



BENEMÉRITA UNIVERSIDAD
AUTÓNOMA DE PUEBLA

FACULTAD DE CIENCIAS DE LA COMPUTACIÓN

REPORTE FINAL
INTRODUCCIÓN A
LA CIENCIA DE
DATOS

NRC: 17925

ING. EN CIENCIAS DE DATOS



PRESENTA

GABRIELA NICOLÁS
TORRES

PROF. JAIME ALEJANDRO ROMERO SIERRA

25 - NOVIEMBRE - 2024

1. Introducción

- Descripción breve del objetivo del proyecto.

“Encontrar la relación que existe entre la salud mental y el uso de la tecnología, especialmente las redes sociales y como afecta en nuestro entorno laboral.”

- Justificación y contexto: ¿por qué es importante resolver o estudiar esta problemática?

“Es relevante abordar la relación existente entre las redes sociales y la salud mental, ya que, el uso en demasía de estas plataformas puede mermar la autoestima, agudizar la ansiedad y fomentar el desarrollo de la depresión. De esta manera, al comprender mejor esta relación es factible fomentar el uso consciente de las redes sociales, disminuyendo los efectos negativos anteriormente mencionados. De este modo, se pueden “fijar” comportamientos digitales en pro del bienestar mental, los cuales guíen el uso de las redes sociales y permitan incorporarlas en la vida práctica y cotidiana. Esta problemática nos afecta en diversas áreas de nuestra vida, una de ellas es en el entorno de trabajo y es importante analizar esta área para determinar factores que influyen y cómo ayudar a sobrellevar toda la situación.”

- Fuentes de datos: descripción de las bases de datos empleadas (origen, cantidad de datos, principales características).

“La base fue extraída de la página kaggle. El total de datos son 8791. La mayoría de los datos son horas, relacionadas a los niveles de uso de tecnologías, de igual forma edades de los individuos, niveles de estrés, ambiente laboral, horas de sueño, género”

2. Metodología

- **Proceso de limpieza de datos:** Explicación de cómo limpiaron los datos, tipos de datos ausentes y su manejo, eliminación de duplicados, manejo de valores atípicos, etc.
- I. Datos ausentes: Se rellenaron valores faltantes en columnas categóricas con valores "neutrales" (como "Otro" para Género y "Neutral" para Estado de Salud Mental, Nivel de Estrés e Impacto en el Entorno Laboral). Para la columna "Acceso a Sistemas de Soporte", los valores ausentes se eliminaron al no tener un valor adecuado de imputación, ya que los datos eran de tipo binario ("Sí" o "No").
- II. Eliminación de duplicados: Después de limpiar los valores faltantes, se verificó que no quedaran registros duplicados, asegurando que cada entrada del conjunto de datos fuera única.
- III. Valores atípicos: Las estadísticas descriptivas mostraron valores en un rango aceptable, por lo que no se observaron valores atípicos extremos o inusuales. El análisis se centró en mantener los valores dentro de rangos normales.
- IV. Tipos de datos: Finalmente, se ajustaron los tipos de datos (como cambiar "Edad" a tipo entero) para facilitar el análisis y asegurarse de que cada columna tuviera el tipo de dato adecuado.

El conjunto de datos resultante se guardó en un archivo limpio, listo para análisis posteriores.

- **Análisis Exploratorio de Datos (EDA)**

El EDA es una fase clave para entender la estructura, patrones y relaciones dentro de los datos antes de la construcción del modelo. Esta sección debería dividirse en subapartados que aborden distintos aspectos del análisis exploratorio:

1) Descripción General de los Datos

- Visión General: El dataset cuenta con 8791 registros y diversas variables que describen aspectos del uso de tecnología y su relación con la salud mental. Estas incluyen horas de uso de tecnología, edad de los individuos, niveles de estrés, ambiente laboral, horas de sueño, género, entre otras.
- Tipos de Variables:
 - Catóricas: Género, Estado de Salud Mental, Nivel de Estrés, Impacto en el Entorno Laboral, y Acceso a Sistemas de Soporte.
 - Numéricas: Edad, Horas de uso de la tecnología, Horas de uso de las redes sociales, Horas de juego, Horas de tiempo en pantalla, Horas de sueño, Horas de actividad física
- Resumen Estadístico:

Unnamed: 0	Edad	Horas de uso de la tecnología	Horas de uso de las redes sociales	Horas de juego	Horas de tiempo en pantalla	Horas de sueño	Horas de actividad física
count	8791.000000	8791.000000	8791.000000	8791.000000	8791.000000	8791.000000	8791.000000
mean	4395.000000	41.448868	6.438605	3.971663	2.529769	7.972097	6.492429
std	2537.887442	13.805403	3.133297	2.295750	1.420544	3.968671	1.439642
min	0.000000	18.000000	1.000000	0.000000	0.000000	1.000000	0.000000
25%	2197.500000	29.000000	3.780000	2.010000	1.335000	4.640000	5.270000
50%	4395.000000	41.000000	6.444082	3.973487	2.529028	7.972650	6.493213
75%	6592.500000	53.000000	9.080000	5.970000	3.760000	11.300000	7.730000
max	8790.000000	65.000000	12.000000	8.000000	5.000000	15.000000	9.000000

La mediana de edad es: 41.8
La mediana de horas de uso de redes sociales es: 3.973487447461093
La mediana de horas de actividad física es: 5.004031782065835
La mediana de horas de tiempo en pantalla es: 7.97265033666125
La mediana de horas de sueño es: 6.493213189312109
La mediana de horas de juego es: 6.444081563103487
La mediana de horas de uso de la tecnología es: 6.444081563103487

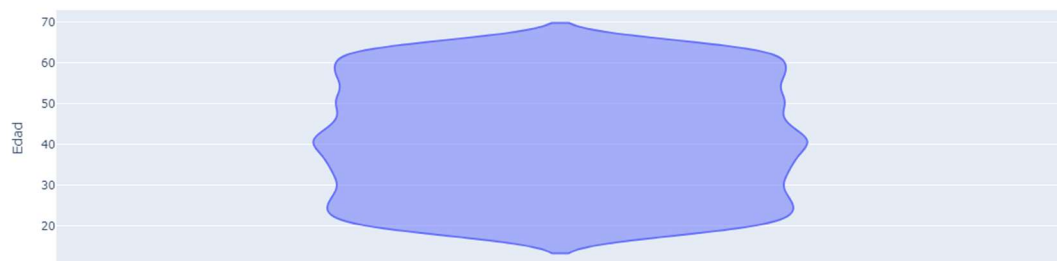
Frecuencia de categorías para las variables categóricas.

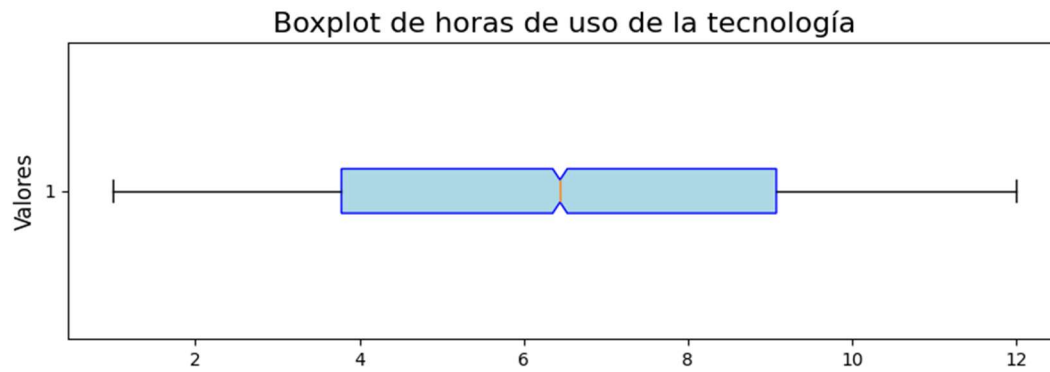
- Género: 33% son hombres, 32.4% son mujeres, y 34.7% son de otro género.
- Estado de Salud Mental: 25% Excelente, 25.8% Neutral, 24.6% Buena, 24.6% Mala.
- Nivel de Estrés: 32.2% Alto, 34% Medio, 32.9% Bajo.
- Impacto en el Entorno Laboral: 33.7% es neutral, 33.1% es positivo, 33.1% es negativo.
- Acceso a Sistemas de Soporte: el 50% sí tiene acceso a sistemas de soporte, y el otro 50% no tiene acceso a sistemas de soporte.

2) Visualización y Distribución de Variables Individuales

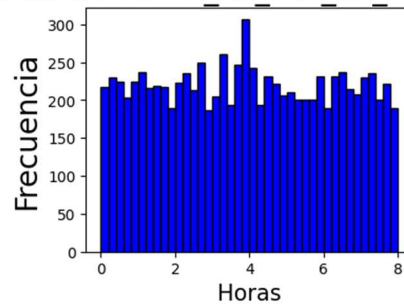
- Variables Numéricas:

Diagrama de Violín-Edad

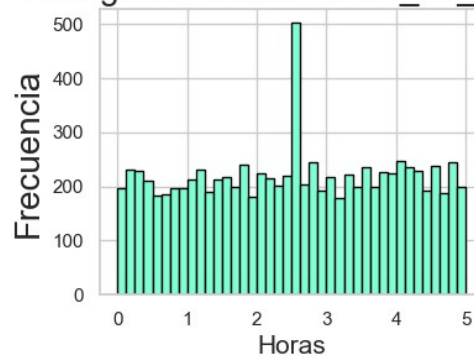




Histograma de las Horas_de_uso_de_las_redes_sociales

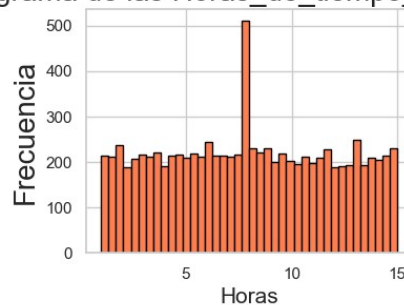


Histograma de las Horas_de_juego



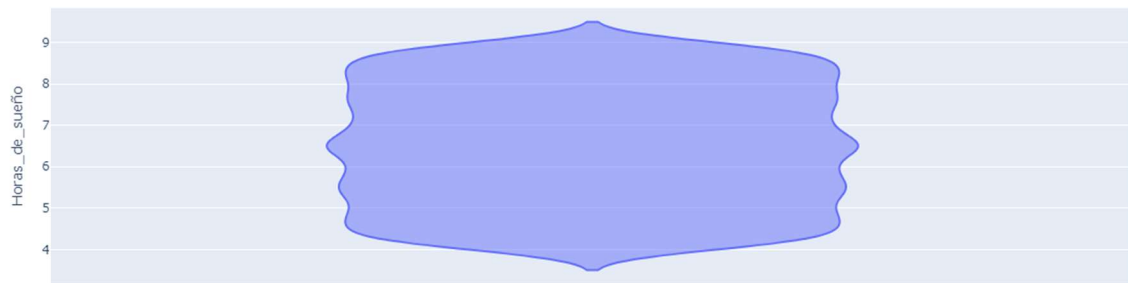
Hay una frecuencia particularmente alta alrededor del valor "3" (según el eje X). Esto sugiere que muchas personas jugaron aproximadamente 3 horas.

Histograma de las Horas_de_tiempo_en_pantalla

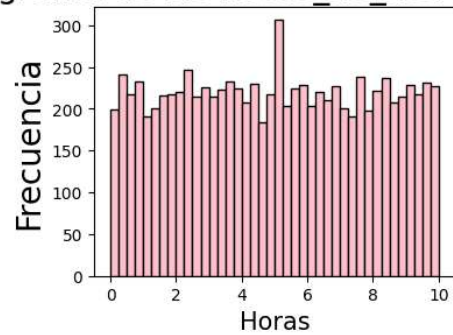


Nuevamente se aprecia un incremento significativo en este histograma, eso debido a que cuando se realizó la limpieza de la base de datos, los espacios vacíos se rellenaron con la media.

Diagrama de Violín-Horas de sueño

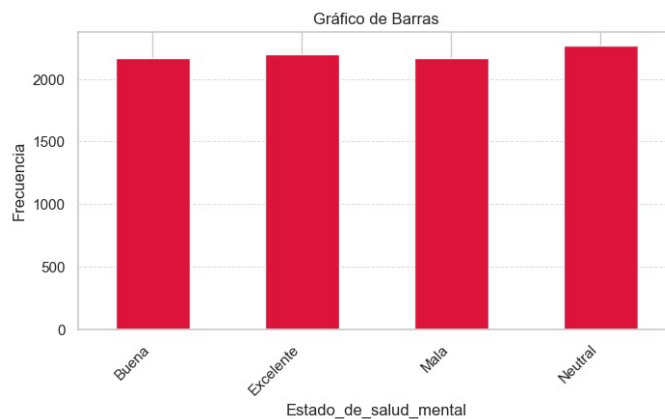
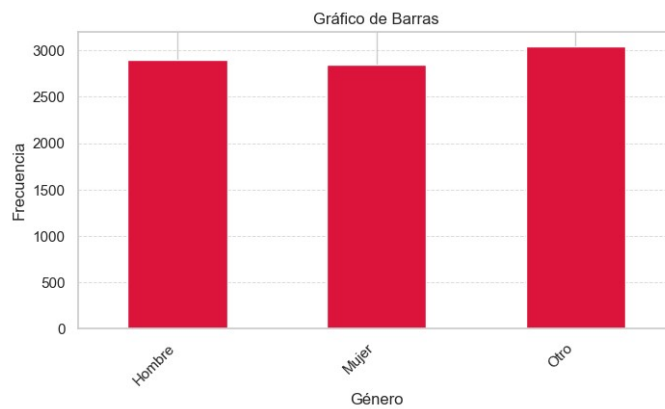


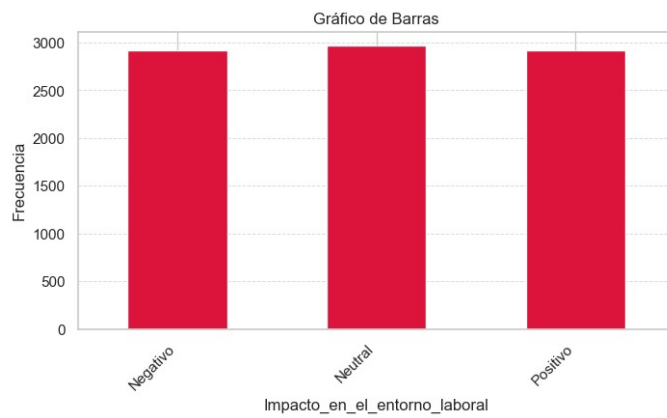
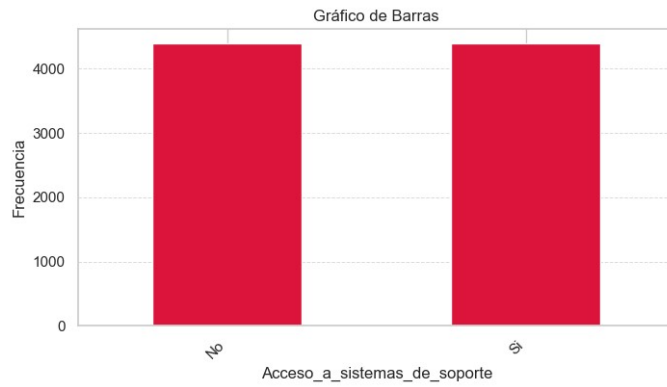
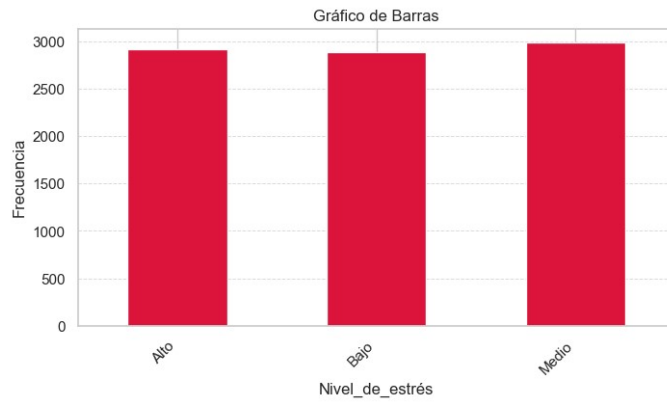
Histograma de las Horas_de_actividad_física



En estos gráficos podemos apreciar que todos nuestros datos tienen una distribución muy equitativa, no existen outliers ni espacios vacíos.

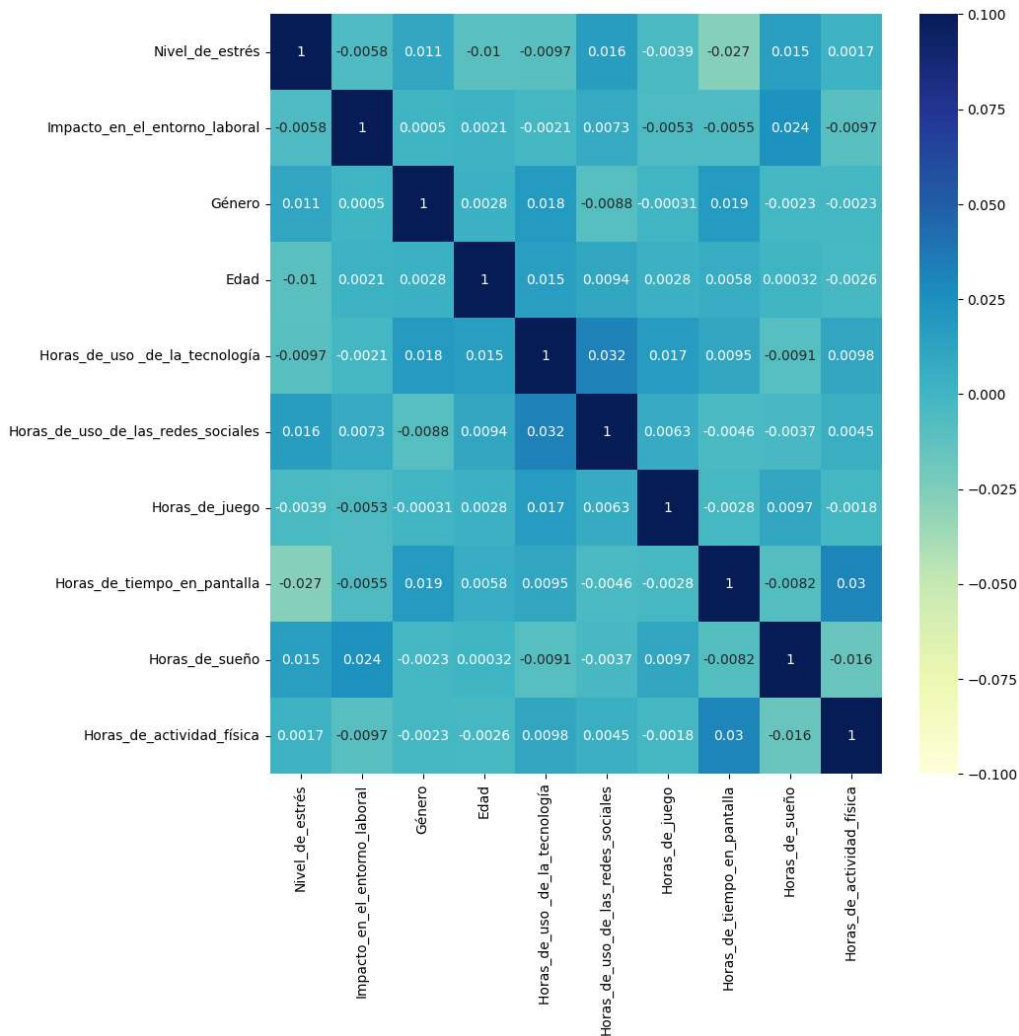
b) Variables Categóricas: Todas se encuentran distribuidas de forma muy equilibrada.



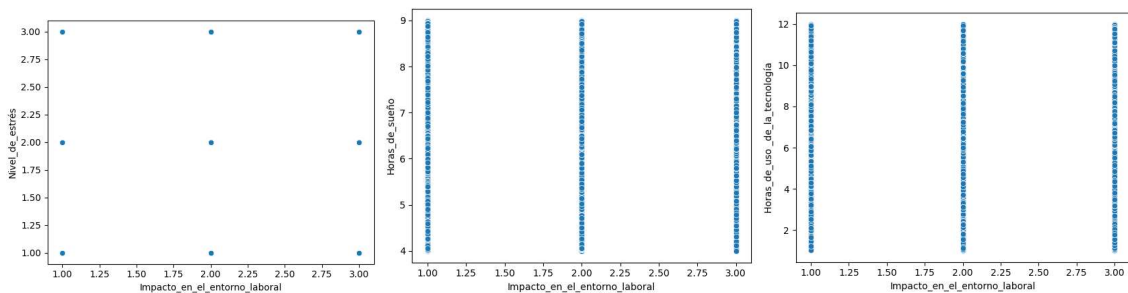


3) Correlación entre Variables

- a) Matriz de Correlación: Podemos apreciar en esta matriz que no existe ninguna correlación significativa entre las variables. Las variables de tipo objeto las cambiamos a numéricas para que pudiéramos considerarlas en este gráfico.



- b) Parejas de Variables: Realmente como antes mencionaba, no existe una correlación fuerte entre las variables. Comparando el impacto en el entorno laboral con variables como nivel de estrés, horas de sueño y horas de uso de la tecnología se pueden apreciar de una forma más clara todo.



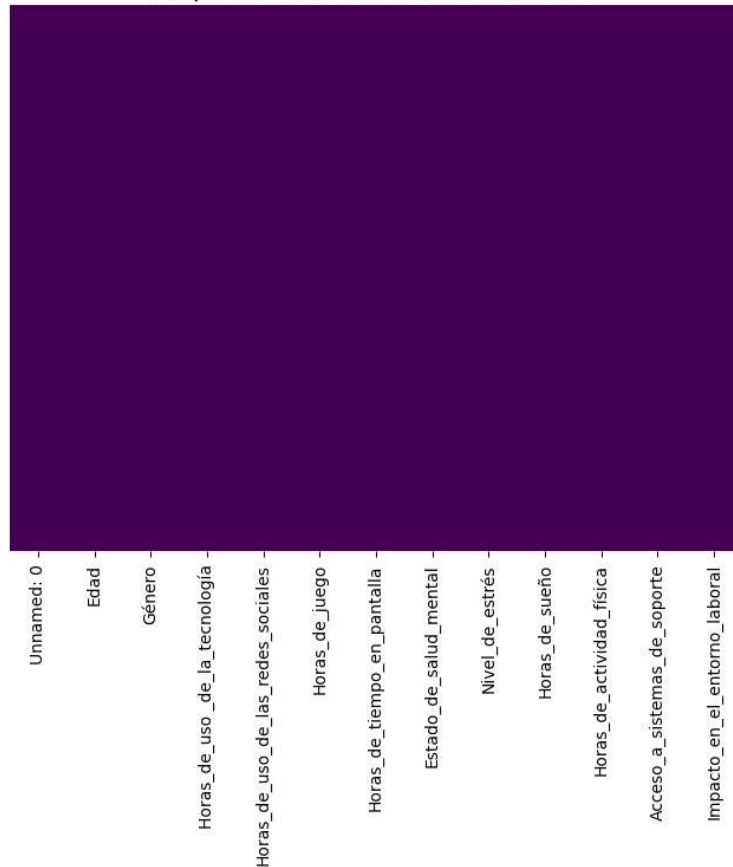
4) Análisis de Valores Atípicos (Outliers)

- a) Identificación de Outliers: Realizando el método del Rango Inter cuartílico (IQR), determinamos que no existen outliers en nuestra base de datos. Los comprobamos haciendo boxplots.

5) Análisis de Valores Faltantes

- a) Identificación de Datos Faltantes:

Mapa de Calor de Datos Faltantes

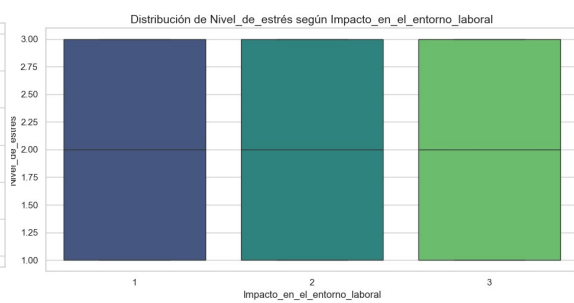
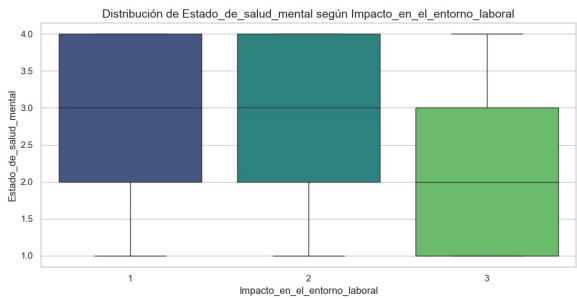
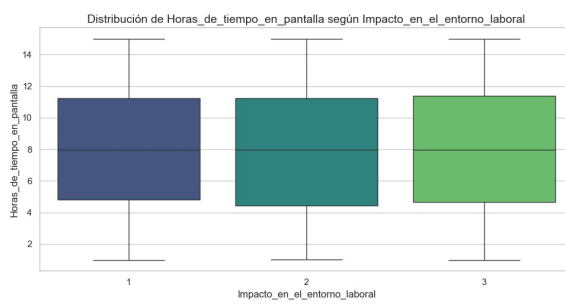
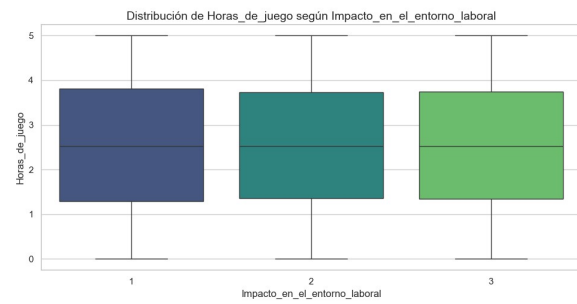
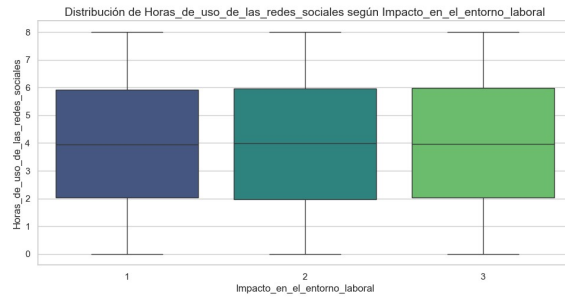
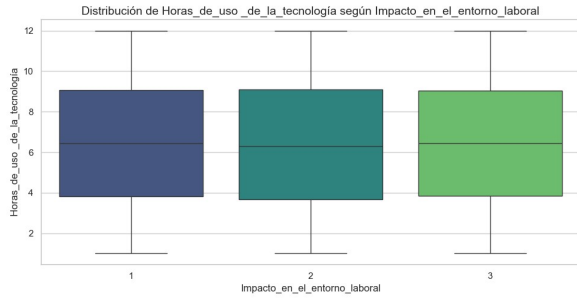
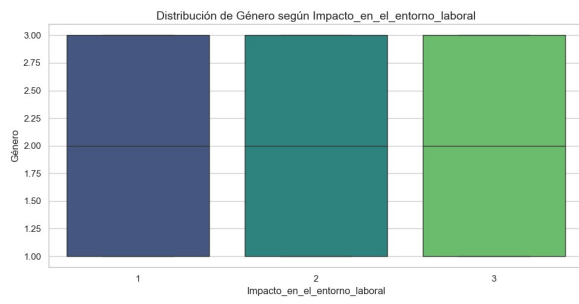
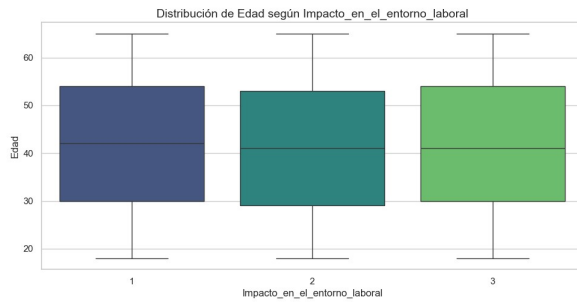


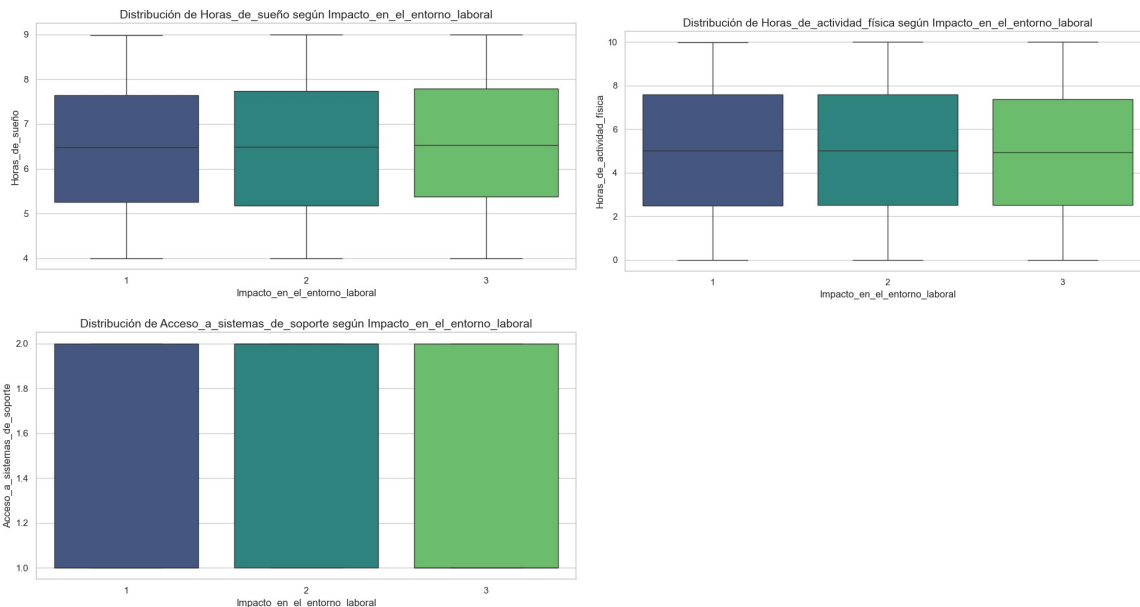
```
Proportión de valores faltantes por columna:  
Unnamed: 0      0.0  
Edad            0.0  
Género          0.0  
Horas_de_uso_de_la_tecnología  0.0  
Horas_de_uso_de_las_redes_sociales  0.0  
Horas_de_juego  0.0  
Horas_de_tiempo_en_pantalla  0.0  
Estado_de_salud_mental      0.0  
Nivel_de_estrés      0.0  
Horas_de_sueño      0.0  
Horas_de_actividad_física  0.0  
Acceso_a_sistemas_de_soporte  0.0  
Impacto_en_el_entorno_laboral  0.0  
dtype: float64
```

No existen valores faltantes, en el proceso de limpieza de datos usamos la estrategia de imputación con el promedio para los valores que nos hacían falta.

6) Relación entre Variables Categóricas y Numéricas

a) Análisis Comparativo:





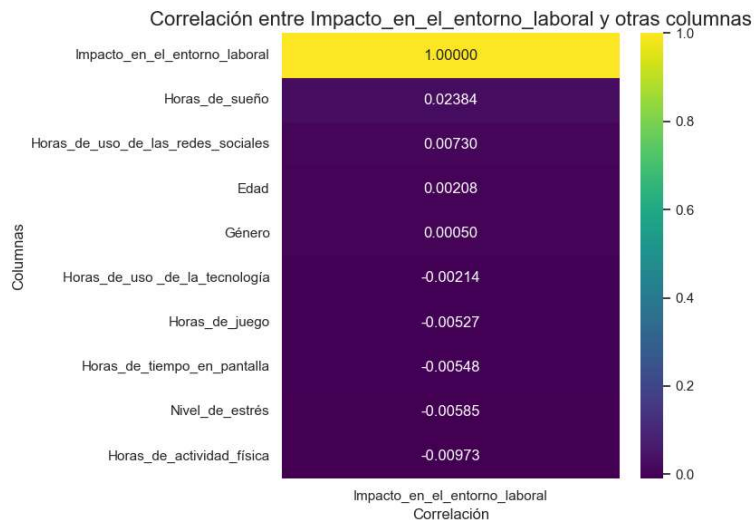
Realizamos boxplots comparando la columna “Impacto en el entorno laboral” con cada columna de la base de datos, estos fueron los resultados explicados con más claridad.

1. Edad vs Impacto en el entorno laboral:
 - No hay grandes diferencias entre los grupos.
 - Parece que la edad tiene una distribución similar en las categorías Positivo, Neutral y Negativo, con una mediana en el rango medio. Esto indica que el impacto en el entorno laboral no está fuertemente relacionado con la edad.
2. Horas de uso de redes sociales vs Impacto en el entorno laboral:
 - Similitud entre los grupos: Las distribuciones de las horas de uso de redes sociales son bastante similares entre los tres grupos (Positivo, Neutral, Negativo), con medianas y rangos intercuartiles casi idénticos.
 - Outliers: Hay algunos valores extremos en los tres grupos, pero no parecen concentrarse en un grupo específico.
 - Conclusión principal: El tiempo que las personas dedican a redes sociales no parece influir directamente en su percepción del impacto laboral. Es posible que otros factores relacionados con la calidad del uso de redes sociales, y no la cantidad de horas, tengan un mayor impacto.
3. Género vs Impacto en el entorno laboral:
 - El grupo "Hombre" (1) parece ser más frecuente en todas las categorías, mientras que "Mujer" (2) y "Otro" (3) tienen una representación menor.
 - No hay diferencias notables en términos de impacto laboral entre los géneros.
4. Horas de uso de redes sociales vs Impacto en el entorno laboral:
 - La mediana y el rango de horas de uso de redes sociales son similares entre los tres grupos.
 - Esto sugiere que el impacto laboral no está significativamente influenciado por las horas en redes sociales.
5. Horas de juego vs Impacto en el entorno laboral:
 - Los grupos Positivo, Neutral y Negativo tienen distribuciones bastante similares.
 - Esto indica que el tiempo dedicado a jugar no parece estar relacionado con el impacto en el entorno laboral.
6. Horas de tiempo en pantalla vs Impacto en el entorno laboral:

- Aunque hay ligeras variaciones, las distribuciones son consistentes entre los grupos.
 - Parece que el tiempo en pantalla no tiene un impacto directo en el entorno laboral.
7. Estado de salud mental vs Impacto en el entorno laboral:
- La mediana en el grupo Positivo se encuentra en "Excelente" (1).
 - En el grupo Neutral, la mediana se inclina hacia "Buena" (2).
 - En el grupo Negativo, la distribución incluye más valores altos, como "Mala" (3).
 - Conclusión: Un impacto positivo en el entorno laboral está asociado con un mejor estado de salud mental, mientras que un impacto negativo está relacionado con una peor salud mental.
8. Nivel de estrés vs Impacto en el entorno laboral:
- El grupo Positivo tiene una mediana en "Bajo" (1).
 - Los grupos Neutral y Negativo tienen medianas en "Medio" (2) y "Alto" (3), respectivamente.
 - Conclusión: Un impacto positivo en el entorno laboral está relacionado con un menor nivel de estrés, mientras que un impacto negativo está vinculado a un mayor nivel de estrés.
9. Horas de sueño vs Impacto en el entorno laboral:
- Las distribuciones son similares entre los grupos, aunque el grupo Negativo parece tener una leve reducción en la mediana de horas de sueño.
 - Conclusión: Aunque la relación no es fuerte, podría haber un ligero vínculo entre menos horas de sueño y un impacto negativo en el entorno laboral.
10. Horas de actividad física vs Impacto en el entorno laboral:
- Los grupos Positivo y Neutral tienen distribuciones más centradas que el grupo Negativo, que parece tener valores más dispersos.
 - Conclusión: La actividad física regular podría estar ligeramente asociada con un impacto positivo en el entorno laboral.
11. Acceso a sistemas de soporte vs Impacto en el entorno laboral:
- En el grupo Positivo, la mayoría tiene "Sí" (1) como acceso a sistemas de soporte.
 - En el grupo Negativo, hay una mayor proporción de "No" (2).
 - Conclusión: El acceso a sistemas de soporte parece estar relacionado con un impacto positivo en el entorno laboral.

7) Observaciones y Hallazgos Importantes

- a) La variable que tenemos como foco de investigación es "Impacto en el entorno laboral", realizamos un heatmap de esta columna en específico con el resto de columnas. Los resultados que nos muestran son los siguientes:
- No existe una correlación fuerte entre la columna de referencia y el resto de las columnas que comparamos, la que más correlación tiene es "Horas de sueño" con 0.02, aunque no es alta, es la que tiene mayor valor.



b) Resumir Hallazgos Clave:

- Impacto del uso de redes sociales: Aunque no se ha encontrado una correlación directa y fuerte entre el tiempo dedicado a las redes sociales y su impacto en el trabajo, el análisis indica que los efectos cualitativos, como el tipo de uso y la calidad de las interacciones, podrían tener más influencia que simplemente la cantidad de horas. Esto es importante, ya que subraya la necesidad de realizar estudios más detallados sobre cómo las redes sociales afectan tanto los aspectos laborales como los personales.
- Relaciones detectadas entre variables: el análisis reveló que un mejor estado de salud mental y niveles más bajos de estrés están relacionados con resultados laborales positivos. Esto subraya la importancia de implementar políticas laborales que prioricen el bienestar psicológico. De igual forma encontramos que la disponibilidad de sistemas de apoyo mostró una conexión evidente con resultados positivos en el ámbito laboral, lo que resalta la necesidad de promover redes de apoyo en los entornos laborales y personales.
- Factores con baja influencia: Variables como la edad, el género, las horas dedicadas a jugar y las horas de pantalla no presentan una relación significativa con el impacto en el entorno laboral, lo que sugiere que no son los principales determinantes en este contexto.
- Horarios críticos: A pesar de que las horas de sueño y la actividad física mostraron relaciones débiles con el impacto en el trabajo, el análisis indica que una disminución en las horas de sueño podría estar levemente relacionada con efectos negativos en el ámbito laboral. Esto resalta la importancia de tener en cuenta la calidad del sueño como un componente en las estrategias de bienestar en el trabajo.

3. Dashboard

- Capturas y Explicación del Dashboard:

Dashboard de Salud Mental e Impacto Laboral

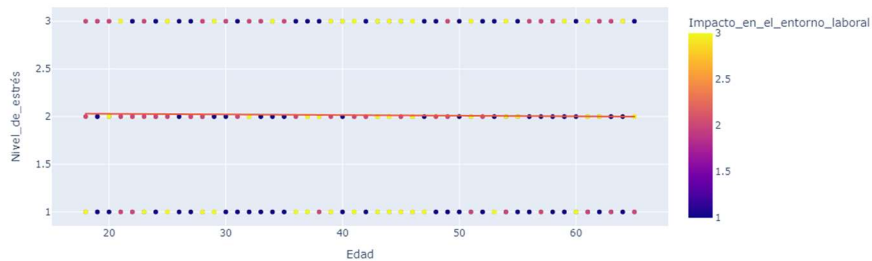
Mujer

Hombre

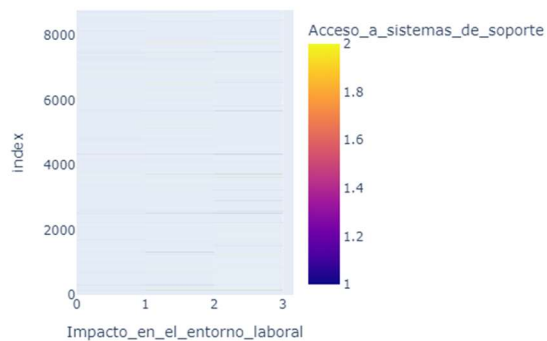
Mujer

Otro

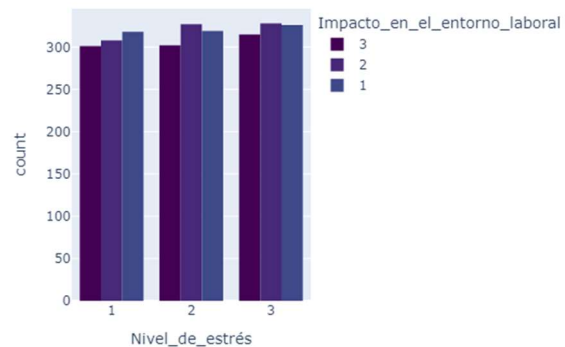
Relación entre Edad y Nivel de Estrés



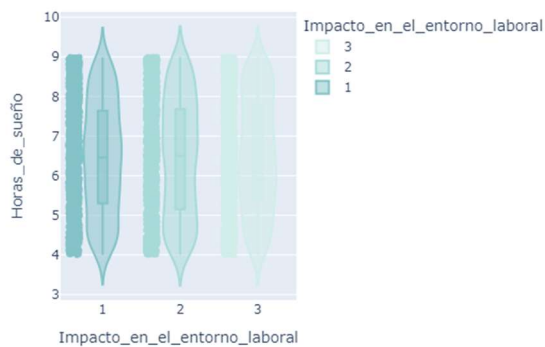
Acceso a Sistemas de Soporte vs Impacto Laboral



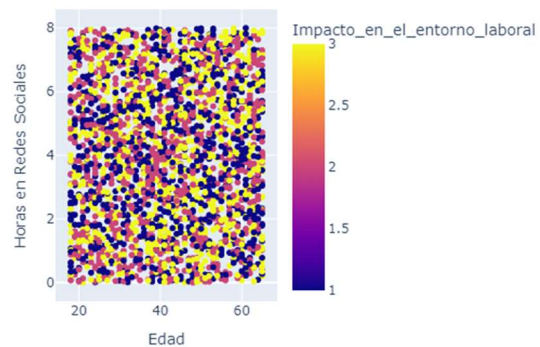
Niveles de Estrés vs Impacto Laboral



Niveles de Sueño vs Impacto en el Entorno Laboral



Relación entre Edad y Uso de Redes Sociales



- [Dash](#)

Esta es una herramienta interactiva para investigar y analizar datos relacionados con la salud mental y su impacto en el entorno laboral en función de la investigación que realiza. La siguiente es una guía general del significado de cada componente según el contenido del dashboard:

i. Interactividad del Dashboard.

- Filtros por género: Permite seleccionar entre géneros como hombre, mujer u otro, para observar cómo varían los datos de salud mental y su impacto laboral según el grupo seleccionado.
- Gráficas dinámicas: Cada vez que se selecciona un género o cualquier filtro, las gráficas cambian automáticamente para reflejar los datos relevantes.

ii. Gráficas y su significado.

El impacto en el entorno laboral está representado por una escala donde 1 es positivo, 2 es neutral y 3 es negativo.

- Relación entre Edad y Uso de Redes Sociales:
 - Muestra cómo la edad influye en el tiempo que las personas dedican a las redes sociales. Puede ser útil para analizar si hay una correlación entre la edad y los hábitos digitales.
 - Las personas con más horas de uso de redes sociales tienden a estar distribuidas en las tres categorías de impacto laboral.
- Horas de Sueño vs Impacto en el Entorno Laboral:
 - Relaciona las horas de sueño con el impacto que estas tienen en el desempeño laboral.
 - Puede resaltar cómo la calidad del sueño afecta el rendimiento laboral, las personas con menos horas de sueño podrían tender a experimentar más impacto negativo.
- Niveles de Estrés vs Impacto Laboral (Gráfico de barras):
 - Compara diferentes niveles de estrés y cómo estos afectan el desempeño en el entorno laboral. Este gráfico es útil para observar si el estrés está relacionado directamente con un mayor impacto negativo.
 - Los tres niveles de estrés muestran distribuciones muy similares en términos de impacto en el trabajo. Esto sugiere que los niveles de estrés pueden no ser el único determinante del impacto que las personas sienten en su entorno laboral. Es decir, el efecto neutral fue consistente en todos los niveles de estrés, lo que puede reflejar el equilibrio general de los efectos.
- Acceso a Sistemas de Soporte vs Impacto Laboral:
 - Este gráfico analiza cómo el acceso a sistemas de soporte (1 = Sí, 2 = No) influye en el impacto en el entorno laboral, acceso a sistemas de soporte como terapia, apoyo familiar o comunitario, tienen un impacto laboral diferente. Puede indicar la importancia de ofrecer estos recursos.
 - Si bien los datos no revelan patrones claros cuando se visualizan, es posible examinar si el acceso a sistemas de apoyo ayuda a reducir los impactos negativos.

- Edad vs Niveles de Estrés:
 - Relaciona las diferentes edades con los niveles de estrés, categorizado en bajo (1), medio (2) y alto (3). Ayuda a identificar grupos etarios que podrían ser más propensos a altos niveles de estrés.
 - Los niveles de estrés están distribuidos de manera uniforme a lo largo de las diferentes edades. Estos son variables, su efecto en el entorno laboral parece mantenerse predominantemente en la categoría "neutral" (2), con algunos casos de impacto negativo (3).

iii. Datos Analizados

Los datos que estás visualizando provienen de tu base de datos y están organizados para responder preguntas clave como:

- ¿Qué factores influyen más en el impacto laboral?
 - ¿Cómo se relacionan los hábitos de sueño, estrés y uso de redes sociales con la salud mental en el trabajo?
- ### iv. ¿Qué puedes aprender del Dashboard?
- Identificar patrones claros y posibles correlaciones entre variables.
 - Detectar áreas específicas donde los empleadores o instituciones podrían intervenir para mejorar el bienestar laboral (ej., promoviendo soporte emocional o prácticas para reducir el estrés).
 - Generar hipótesis para futuros estudios o intervenciones basadas en los datos observados.

4. Conclusiones y Futuras Líneas de Trabajo

- Resumen de los hallazgos principales.
 - Los indicadores más relevantes del impacto laboral en la percepción son la salud mental y el nivel de estrés, lo que enfatiza la necesidad de trabajar en ambientes reducidos en estrés y promover el bienestar psicológico.
 - El acceso a sistemas de soporte es crítico para la percepción positiva, lo que significa que las compañías deben avanzar políticas que apoyen el bienestar social y psicológico.
 - Las horas de uso de redes sociales, aunque no tienen una correlación fuerte con el impacto laboral, podrían estar asociadas a otros factores no explorados, como la calidad de las interacciones digitales.

Estos descubrimientos realizados en la investigación, nos llevan a comprobar una nuestras hipótesis iniciales, los niveles de estrés influyen en su entorno laboral, por ello es importante tomar acción en y procurar espacios donde nuestros niveles de estrés no se vean elevados, o realizar actividades que nos ayuden a regularla; tales como el ejercicio regular, prácticas de mindfulness, y establecer una buena calidad de sueño.

- Posibles mejoras:
 - Investigación cualitativa: analizar cómo el uso de las redes sociales, y más específicamente el uso para determinados propósitos (por ejemplo, ocio, trabajo, conexión social), impacta en la salud mental y el desempeño laboral.

- Estudios longitudinales: desarrollar estudios a largo plazo que permitan tener una continuidad en cuanto al conocimiento y la comparativa de los cambios respecto al impacto laboral y su relación con los hábitos tecnológicos y los hábitos de salud.
- Ampliación del dataset: considerar variables adicionales como los distintos tipos de empleo, las condiciones laborales específicas y la calidad de las interacciones en redes sociales para analizar la información de manera más amena y enriquecer los análisis.
- Políticas fundamentadas en evidencia: Crear intervenciones que incorporen talleres sobre manejo del estrés, hábitos de sueño saludables y un uso consciente de la tecnología.

Estos hallazgos y recomendaciones no solo enriquecen la comprensión del problema presentado, sino que también establecen las bases para intervenciones prácticas y sugerencias que fomenten un equilibrio saludable entre la tecnología y el bienestar mental.

5. Referencias

- Base de datos: mental_health_and_technology_usage_2024. [Kaggle: Your Machine Learning and Data Science Community](#)
- [Optimizing Green Brand Equity: The Integrated Branding and Behavioral Perspectives - Minh-Tri Ha, 2021](#)
- [Información en español sobre la salud mental - National Institute of Mental Health \(NIMH\)](#)
- Python, usando las librerías: numpy, matplotlib, seaborn.

6. Anexos

- Base de datos limpia que se utilizó: raw.githubusercontent.com/Gabby14-nt/PROYECTO_CDD/refs/heads/main/Base_limpia_proyecto.csv
- Video de youtube: <https://youtu.be/89YJuLsvRf8?t=0&si=pAie9bQ6K0khOMxR>