



# Tecnológico de Monterrey

**Instituto Tecnológico y de Estudios Superiores de Monterrey Campus Puebla**

## **Evidencia 2. Evaluación integradora**

### **Equipo 5:**

José Manuel Morales Escalante

Josué Ibáñez Santana

Sergio Giles Bracamontes

Ismael Martinez Duran

### **Materia:**

Analítica de datos y herramientas de inteligencia artificial I

### **Fecha:**

20 de marzo de 2025

## Reporte Informativo: Análisis de Datos de Interacciones de Wuupi

### Introducción

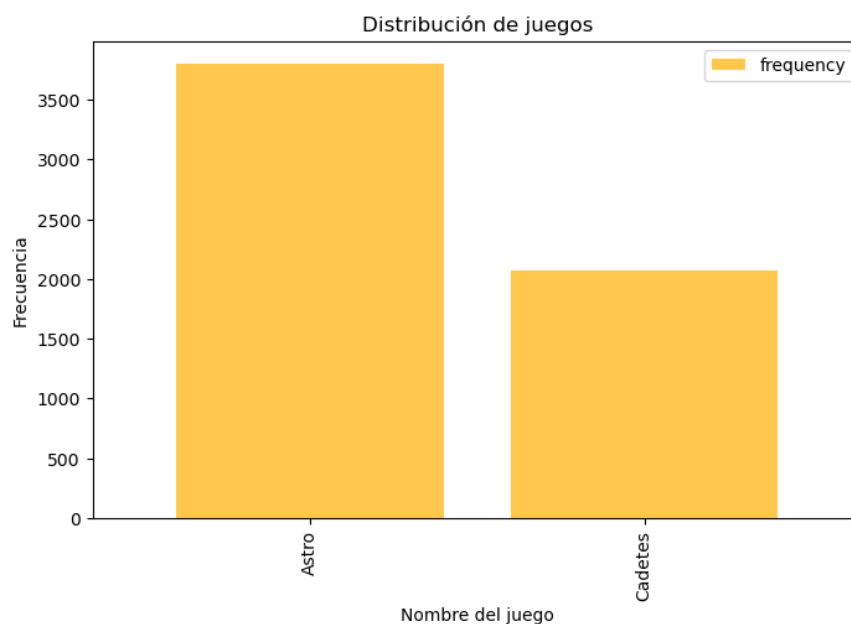
El presente documento aborda un análisis en el que exploramos y describimos los datos recopilados, centrándonos en la distribución de las interacciones por tipo de juego y nivel de dificultad, así como en la identificación de valores nulos y su tratamiento. Para facilitar la comprensión de los datos, se emplearon gráficas que permiten visualizar de manera clara y efectiva la distribución de las interacciones según diferentes variables. Este enfoque gráfico complementa el análisis cuantitativo, ofreciendo una perspectiva más intuitiva de los patrones identificados.

### Análisis de Datos

El conjunto de datos contiene información sobre diversas variables, como el tipo de juego, el nivel de dificultad, el tiempo de interacción y otros aspectos relacionados con la experiencia del usuario. Se identificaron valores nulos en algunas columnas, como "botón correcto", "tiempo de interacción", "número de interacción", "auto push", "tiempo de lección" y "tiempo de sesión". Estos valores nulos fueron sustituidos por "-1" en las columnas cuantitativas y por "Sin registro" en las cualitativas, lo que permitió mantener la integridad del conjunto de datos sin alterar su naturaleza significativa.

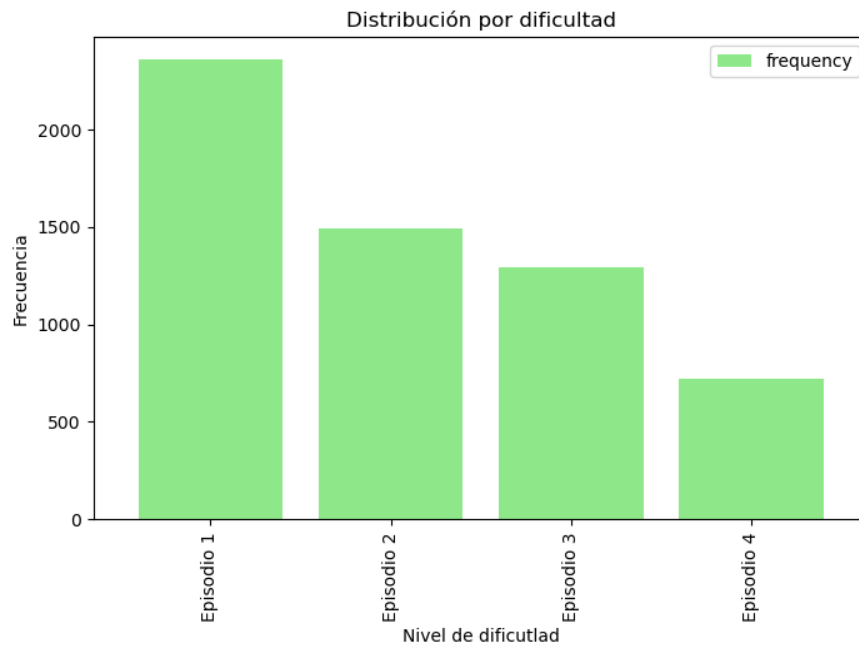
### Reporte de los hallazgos

#### Análisis Univariado de Variables Categóricas



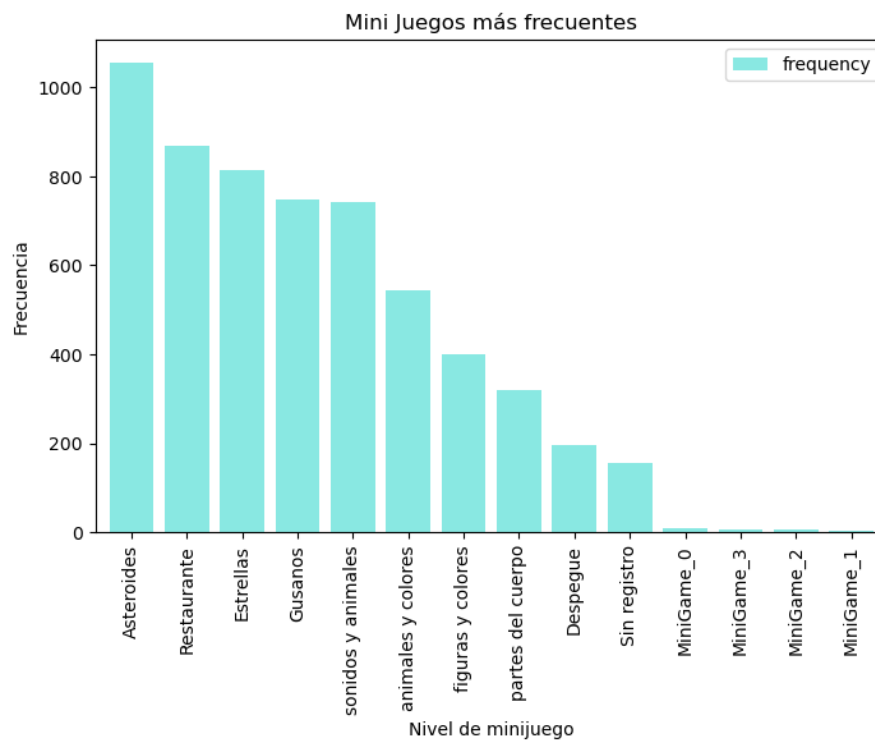
Se puede notar en la gráfica que la mayoría de las interacciones registradas pertenecen al juego "Astro". Esto puede representar cierta preferencia por parte de los terapeutas debido a distintos motivos; tal vez el juego resulta más desafiante, tiene una mejor aceptación por los pacientes, resulta más entretenido, entre otras posibilidades.

## Análisis sobre la Distribución por dificultad



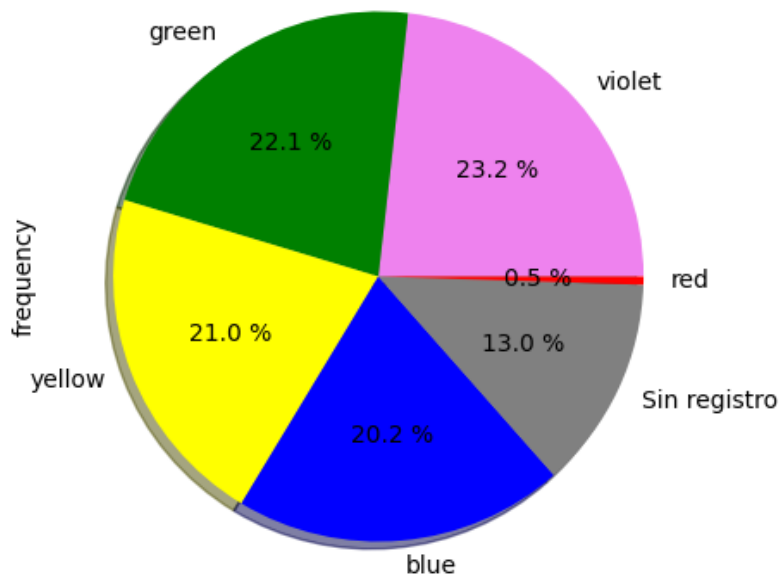
El Episodio 1 es el más frecuente, seguido del Episodio 2. Los episodios más avanzados (3 y 4) tienen menos registros. Lo que indica que los usuarios pasan más tiempo en los niveles iniciales. A medida que va aumentando la dificultad, se reduce considerablemente la cantidad de jugadores en ese nivel de dificultad (episodio 4).

## Análisis sobre los Minijuegos más frecuentes



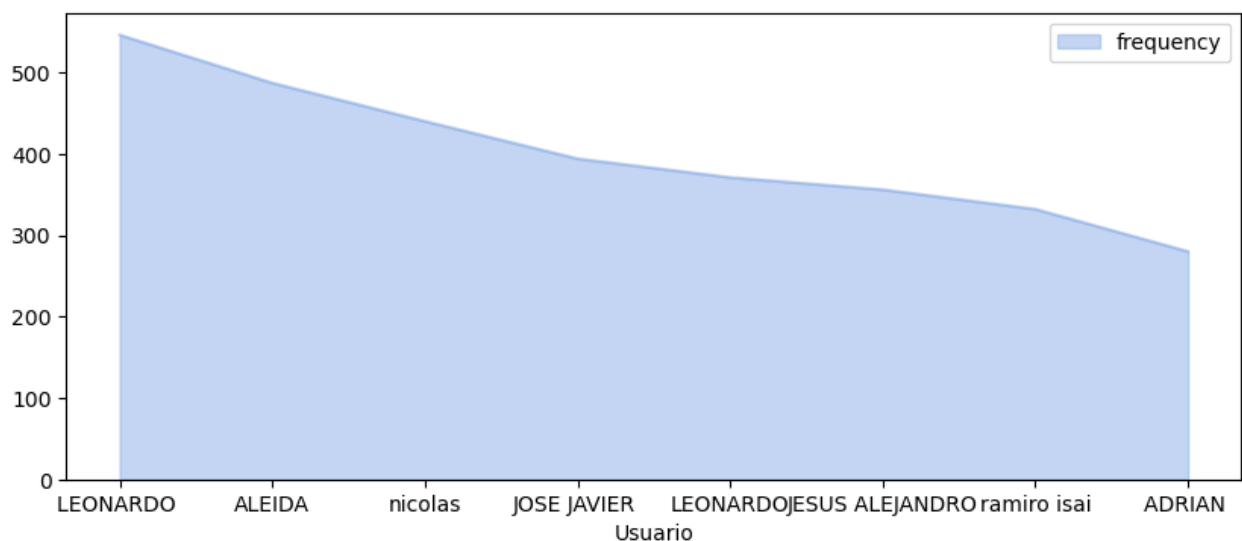
Los mini juegos "Asteroides" y "Estrellas" son los más frecuentes en el dataset, aunque podemos ver que existen ciertos registros con un nombre particular, como lo son *MiniGame\_0*, *1*, *2* y *3*, por lo que es necesario aclarar con el socio a que se refieren estos registros.

### Frecuencia de colores presionados



En base a la información obtenida por el socio, sabemos que los 4 colores principales (yellow, blue, green, violet) se generan de forma aleatoria durante el juego y no existe cierta preferencia sobre algún color. También sabemos que el color **rojo** es poco frecuente ya que su instrucción en el juego es que no se debe presionar, por lo que ese 0.5% representa que el usuario cometió un error.

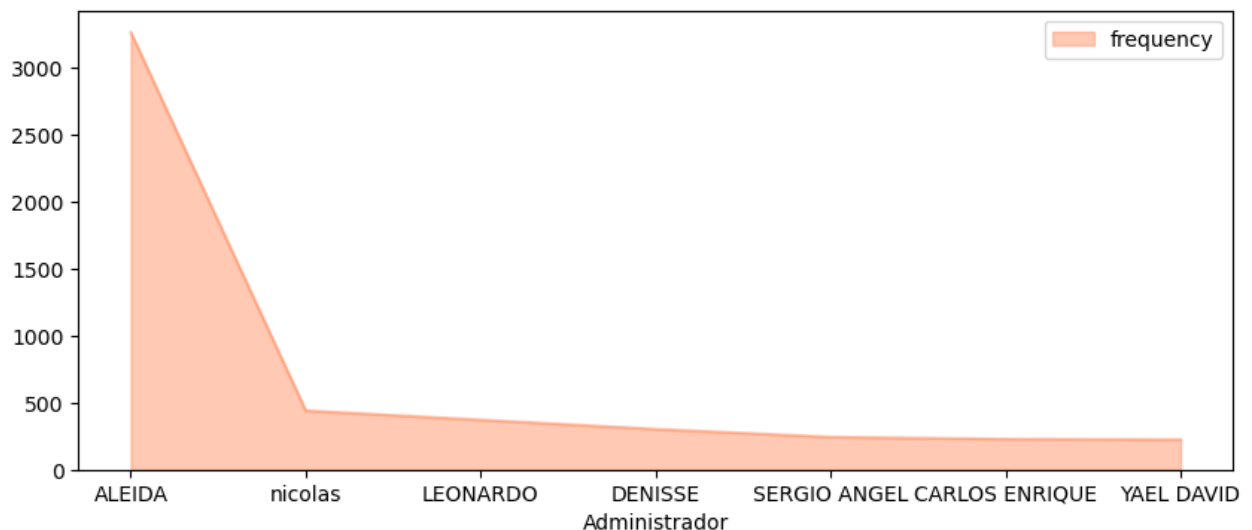
### Análisis sobre los usuarios con más registros en la base de datos



Esta gráfica es útil para identificar cuáles usuarios tienen más registros en la base de datos. Esto se puede traducir como que son los usuarios que han recibido esta terapia durante un

mayor tiempo o durante sesiones más largas, por lo cual sería útil para evaluar su desarrollo gracias a la terapia.

### Análisis sobre los administradores con más registros en la base de datos



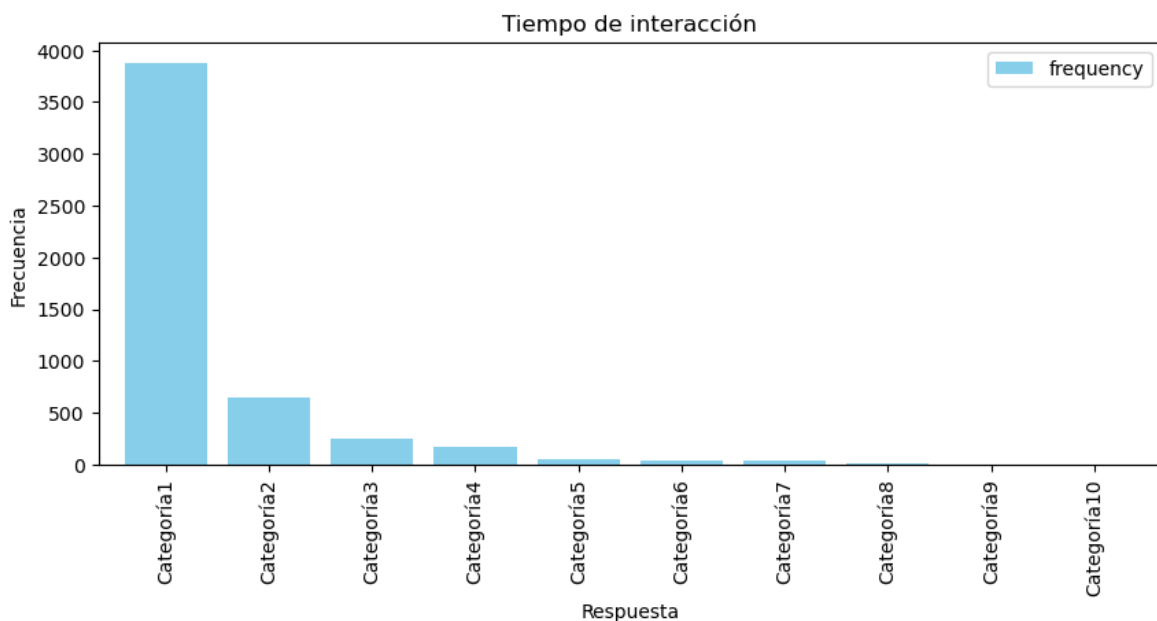
Esta gráfica es similar pero con la diferencia que los registros hacen referencia al administrador que supervisa y lleva a cabo las sesiones con los pacientes. Podemos observar que existe una gran diferencia entre ALEIDA y todos los demás administradores; esto puede ser debido a que se utiliza el administrador de ALEIDA para generar sesiones de prueba, por ejemplo, pero es importante consultar esta información con el socio.

### Categorización de Variables Numéricas (Regla de Sturges)

#### Análisis sobre el tiempo de interacción

CATEGORÍA 1	0 - 11.79 s
CATEGORÍA 2	11.80 - 24.57 s
CATEGORÍA 3	24.58 - 37.36 s
CATEGORÍA 4	37.37 - 50.14 s
CATEGORÍA 5	50.15 - 62.92 s
CATEGORÍA 6	62.93 - 75.71 s

CATEGORÍA 7	75.72 - 88.49 s
CATEGORIA 8	88.50 - 101.27 s
CATEGORIA 9	101.28 - 114.06 s
CATEGORIA 10	114.07 - 126.84 s
CATEGORIA 11	126.85 - 139.62 s
CATEGORIA 12	139.63 - 152.41 s
CATEGORIA 13	152.42 - 165.19s

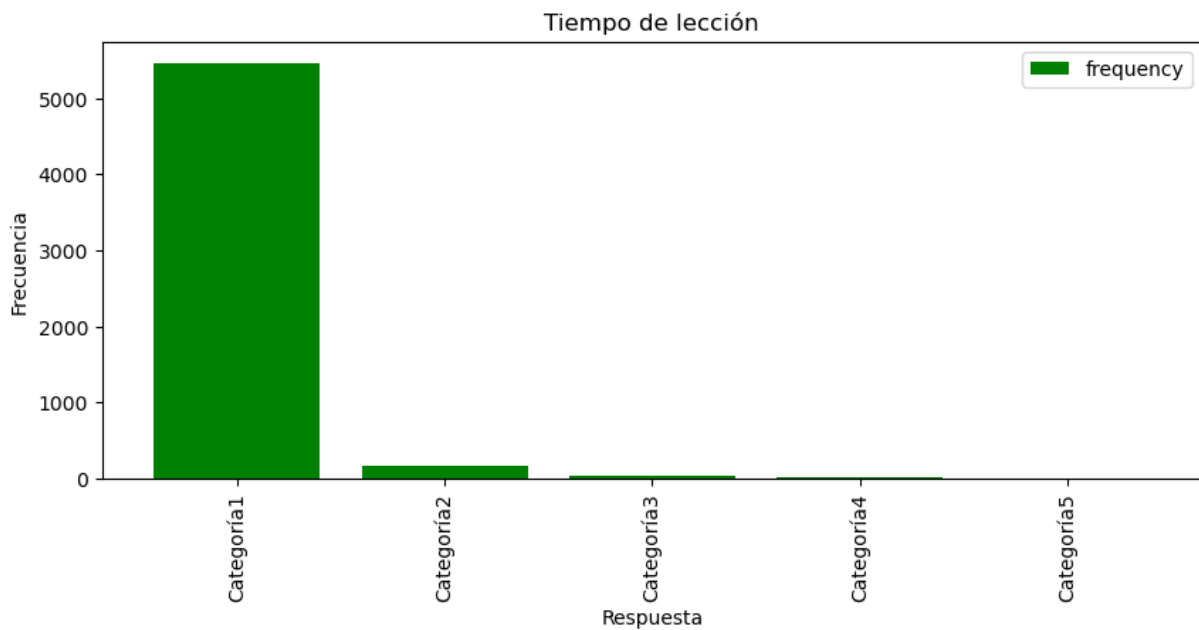


Podemos observar que la mayoría de los datos se encuentran en la categoría 1, es decir, la mayor parte de los tiempos de interacción registrados se encuentran entre 0 y 11.79 segundos, aunque podemos ver que la distribución llega hasta la categoría 4, cuyo límite es 50.14 segundos.

El resto de categorías contiene muy pocos valores y por eso no se logran representar exitosamente en el gráfico.

### **Análisis sobre el tiempo de lección**

CATEGORÍA 1	0 - 59.18 s
CATEGORÍA 2	59.19 - 119.36 s
CATEGORÍA 3	119.37 - 179.54 s
CATEGORÍA 4	179.55 - 239.72 s
CATEGORÍA 5	239.73 - 299.9 s
CATEGORÍA 6	299.91 - 360.08 s
CATEGORÍA 7	360.09 - 420.25 s
CATEGORIA 8	420.26 - 480.43 s
CATEGORIA 9	480.44 - 540.61 s
CATEGORIA 10	540.62 - 600.79 s
CATEGORIA 11	600.80 - 660.97 s
CATEGORIA 12	660.98 - 721.15 s
CATEGORIA 13	721.16 - 781.33 s



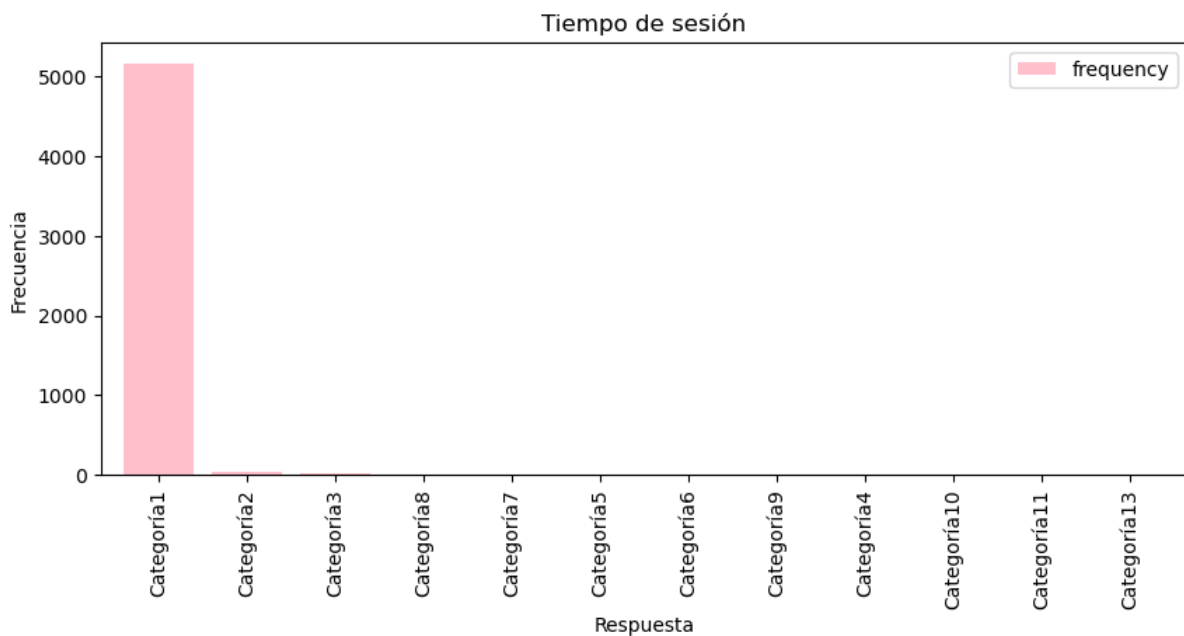
En esta variable, la cual refleja el tiempo destinado a una sola lección, podemos ver un comportamiento similar a la anterior gráfica, donde la mayoría de datos se almacenan en la categoría 1 (de 0 a 59.18 segundos) y unos cuantos en las categorías 2 y 3. El resto de categorías alberga una limitada cantidad de registros y por eso no es posible visualizarlo en la gráfica, aún considerando que el registro más alto es de 781.33 segundos.

### **Análisis sobre el tiempo de sesión**

CATEGORÍA 1	0 - 164.04 s
CATEGORÍA 2	164.05 - 329.07 s
CATEGORÍA 3	329.08 - 494.11 s
CATEGORÍA 4	494.12 - 659.14 s
CATEGORÍA 5	659.15 - 824.18 s
CATEGORÍA 6	824.19 - 989.21 s
CATEGORÍA 7	989.22 - 1154.24 s



CATEGORIA 8	1154.25 - 1319.28 s
CATEGORIA 9	1319.29 - 1484.31 s
CATEGORIA 10	1484.31 - 1649.35 s
CATEGORIA 11	1649.35 - 1814.38 s
CATEGORIA 12	1814.39 - 1979.41 s
CATEGORIA 13	1979.41 - 2144.45 s



Por último, tenemos el tiempo que refleja la duración total de una sesión realizada por cada paciente. En esta gráfica la distribución es todavía más desbalanceada y la categoría 1, que va de 0 a 164.04 segundos de duración, almacena la mayor cantidad de datos, dejando solo unos cuantos registros en las categorías 2 y 3 respectivamente. Esto refleja que la gran mayoría de los pacientes tienen un comportamiento similar y se adaptan de buena manera a las sesiones en las que se emplea esta terapia.

## **Conclusión**

El presente análisis nos permitió examinar de manera sistemática el conjunto de datos sobre interacciones en juegos educativos. Mediante técnicas de limpieza de datos, se abordó el tratamiento de valores nulos, sustituyéndolos adecuadamente según su naturaleza cuantitativa o cualitativa. El análisis univariado, apoyado en representaciones gráficas, reveló distribuciones significativas en las variables clave: tipo de juego, nivel de dificultad y mini juegos.

Los resultados cuantitativos mostraron una marcada concentración de interacciones en el juego "Astro" frente a "Cadetes", así como una disminución progresiva de participación conforme aumenta la dificultad, siendo el Episodio 1 el más frecuente. Los mini juegos "Asteroides", "Restaurante" y "Estrellas" emergieron como los más utilizados.

Este trabajo demostró la importancia de un procesamiento riguroso de datos y la aplicación de análisis estadísticos para extraer información importante sobre patrones de interacción. Los métodos empleados y los resultados obtenidos constituyen una base para futuros análisis más profundos que puedan correlacionar estas métricas con indicadores de aprendizaje.