

PRACTICA 1.2: Manipulación de datos

Paso 1: Carga del Dataset

¿Cuántos registros y columnas tiene el dataset?

El dataframe tiene: 1200 registros y 21 columnas.

¿Qué tipo de datos predominan?

Los datos que predominan son enteros (integers).

Paso 2: Exploración Inicial

¿Existen columnas que podrían ser categóricas pero aparecen como tipo object?

Si, son: 'Gender', 'Education_Level', 'Employment_Status', 'Medication_Use', 'Substance_Use', estas se han usado como tipo 'object'.

¿Hay valores que parecen inconsistentes o fuera de rango?

Los valores de la mayoría de las columnas parecen estar dentro de los rangos esperados, con la excepción de Sleep_Hours, la cual tiene un valor mínimo de 2 horas de sueño.

Paso 3: Limpieza de Datos

¿Qué porcentaje del dataset contiene valores nulos?

0

¿Qué estrategia decidiste aplicar para tratar los valores nulos y por qué?

Ya que los valores nulos en las columnas Medication_Use y Substance_Use son bastante significativos, la imputación me ha parecido una buena estrategia. He imputado con el valor 'Unknown'.

Paso 4: Análisis Estadístico

¿Cuál de los tres puntajes tiene mayor variabilidad?

La columna con mayor variabilidad es: Anxiety_Score con una desviación estándar de 5.911137901685593

¿Qué observas sobre la diferencia entre media y mediana en cada caso? ¿Qué te dice esto sobre la distribución?

Para Anxiety_Score, tiene un valor más bajo en la media que en la mediana, la diferencia es pequeña, sin embargo podemos decir que la distribución es prácticamente simétrica. En Depression_Score, la diferencia más grande indica una distribución sesgada hacia la derecha. Esto puede significar que hay un pequeño número de personas con puntajes de depresión muy altos, lo que afecta la media. En el caso de Stress_Level, se percibe una diferencia casi nula entre la media y la mediana, sabemos que la distribución es casi simétrica, lo que sugiere que la mayoría de los registros tienen puntajes de estrés similares.

Paso 5: Filtrado de Datos

¿Qué porcentaje del total presenta niveles altos de ansiedad?

Número de personas con alta ansiedad: 0

¿Existe una relación visible entre las horas de sueño y el nivel de estrés?

Las personas que duermen menos de 5 horas tienen el nivel de estrés más alto, seguidas de las que duermen entre 5 y 7 horas, y por último las que duermen más de 7 horas. La variabilidad en los niveles de estrés es bastante consistente entre los tres grupos, lo que sugiere que el número de horas de sueño influye ligeramente en el nivel de estrés, pero la relación no es muy fuerte.

Paso 6: Ordenamiento

¿Qué características comunes encuentras en estas 10 personas?

Todas las personas tienen el puntaje más alto posible de satisfacción de vida, el cual es de 9, lo que sugiere que se sienten muy satisfechas con su vida en general. De igual manera, todas estas personas tienen el puntaje más bajo de soledad, que es de 1, lo que significa que se sienten menos solas y están bien acompañadas o apoyadas.

¿Qué hipótesis podrías plantear sobre el vínculo entre satisfacción y soledad?

La satisfacción de vida y la soledad parecen estar conectadas por el apoyo social. Las personas con mayor apoyo social tienden a tener menos soledad y mayor satisfacción de vida, lo que sugiere que las interacciones sociales de calidad son un factor crucial. Incluso cuando la actividad física no es alta, el sentimiento de estar rodeado de personas de apoyo parece ser clave para reducir la soledad y aumentar la satisfacción.

Paso 7: Guardado de Resultados

¿Cuántas personas cumplen con ambas condiciones?

Ninguna, el resultado sugiere que ninguna persona con bajo estrés está recibiendo terapia.

¿Qué conclusiones puedes hacer sobre la relación entre terapia y nivel de estrés en este conjunto?

Los resultados indican que las personas con buen bienestar emocional no sienten la necesidad de tratamiento terapéutico. Por otro lado, las personas con estrés alto podrían ser las que buscan más frecuentemente terapia, lo que sugiere que la terapia está más asociada con altos niveles de estrés. También se considera un posible sesgo en los datos, ya que si pocas personas reciben terapia o la mayoría con bajo estrés no lo hace, esto podría ser una limitación de la información que se tiene. Finalmente, es posible que simplemente no haya muchas personas que cumplan ambas condiciones a la vez en la muestra.

Visualización Opcional

¿Hay diferencias notables entre los géneros en cuanto a autoestima?

Aunque los hombres y mujeres tienen una autoestima relativamente alta, las personas no binarias y aquellas que se identifican como "Other" parecen estar enfrentando desafíos adicionales que podrían estar afectando negativamente su autoestima.

¿Qué factores del dataset podrían explicar estas diferencias?

Factores como el apoyo social, el estrés, el nivel educativo, y las condiciones de salud pueden estar contribuyendo a las diferencias observadas. Además, las presiones sociales y normas de género pueden jugar un papel crucial, especialmente en los grupos con menor autoestima.