

¿Jugar videojuegos causan trastornos mentales?



Presenta:

José Vega

José Oteiza

José Cabrera

Tutor:

Juan Felipe Gonzales

Profesor:

Lucas Citta

23 de Febrero de 2022

# Índice de contenido:

<u>Introducción:</u>	<u>3</u>
<u>Problema del Negocio:</u>	<u>4</u>
<u>Data acquisition</u>	<u>5</u>
<u>Los acrónimos que se estarán usando en este informe, significan lo siguiente:</u>	<u>5</u>
<u>EDA</u>	<u>6</u>
<u>Se pasaron todos los valores decimales a valor de número entero</u>	<u>7</u>
<u>Procedemos a ver las relaciones entre los feature escogidos</u>	<u>7</u>
<u>Análisis univariado</u>	<u>8</u>
<u>Podemos observar que la cantidad de personas que juegan va disminuyendo mientras avanza su vida laboral y académica</u>	<u>8</u>
<u>Análisis bivariado</u>	<u>8</u>
<u>Podemos ver una relación entre la edad y el narcisismo. Según nuestro dataset a mayor edad mayor es el nivel de narcisismo en la personas</u>	<u>8</u>
<u>Análisis multivariado</u>	<u>9</u>
<u>Podemos ver cómo los jóvenes de entre 20 a 30 años son más propensos a jugar más de 40 horas a la semana, en su mayoría siendo o estudiantes o desempleados.</u>	<u>9</u>
<u>Planteamiento de un modelo de decisión:</u>	<u>9</u>
<u>Clustering</u>	<u>9</u>
<u>PCA</u>	<u>10</u>
<u>Regresión Lineal</u>	<u>11</u>
<u>Conclusión</u>	<u>13</u>

# Introducción:

El tiempo que se pasa en línea y/o jugando videojuegos ha aumentado en la vida diaria de un gran número de personas en todo el mundo. En consecuencia, se han expresado preocupaciones con respecto al comportamiento adictivo tales como, comportamiento de juego en línea excesivo y compulsivo. Los videojuegos se han relacionado con un aumento de los trastornos de salud mental como la ansiedad, la fobia y una disminución con la escala de la vida.

Como grupo nos hemos propuesto averiguar con gráficas que tan graves son cada una de estas patologías.

# Problema del Negocio:

Los videojuegos han aumentado en la vida diaria de un gran número de personas en todo el mundo. Paralelamente ha aumentado los trastornos de salud mental como la ansiedad, la fobia y una disminución en la satisfacción con la vida.

Entre los usuarios de videojuegos se han relacionado que los trastornos de salud aumentan en torno a la cantidad de tiempo dedicado a los videojuegos

Con el fin de realizar una investigación entre el uso de los videojuegos, bienestar y trastornos psicológicos, encontramos una encuesta internacional con más de 13.000 participantes en más de 100 países la cual analizaremos para poder encontrar un patrón que nos defina la correlación entre los videojuegos y los trastornos de salud antes mencionados

## Data acquisition

Fuentes: La fuente de los datos los obtuvimos de:

<https://www.kaggle.com/divyansh22/online-gaming-anxiety-data>

Donde la investigación junto a su profundidad, lo pudimos encontrar en:

[https://osf.io/vnbxk/?view\\_only=4c54da075e164ea2a5329f5669d03c41](https://osf.io/vnbxk/?view_only=4c54da075e164ea2a5329f5669d03c41) (revisar PDF GamingStudy\_Survey, para entender a profundidad el contexto de la investigación realizada)

Los acrónimos que se estarán usando en este informe, significan lo siguiente:

**GAD** = Generalized Anxiety Disorder

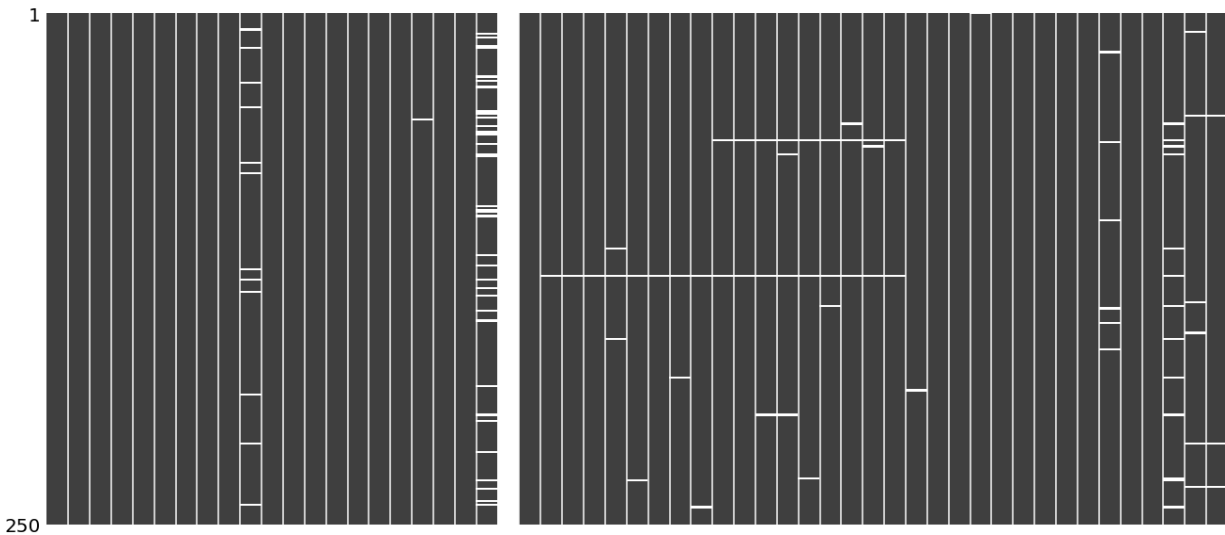
**SWL** = Satisfaction with Life Scale

**SPIN** = Social Phobia Inventory

**SINS** = Single Item Narcissism Scale

# EDA

En esta parte vamos a resumir las principales características de nuestros datos y mirar que nos pueden decir antes de usarlos en una tarea de modelación.



- Empezamos por una vista general de los datos, donde evidenciamos que una columna en nuestro dataset no está aportando nada, por lo cual procedemos a eliminarla. La columna es (**highestleague**)
- Eliminamos 1808 filas dado que tenían al menos un campo en NaN
- Hicimos un estudio de outliers y eliminamos aquellos que a futuro pueden afectar el análisis

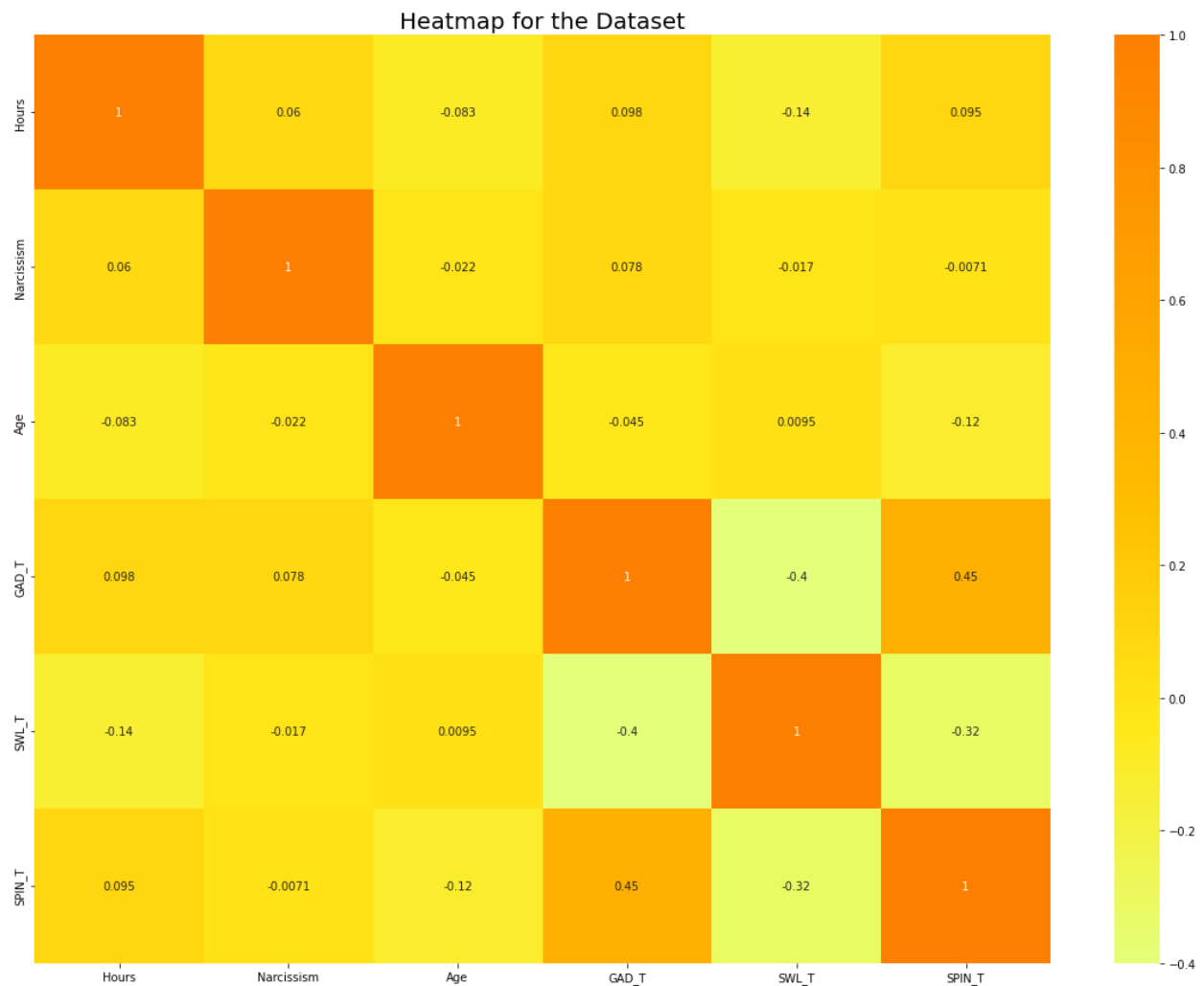
Luego procedemos eliminar las columnas que no nos son de interés para la investigación:

```
df_league_of_legends = df_league_of_legends.drop(columns=[
    'earnings',
    'Reference',
    'accept',
    'Birthplace_IS03',
    'Residence_IS03',
    'Timestamp',
    'Birthplace',
    'highestleague',
    'streams',
    'S. No.',
    'Gender',
    'GAD1',
    'GAD2',
    'GAD3',
    'GAD4',
    'GAD5',
    'GAD6',
    'GAD7',
    'GADE',
    'CUL 1'
```

Se pasaron todos los valores decimales a valor de número entero

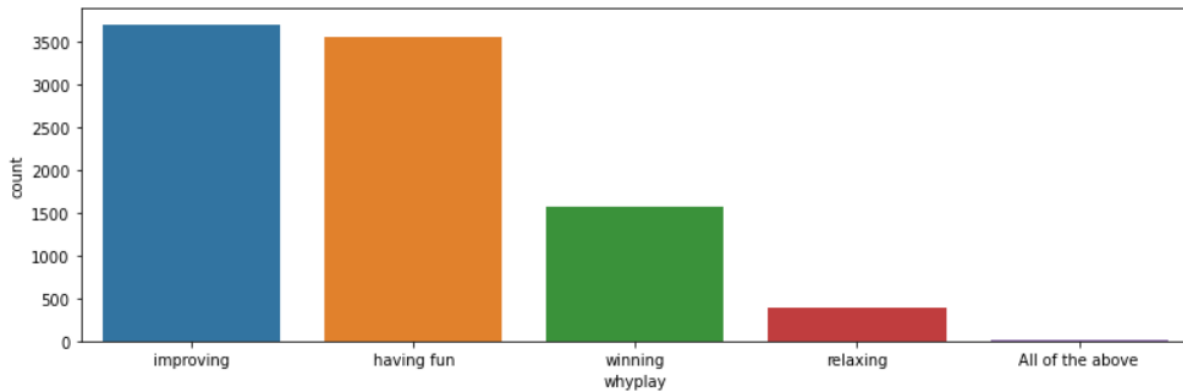
	Game	Platform	Hours	whyplay	League	Narcissism	Age	Work	Degree	Residence	Playstyle	GAD_T	SWL_T	SPIN_T
11	League of Legends	PC	25	having fun	Gold	1	25	Employed	Bachelor (or equivalent)	USA	Multiplayer - online - with real life friends	0	33	3
27	League of Legends	PC	5	having fun	Bronze	3	36	Employed	Bachelor (or equivalent)	Romania	Multiplayer - online - with strangers	5	26	31
33	League of Legends	PC	9	improving	Silver 4	4	23	Employed	High school diploma (or equivalent)	UK	Multiplayer - online - with real life friends	0	26	39
35	League of Legends	PC	7	having fun	Gold	2	25	Student at college / university	Master (or equivalent)	USA	Multiplayer - online - with strangers	1	21	29
48	League of Legends	PC	25	improving	Unranked - Low ELO	3	18	Student at college / university	High school diploma (or equivalent)	USA	Multiplayer - online - with strangers	11	10	17

Procedemos a ver las relaciones entre los feature escogidos



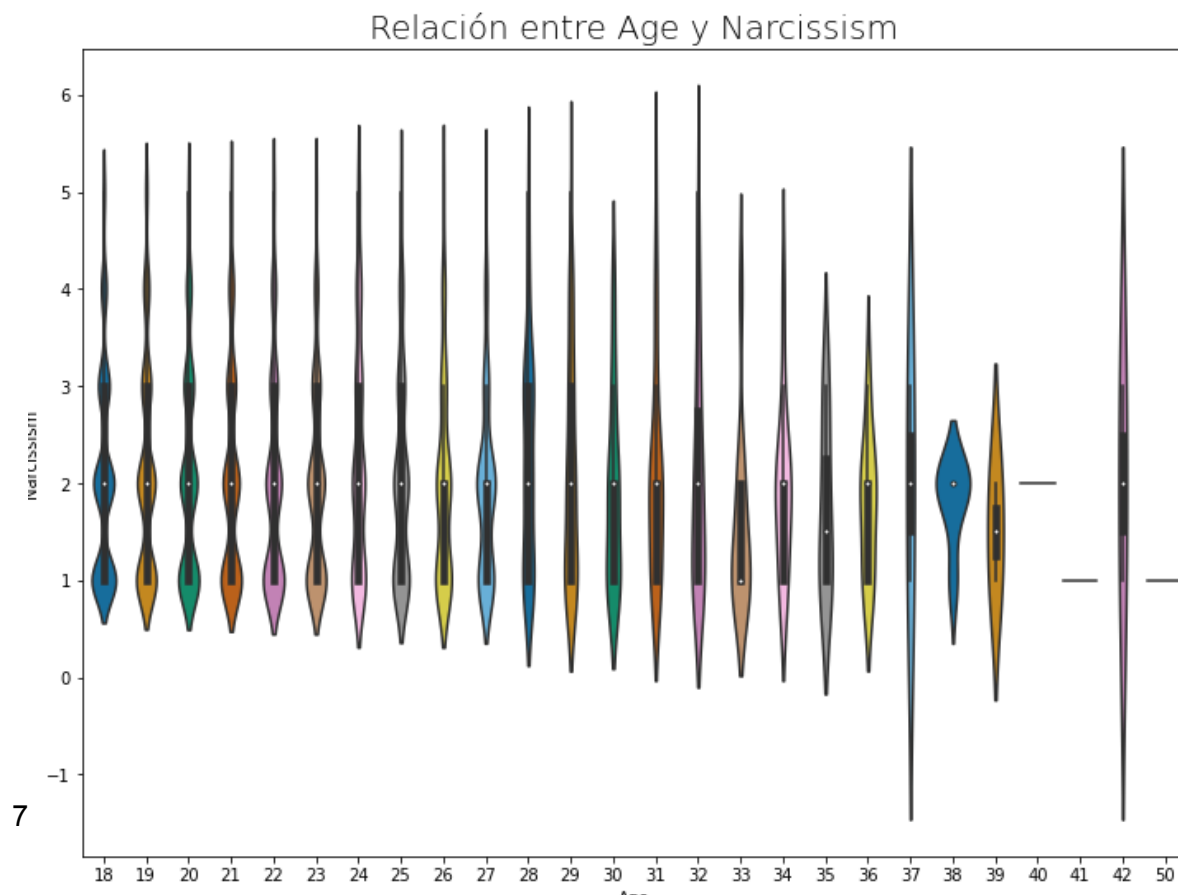
## Analisis univariado

Podemos observar que la cantidad de personas que juegan va disminuyendo mientras avanza su vida laboral y académica



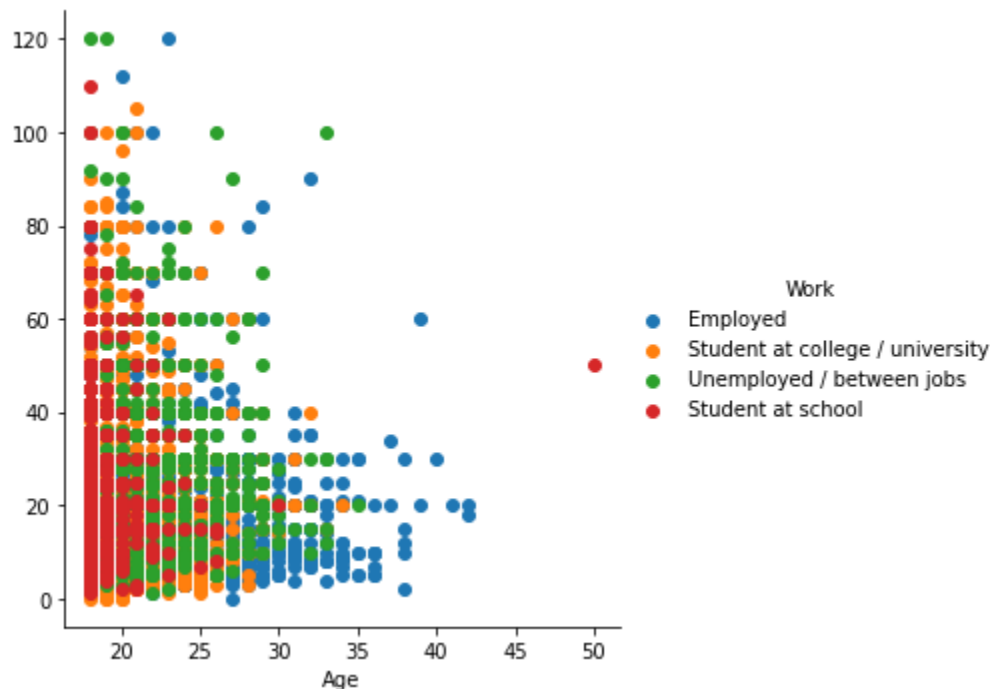
## Analisis bivariado

Podemos ver una relación entre la edad y el narcisismo, Según nuestro dataset a mayor edad mayor es el nivel de narcisismo en la personas



## Análisis multivariado

Podemos ver cómo los jóvenes de entre 20 a 30 años son más propensos a jugar más de 40 horas a la semana, en su mayoría siendo o estudiantes o desempleados.



## Planteamiento de un modelo de decisión:

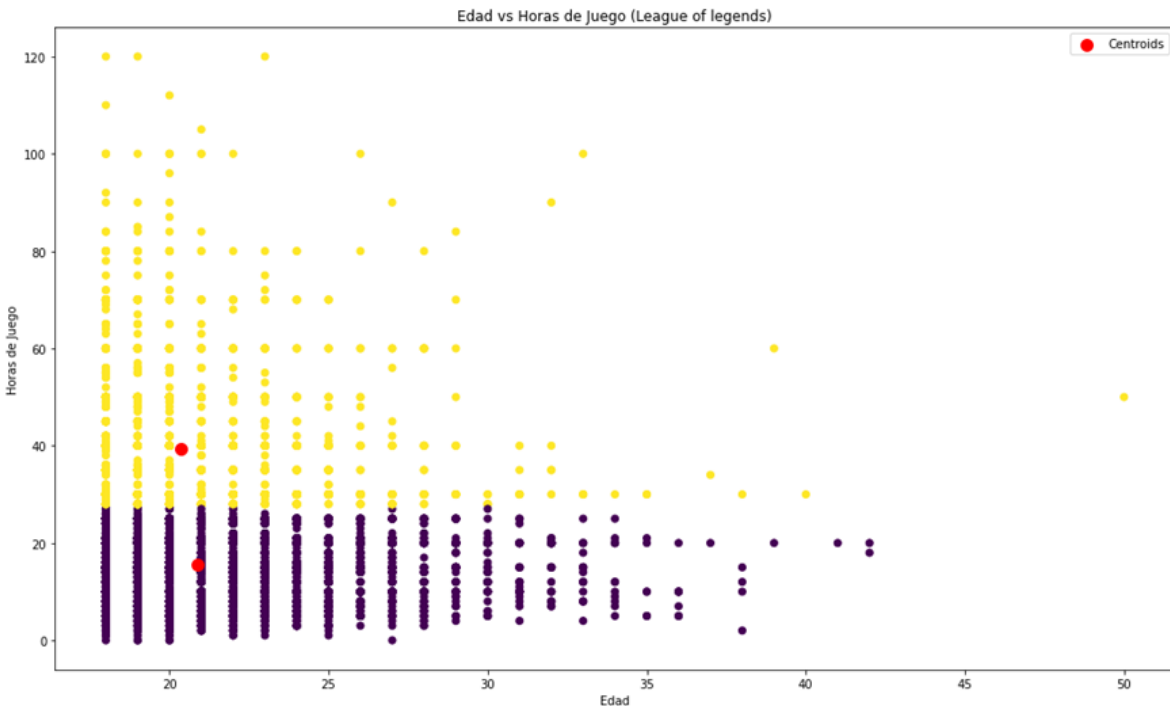
Para el análisis de esta encuesta se plantearon varios modelos de machine learning los cuales vamos a explicar paso a paso.

## Clustering

Para el método de agrupamiento usamos el algoritmo k-means, el objetivo de este es agrupar objetos según sus características.

Para el Primer método agrupamos las horas de juego con la edad del individuo, como se muestra en la siguiente imagen





Como podemos observar, los datos muestran una relación lineal dado que a mayor edad menor son las horas de juego, por consiguiente, no se pueden encontrar grupos claros.

## PCA

Con este método buscamos reducir la dimensionalidad de los datos, para eso realizamos los siguientes pasos.

- Escogimos nuestra variable target (**GAD\_T**) y las variables que van a entrenar el modelo (**SWL\_T, SPIN\_T**)
- Separamos nuestras variables de entrenamiento y test
- Estandarizamos las variables para que tengan un valor entre 0 y 1, ya que de lo contrario las variables con mayor varianza dominarán sobre el resto
- Creamos el modelo y usamos los datos escogidos para entrenarlo
- Evaluamos el modelo con un porcentaje de 14% de asertividad

# Regresión Lineal

Al intentar predecir el GAD\_T, tuvimos que saber cuáles eran los mejores feature para entrar el modelo.

```
importancia_predictores = pd.DataFrame(  
    {'predictor': df_league_of_legends.drop(columns = "GAD_T").columns,  
    'importancia': modelo.feature_importances_  
})  
print("Importancia de los predictores en el modelo")  
print("-----")  
importancia_predictores.sort_values('importancia', ascending=False)
```

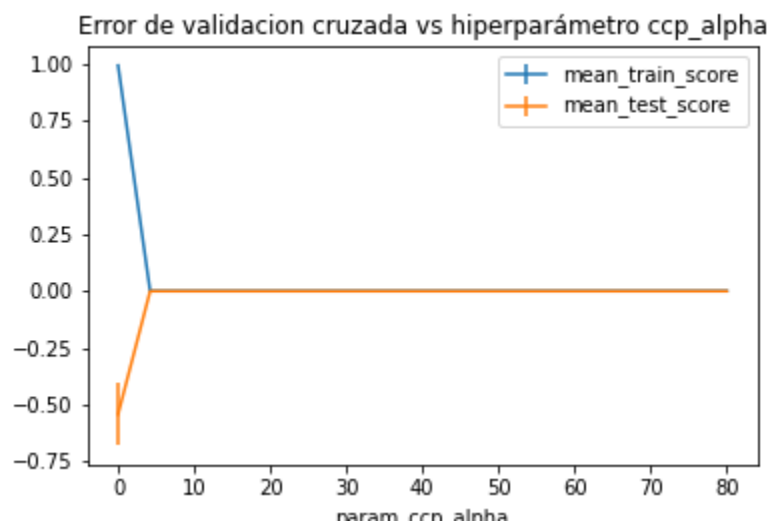
Importancia de los predictores en el modelo  
-----

	predictor	importancia
4	SPIN_T	0.698665
3	SWL_T	0.301335
0	Hours	0.000000
1	Narcissism	0.000000
2	Age	0.000000

Nota: Pudimos encontrar que los feature SPIN\_T y SWL\_T son los mas importante para poder hacer la predicción.

Luego se procedió a calcular el error de validación cruzada vs hiperparámetro ccp\_alpha.

Nota: Con los datos actuales, se dificulta encontrar una diferencia entre ellos. Por eso, el valor



en el gráfico es 0.

Por último, intentamos hacer un árbol de decisiones. Al no tener los datos lo suficientemente funcionales, dicho árbol no se puede desarrollar.

Esto da como resultado que no tenga profundidad.

Profundidad del árbol: 0

Número de nodos terminales: 1

`DecisionTreeRegressor(ccp_alpha=4.2105263157894735, random_state=123)`

samples = 8055  
value = 5.17

**Importante:** Podemos afirmar que este modelo con los datos recolectados no es funcional para responder nuestra duda.

# Conclusión

Habiendo terminado el proceso de escoger datos a analizar (EDA) y limpiar los datos (ETL), nos pudimos dar cuenta que para responder la pregunta “¿Jugar Videojuegos puede causar trastornos mentales?” debemos escoger distintos feature

Debido a métricas obtenidas por los distintos modelos predictivos no tenemos el accuracy (%) adecuado para poder predecir de manera eficaz las preguntas realizadas por el equipo con las variables target escogidas, en una futura versión tendremos mejoras en el modelo escogido.

Como resultado final, podemos encontrar insights interesantes que pueden ayudar a nuestro modelo para obtener mejores métricas que las actuales. Por esto, proponemos dedicar más tiempo de investigación con la finalidad de tener un modelo predictivo que se pueda colocar en producción y con ello, conseguir una posible capitalización de dicho modelo.

Posibles casos de uso para el modelo:

- Evitar crear videojuegos que causan adicciones en los jóvenes, con esto no permitir posibles demandas de entidades de salud mental o estatales.
- Promulgar la necesidad de tomar la soledad o el aislamiento con mayor importancia, para evitar posibles trastornos como la depresión, el estrés, etc.
- Ayudar a los jugadores profesionales (E-Sports) a que obtengan su máximo potencial sin la posibilidad de su agotamiento mental y emocional.
- Avisar a los padres la cantidad de horas máximas que deberían jugar los jóvenes para evitar los trastornos antes mencionados.

Gracias por su atención, sabemos que este es un tema un tanto complejo de tratar... a su vez, vemos importante visibilizar los problemas que puede acarrear el jugar videojuegos. En una siguiente interacción, mostraremos los beneficios que dan los videojuegos.