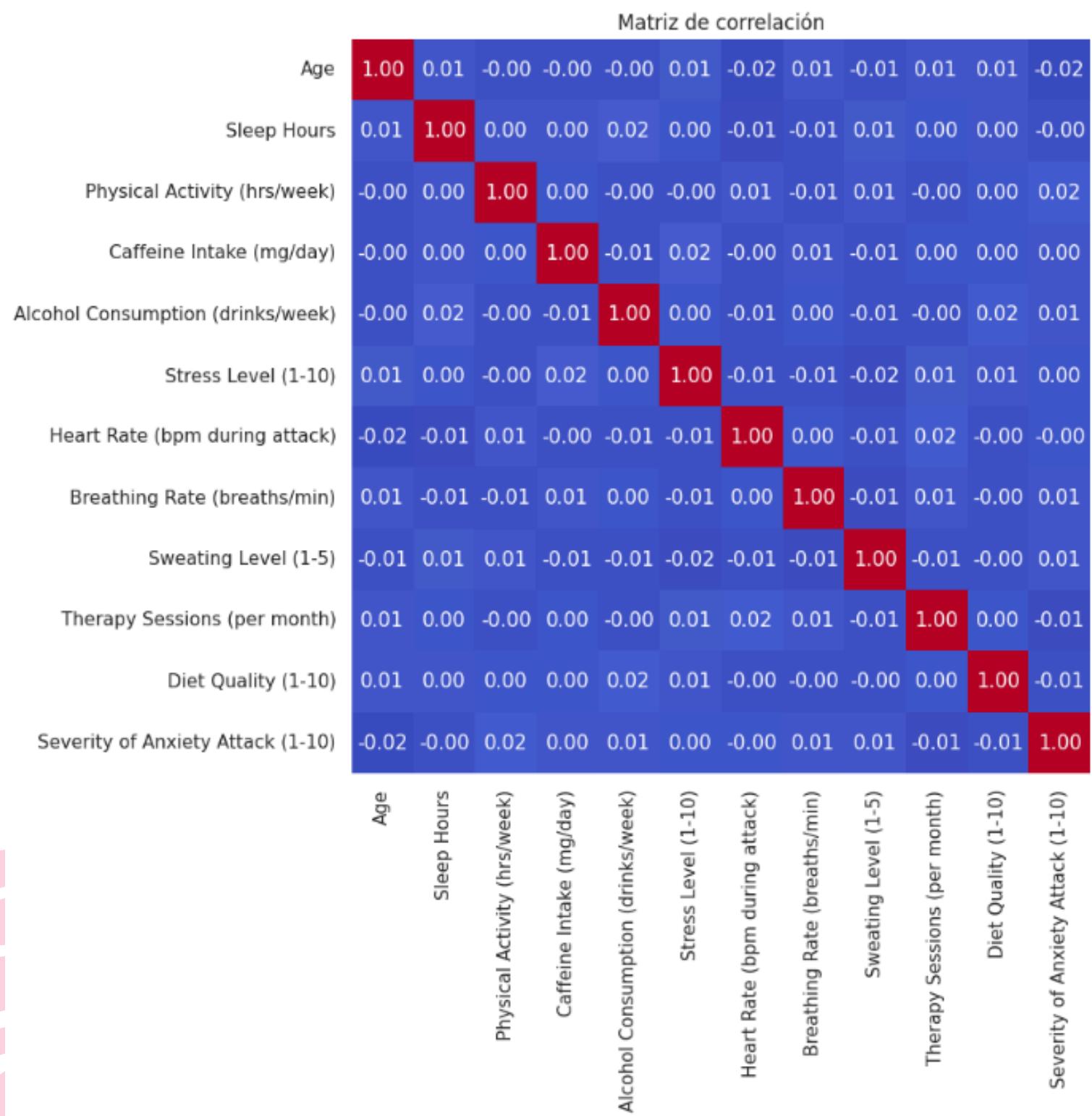
## Distribución estadística de las variables numéricas



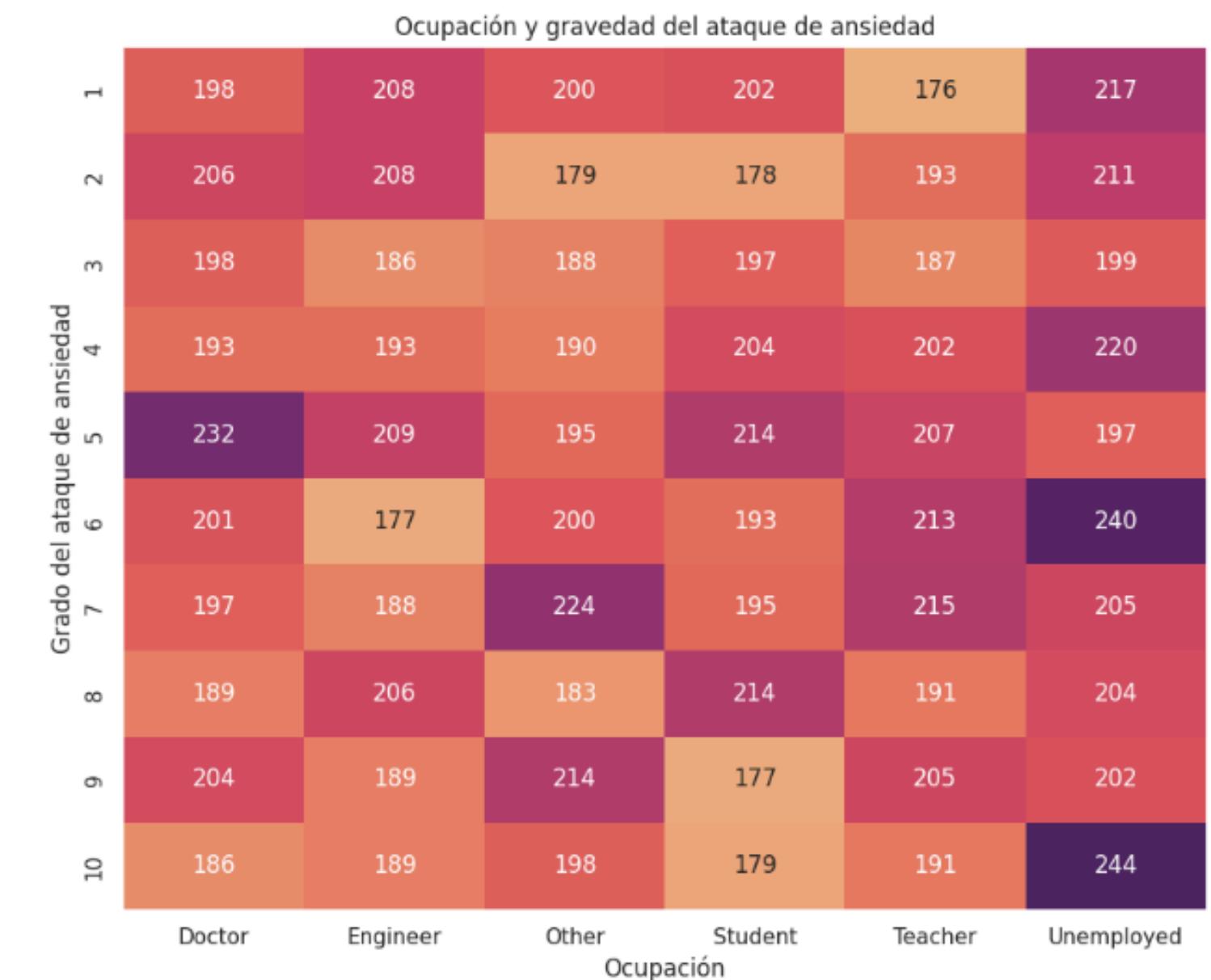
TODA LA DISTRIBUCIÓN  
ESTADÍSTICA ES UNIFORME, DE  
ASPECTO PLANICURTICO.

# EXPLORACIÓN DE LOS DATOS Y DISTRIBUCIÓN

## Correlación de variables numéricas

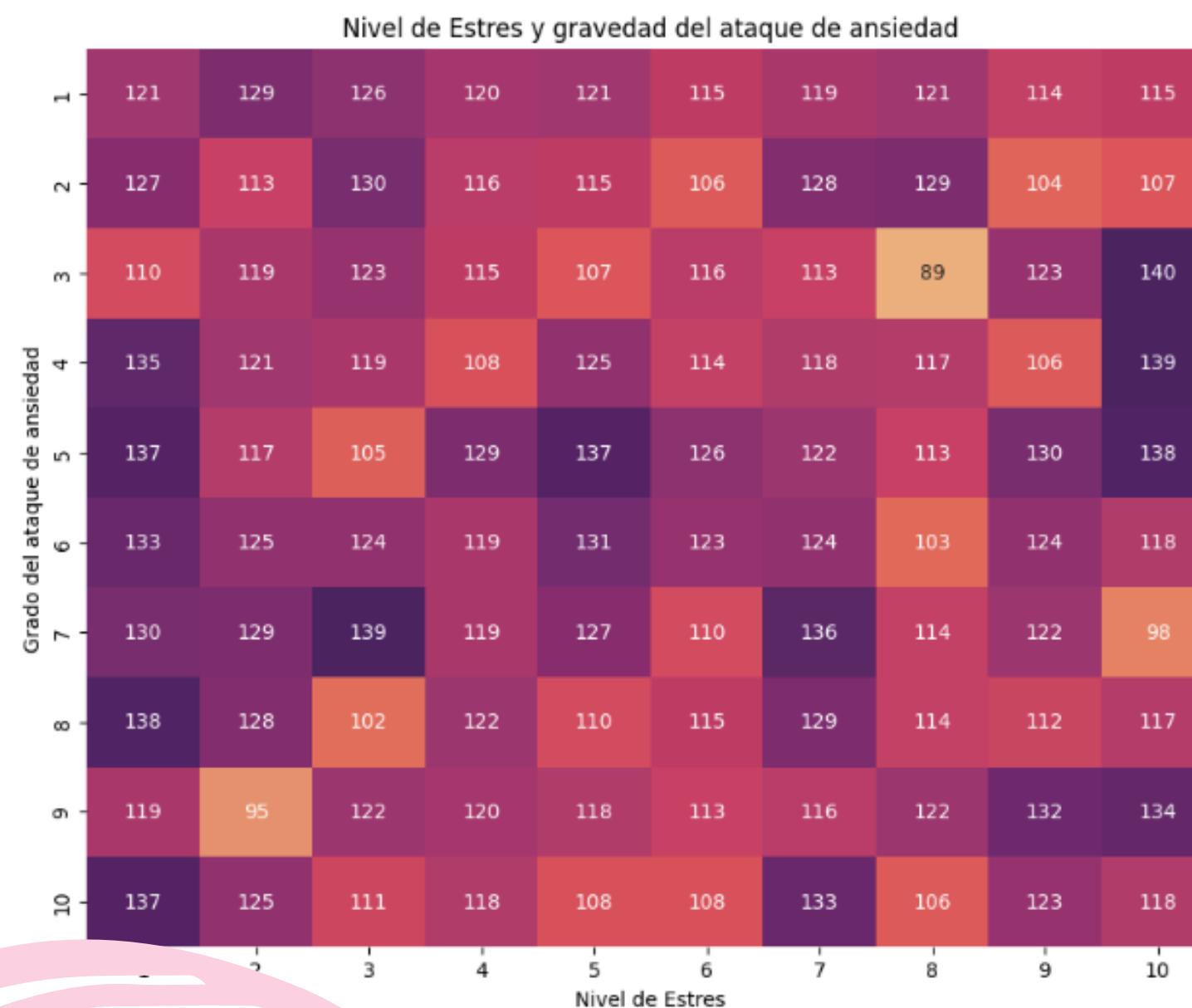


## Correlación entre ocupación y severidad

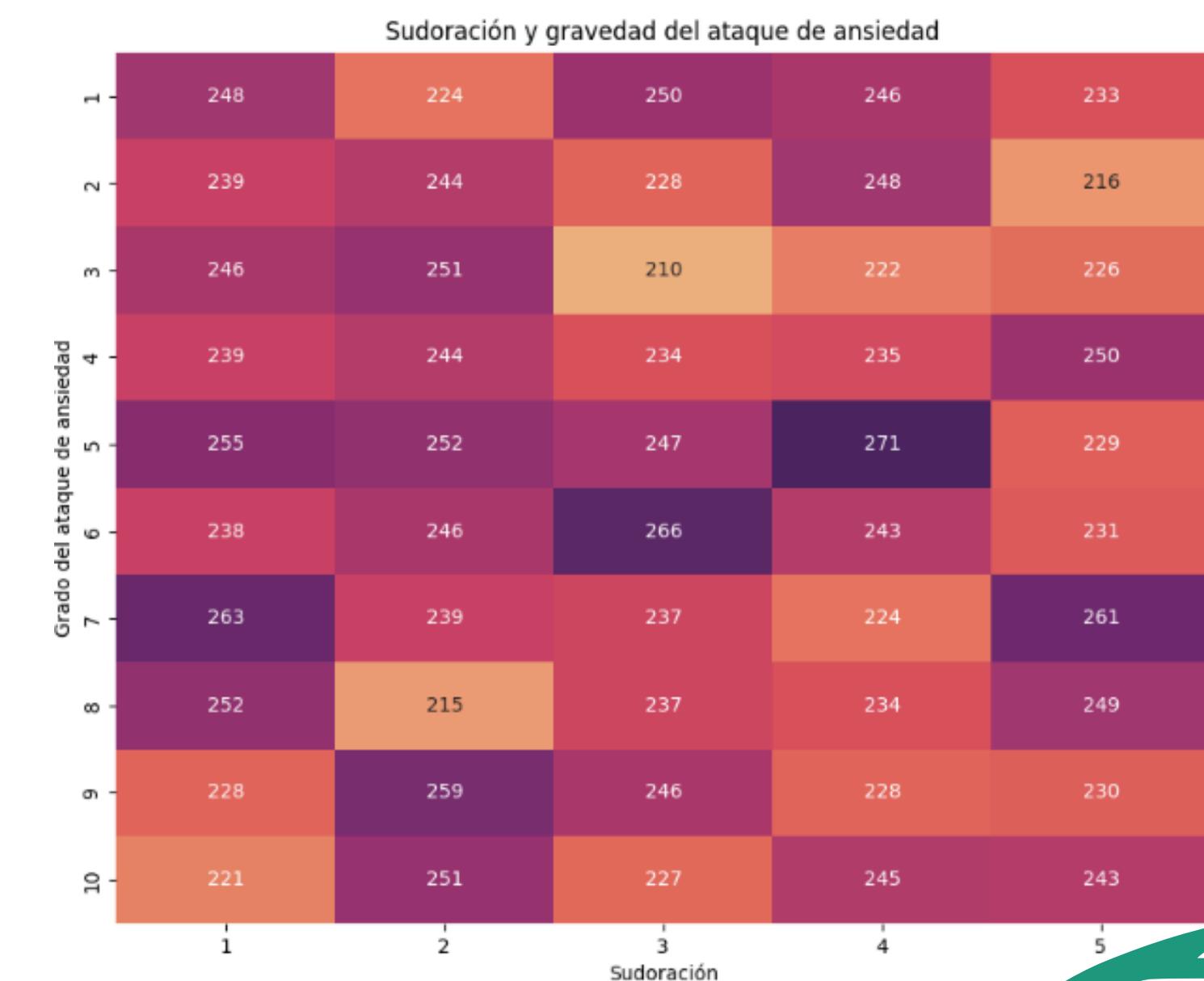


# EXPLORACIÓN DE LOS DATOS Y DISTRIBUCIÓN

## Correlación entre nivel de estrés y severidad

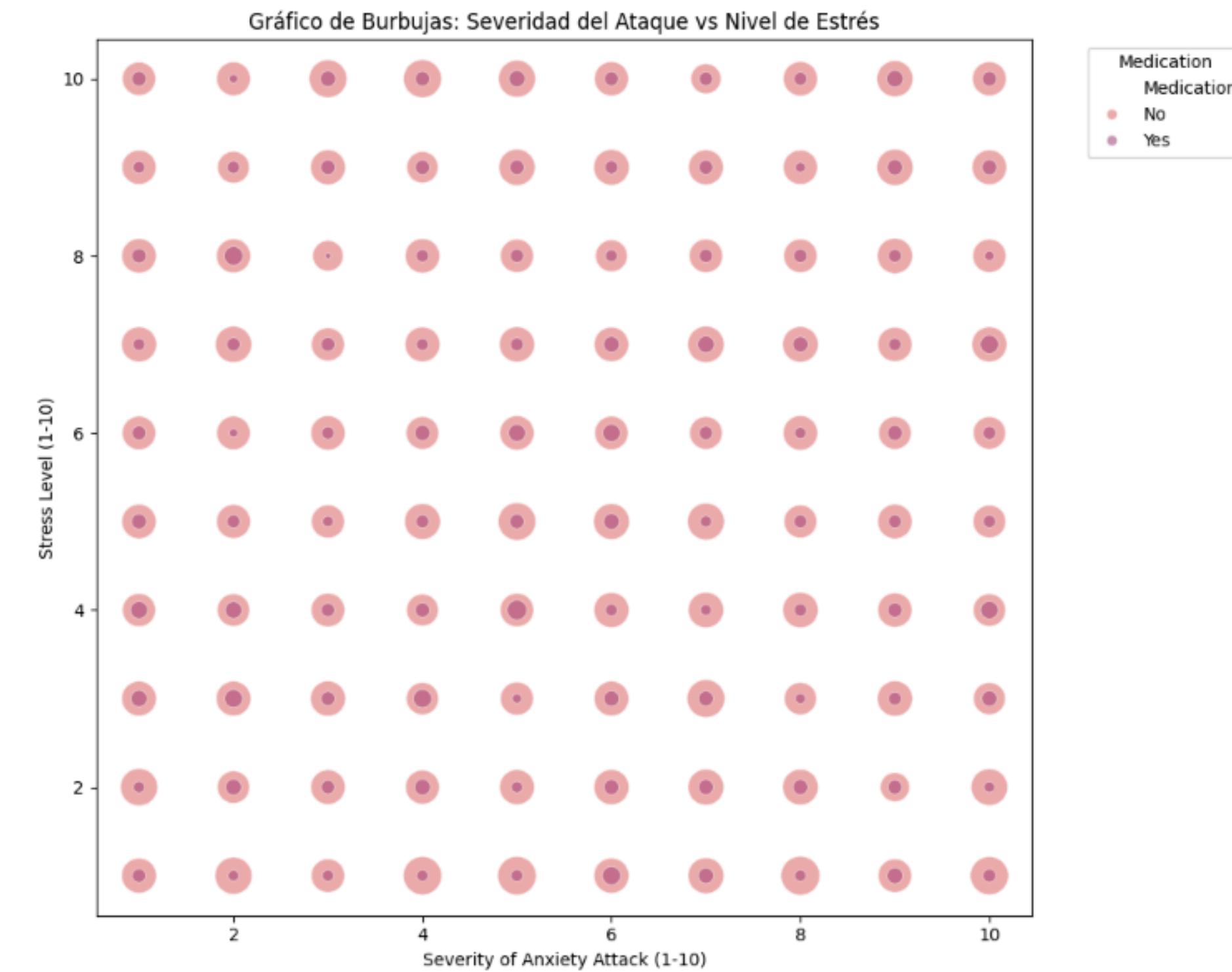


## Correlación entre sudoración y severidad



# EXPLORACIÓN DE LOS DATOS Y DISTRIBUCIÓN

Correlación entre Nivel de estrés y el grado de severidad por presencia o ausencia de medicación



Tras analizar diversos gráficos bivariados y multivariados, no se observa una correlación clara entre las variables. En el contexto de la salud mental, este resultado es esperable, ya que los trastornos psicológicos no pueden definirse únicamente a partir de pruebas fisiológicas objetivas. Su diagnóstico y evolución dependen de una compleja interacción de factores internos, como la genética y la química cerebral, y factores externos, como el entorno social y las experiencias de vida.

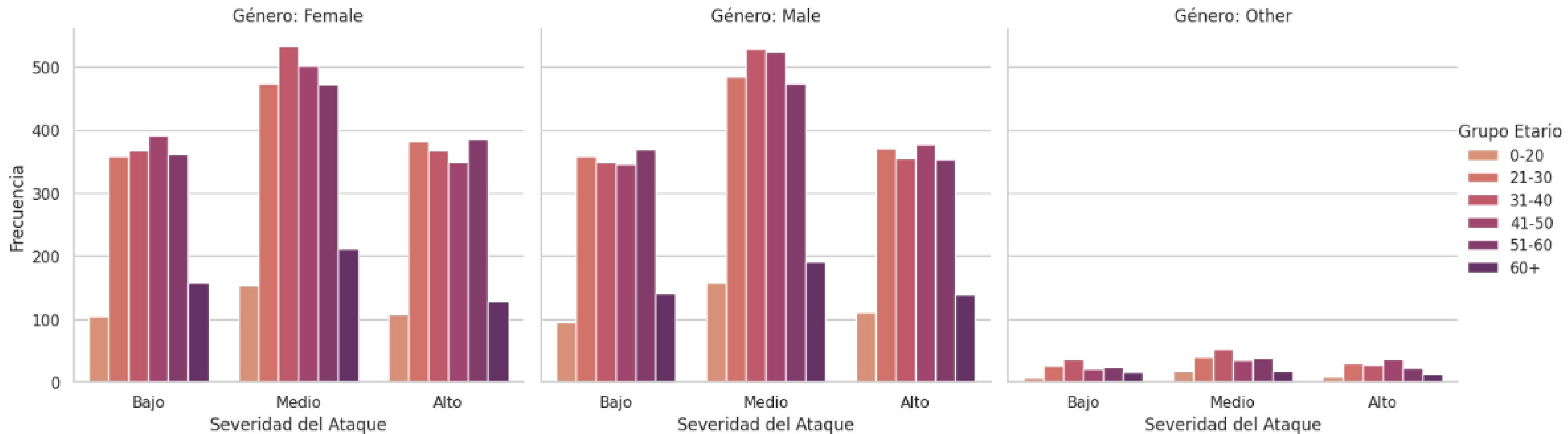


# **¡HABLEMOS DE LAS HIPOTESIS INICIALES!**



# ¿LA SEVERIDAD DE LOS ATAQUES DE ANSIEDAD VARÍA SEGÚN LA EDAD Y EL GÉNERO DE LOS INDIVIDUOS?

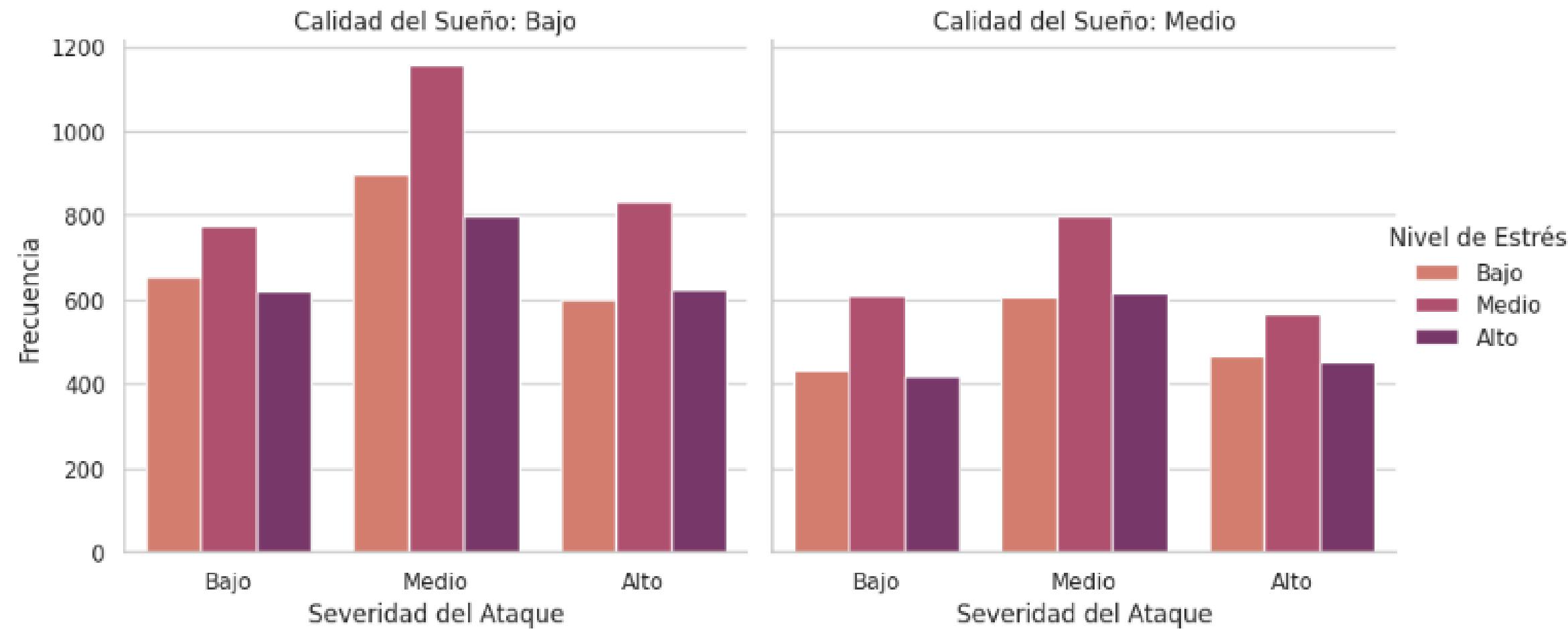
Distribución Etaria por Severidad del Ataque y Género



Entre los datos etarios no se encuentra diferencia sumamente variable entre la severidad de los ataques de ansiedad solo excepto, entre las personas menores a 20 años y mayores a 60 años pero, esto es debido a la muestra de la población de estudio.

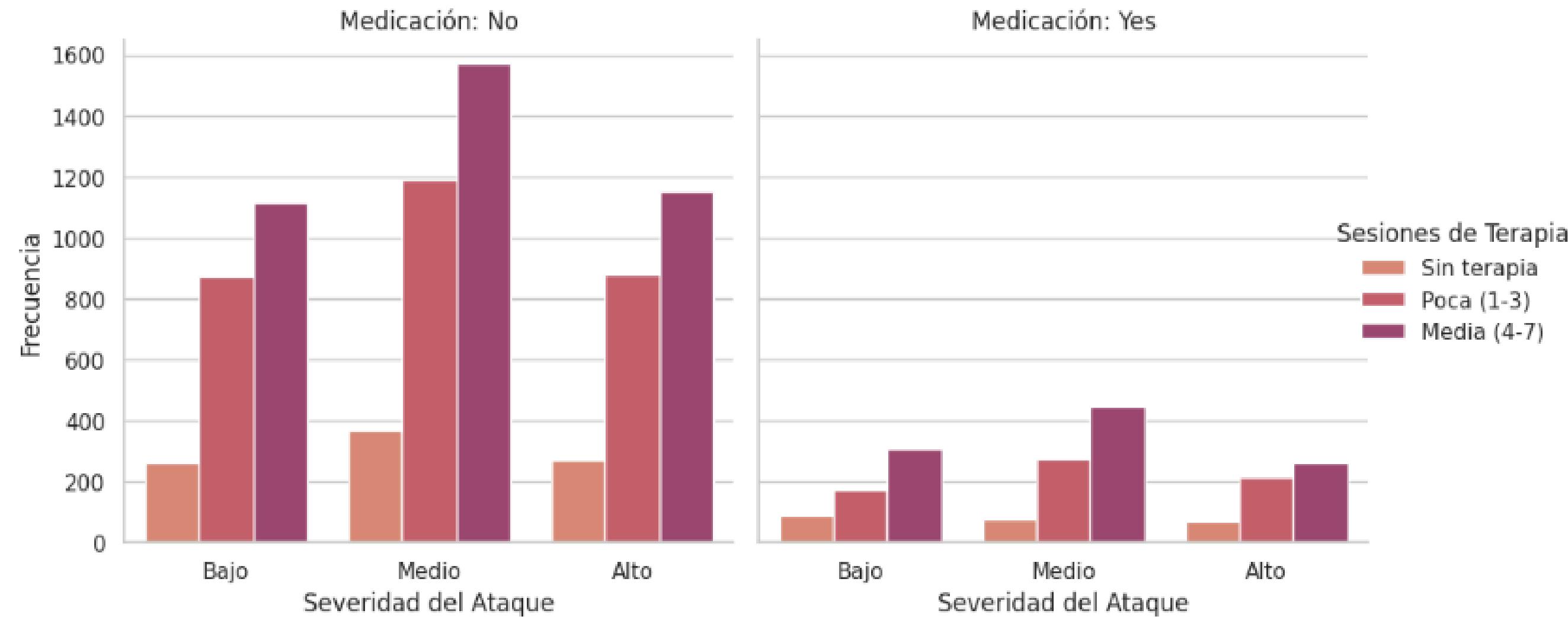
# ¿A MAYOR NIVEL DE ESTRÉS Y MENOR CALIDAD DEL SUEÑO ESTÉN RELACIONADOS CON UNA MAYOR SEVERIDAD DE LOS ATAQUES DE ANSIEDAD?

Relación entre Severidad, Nivel de Estrés y Calidad de Sueño



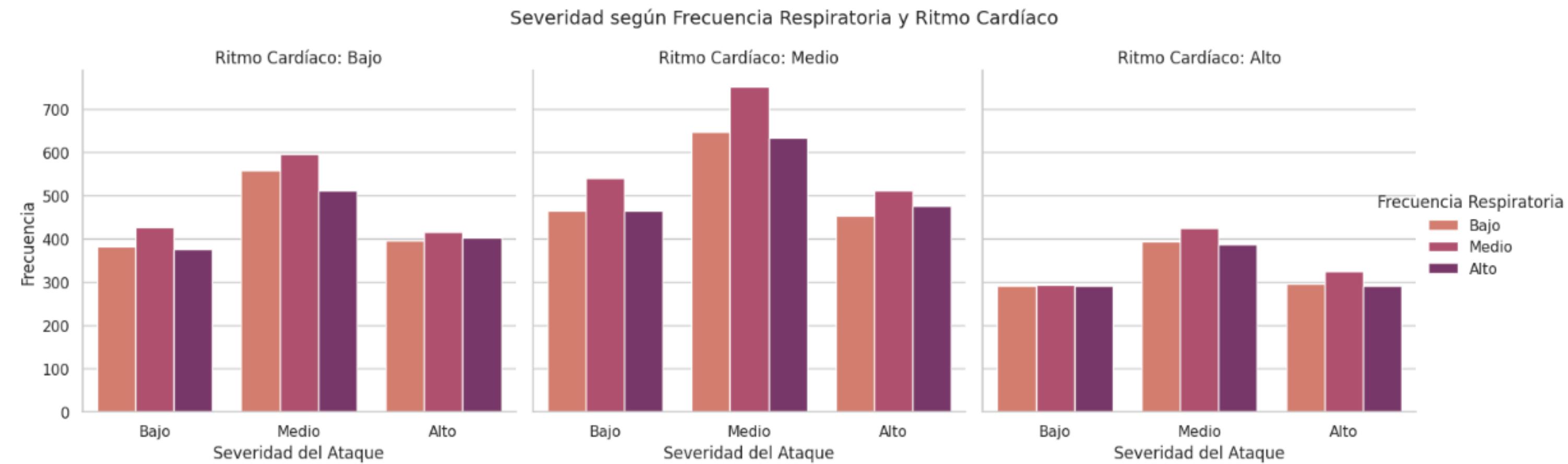
La calidad de sueño afecta a la severidad de los ataques de ansiedad muy por encima del efecto que tiene el nivel de estrés.

# ¿PUEDE QUE LA INTERVENCIÓN MÉDICA (MEDICACIÓN Y TERAPIA) REDUZCA LA SEVERIDAD DE LOS ATAQUES DE ANSIEDAD?



Dentro de los datos estudiados, muchas líneas no tienen medicación, sin embargo, dentro de la muestra que no tienen medicación, se observan los ataques medios principalmente aquellos que tienen mayor cantidad de sesiones terapéutica.

# ¿LA FRECUENCIA CARDÍACA Y RESPIRATORIA DURANTE UN ATAQUE DE ANSIEDAD PUEDEN SER PREDICTORES SIGNIFICATIVOS DE LA SEVERIDAD DE DICHO ATAQUE?



No se observa variación de la frecuencia en cada uno de los rangos de la severidad vs las vitales fisiológicas

# ALGORITMOS DE MACHINE LEARNING



# ALGORITMOS UTILIZADOS

Regresión Logística

XGBoost

MLP Classifier

Árbol de decisión

Optimización con  
GridSearch

SVM

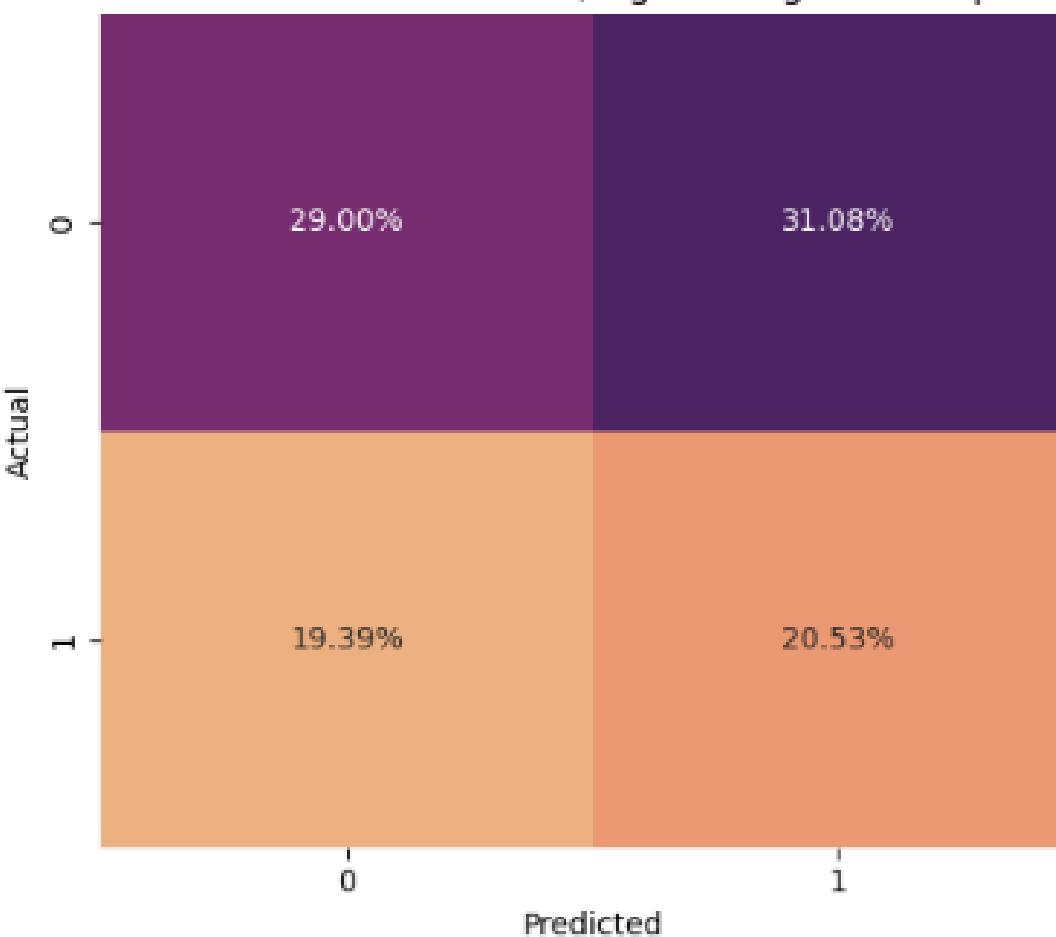
Random Forest

Optimización con  
Randomized Search

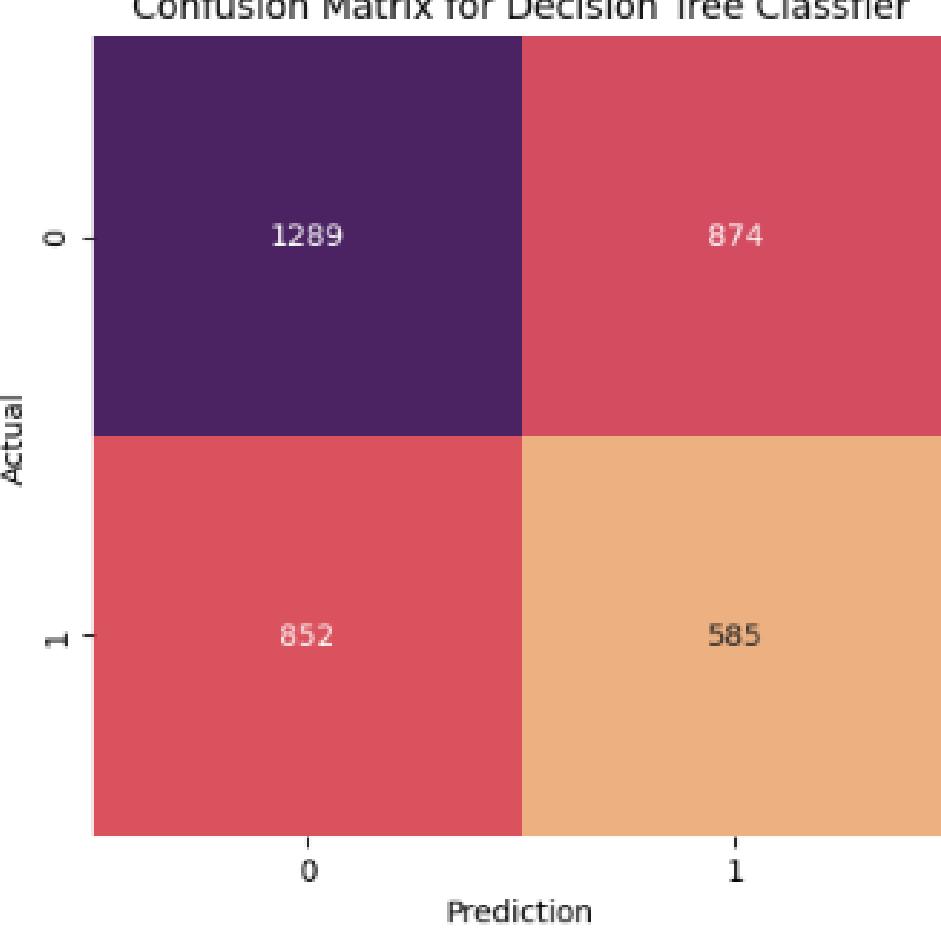
KNN

# MATRICES DE CONFUSIÓN

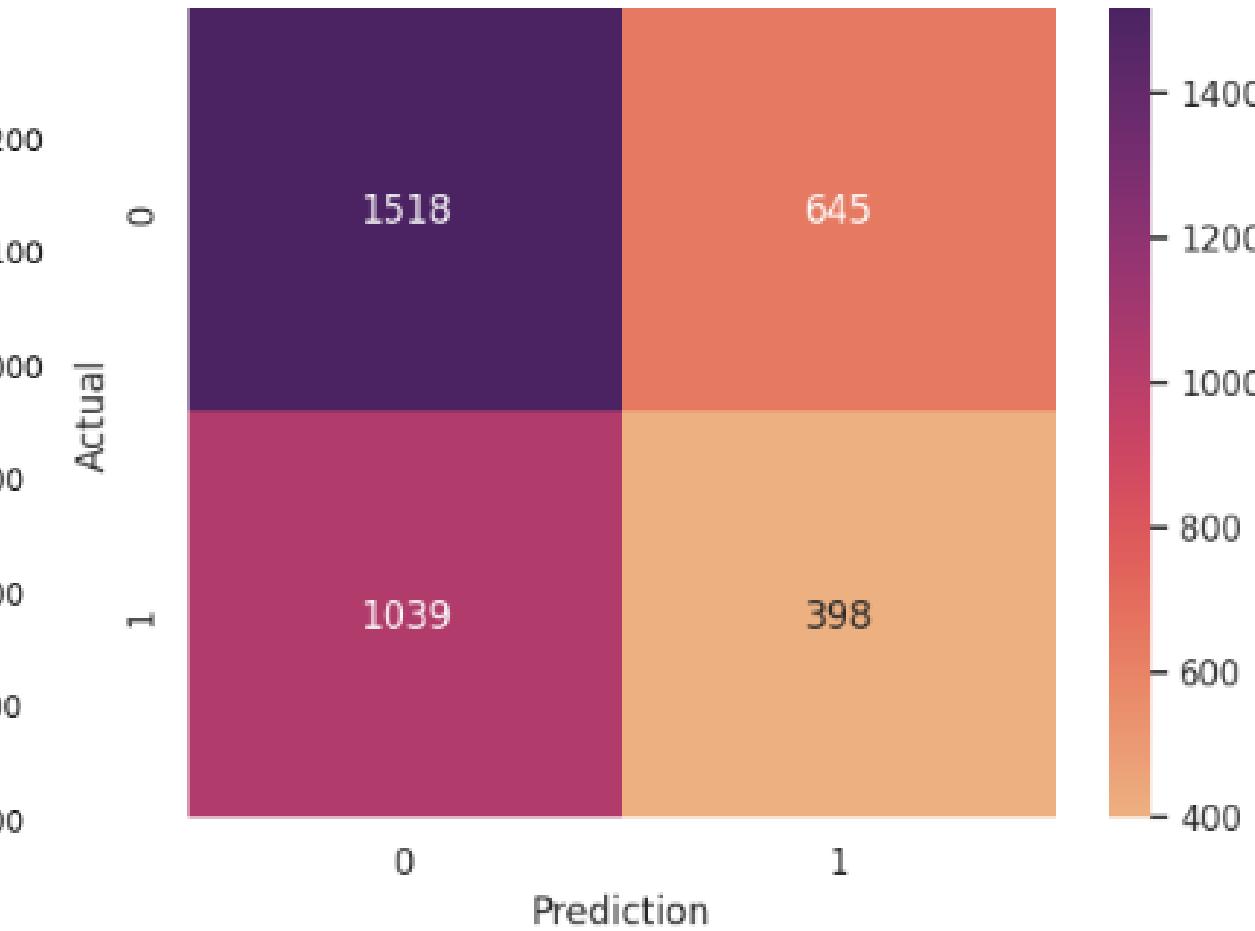
Normalized Confusion Matrix (Logistic Regression Optimized)



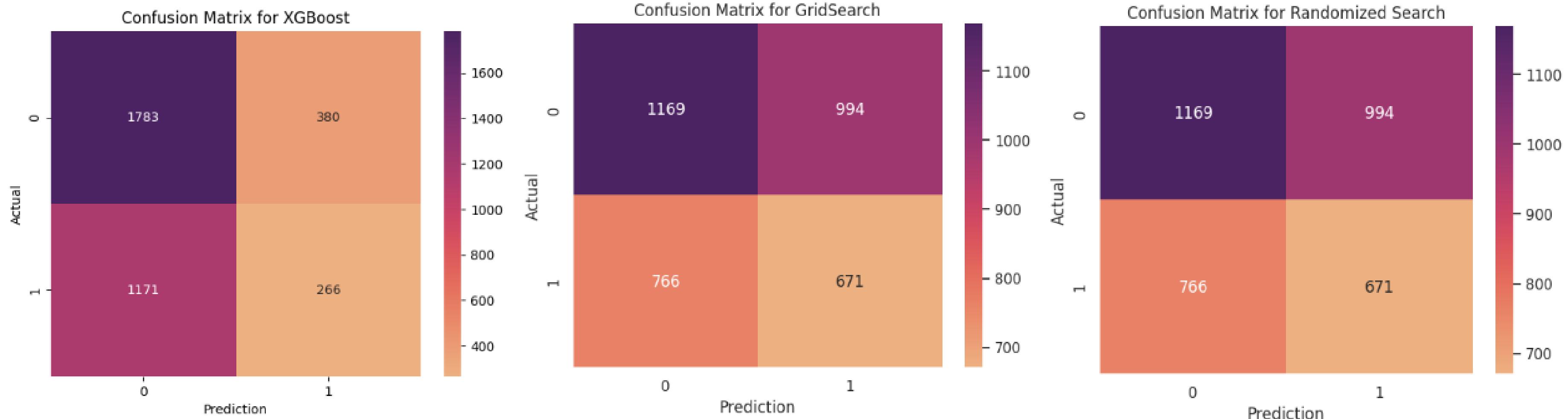
Confusion Matrix for Decision Tree Classifier



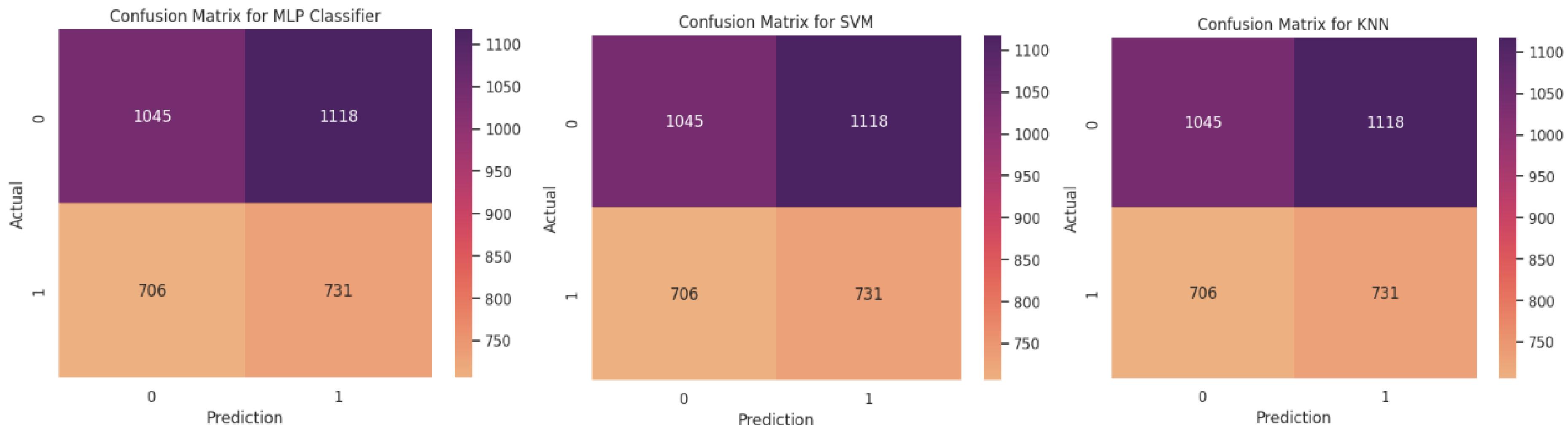
Confusion Matrix for Random Forest



# MATRICES DE CONFUSIÓN



# MATRICES DE CONFUSIÓN



# REPORTE DE CLASIFICACIÓN AVG

	Regresión Logística	Árbol de decisión	Random Forest	XGBoost	Optimización con GridSearch	Optimización con Randomized Search	MLP Classifier	SVM	KNN
Precisión	0.52	0.52	0.51	0.53	0.52	0.52	0.52	0.52	0.52
Recall	0.5	0.52	0.52	0.57	0.51	0.51	0.49	0.49	0.49
F1-score	0.5	0.52	0.51	0.52	0.52	0.52	0.5	0.5	0.5

Si bien, todos los modelos muestran accuracy similares, el mejor para la severidad baja-media en ataques de ansiedad sería XGBoost con un recall de 0.82 y f1-score de 0.7

Luego de implementar varios algoritmos y manejo de hiperparametros no fue posible conseguir una precisión aceptable donde se pueda establecer el uso del modelo. Algunos puntos a tomar en cuenta para este resultado son los siguientes:

- Veracidad de los datos ya que se observo distribución planicurtica y correlación nula entre variables, incluso al evaluar las hipótesis.
- La salud mental no ha sido suficientemente explorada incluso en la actualidad donde un porcentaje muy alto de la población padece algún trastorno psicológico o psiquiátrico y se deben tomar en cuenta muchas variables que afectan el estado de la persona para un diagnostico correcto.

