

# Reporte de hallazgos

**Carlos Fonseca Olivares- A01734538**

**Emily Bueno Romero - A01736939**

**Juan José Zamorano Balderas - A01642396**

**Juan Sebastián Mejía De Gortari - A01722536**

01

# Descripción general

## Lo que podemos encontrar en la base de datos es lo siguiente:

- Total de registros: 5,865
- Total de columnas: 13
- Variables categóricas: Administrador, Usuario, mini juego, color presionado, dificultad, fecha, Juego
- Variables numéricas (cuantitativas): botón correcto, tiempo de interacción, número de interacción, auto push, tiempo de lección, tiempo de sesión

DataAnalytics												
Administrador	Usuario	botón correcto	tiempo de interacción	mini juego	número de interacción	color presionado	dificultad	fecha	Juego	auto push	tiempo de lección	tiempo de sesión
nicolas	nicolas	1	5.399169	Despegue	1	blue	Episodio 1	25/01/2024 09:26:42 a. m.	Astro	0	0	0
nicolas	nicolas	0	1.2834	Despegue	2	violet	Episodio 1	25/01/2024 09:26:46 a. m.	Astro	0	0	0
nicolas	nicolas	1	2.700226	Despegue	3	green	Episodio 1	25/01/2024 09:26:48 a. m.	Astro	0	0	0
nicolas	nicolas	0	3.050262	Despegue	4	green	Episodio 1	25/01/2024 09:26:57 a. m.	Astro	0	0	0
nicolas	nicolas	0	4.750256	Despegue	5	green	Episodio 1	25/01/2024 09:26:58 a. m.	Astro	0	0	0
nicolas	nicolas	0	7.400463	Despegue	6	green	Episodio 1	25/01/2024 09:27:01 a. m.	Astro	0	0	0
nicolas	nicolas	0	13.16735	Despegue	7	green	Episodio 1	25/01/2024 09:27:07 a. m.	Astro	0	0	0
nicolas	nicolas	1	22.46778	Despegue	8	yellow	Episodio 1	25/01/2024 09:27:16 a. m.	Astro	0	0	0
nicolas	nicolas	0	14.51718	Despegue	9	green	Episodio 1	25/01/2024 09:27:37 a. m.	Astro	0	0	0
nicolas	nicolas	0	15.61735	Despegue	10	blue	Episodio 1	25/01/2024 09:27:38 a. m.	Astro	0	0	0
nicolas	nicolas	0	16.80135	Despegue	11	green	Episodio 1	25/01/2024 09:27:39 a. m.	Astro	0	0	0
nicolas	nicolas	1	18.16748	Despegue	12	violet	Episodio 1	25/01/2024 09:27:40 a. m.	Astro	0	0	0
nicolas	nicolas			Despegue			Episodio 1	25/01/2024 09:27:40 a. m.	Astro		63.80254	
nicolas	nicolas	0	9.223312	Asteroides	1	blue	Episodio 1	25/01/2024 09:28:28 a. m.	Astro	0	0	0
nicolas	nicolas	1	10.40688	Asteroides	2	blue	Episodio 1	25/01/2024 09:28:29 a. m.	Astro	0	0	0
nicolas	nicolas	0	1.184384	Asteroides	3	violet	Episodio 1	25/01/2024 09:28:31 a. m.	Astro	0	0	0
nicolas	nicolas	1	4.400641	Asteroides	4	violet	Episodio 1	25/01/2024 09:28:34 a. m.	Astro	0	0	0
nicolas	nicolas	0	2.050367	Asteroides	5	blue	Episodio 1	25/01/2024 09:28:37 a. m.	Astro	0	0	0
nicolas	nicolas	1	2.534215	Asteroides	6	blue	Episodio 1	25/01/2024 09:28:37 a. m.	Astro	0	0	0
nicolas	nicolas	1	2.033403	Asteroides	7	yellow	Episodio 1	25/01/2024 09:28:40 a. m.	Astro	0	0	0
nicolas	nicolas			Asteroides			Episodio 1	25/01/2024 09:28:40 a. m.	Astro		21.64027	
nicolas	nicolas	0	21.19685	Estrellas	1	violet	Episodio 1	25/01/2024 09:30:06 a. m.	Astro	0	0	0
nicolas	nicolas	1	27.01376	Estrellas	2	violet	Episodio 1	25/01/2024 09:30:12 a. m.	Astro	0	0	0
nicolas	nicolas	1	2.066363	Estrellas	3	yellow	Episodio 1	25/01/2024 09:30:16 a. m.	Astro	0	0	0
nicolas	nicolas	0	0.1333601	Estrellas	4	violet	Episodio 1	25/01/2024 09:30:19 a. m.	Astro	0	0	0
nicolas	nicolas	1	1.750274	Estrellas	5	violet	Episodio 1	25/01/2024 09:30:20 a. m.	Astro	0	0	0
nicolas	nicolas	1	0.600757	Estrellas	6	yellow	Episodio 1	25/01/2024 09:30:24 a. m.	Astro	0	0	0
nicolas	nicolas			Estrellas			Episodio 1	25/01/2024 09:30:26 a. m.	Astro		40.84766	
nicolas	nicolas	1	5.166974	Gusanos	1	violet	Episodio 1	25/01/2024 09:31:29 a. m.	Astro	0	0	0
nicolas	nicolas	0	5.750237	Gusanos	2	violet	Episodio 1	25/01/2024 09:31:41 a. m.	Astro	0	0	0
nicolas	nicolas	1	13.03398	Gusanos	3	blue	Episodio 1	25/01/2024 09:31:48 a. m.	Astro	0	0	0
nicolas	nicolas	1	2.067	Gusanos	4	green	Episodio 1	25/01/2024 09:31:56 a. m.	Astro	0	0	0
nicolas	nicolas	0	0.9999382	Gusanos	5	yellow	Episodio 1	25/01/2024 09:32:03 a. m.	Astro	0	0	0
nicolas	nicolas	0	7.867492	Gusanos	6	yellow	Episodio 1	25/01/2024 09:32:10 a. m.	Astro	0	0	0
nicolas	nicolas	0	17.91809	Gusanos	7	green	Episodio 1	25/01/2024 09:32:20 a. m.	Astro	0	0	0
nicolas	nicolas	0	31.38541	Gusanos	8	yellow	Episodio 1	25/01/2024 09:32:34 a. m.	Astro	0	0	0
nicolas	nicolas	0	31.60169	Gusanos	9	green	Episodio 1	25/01/2024 09:32:34 a. m.	Astro	0	0	0

# Limpieza de datos

## Base limpia, resultados limpios

En el análisis realizado, se implementaron técnicas de limpieza de datos para garantizar que estén completos y que sean consistentes. En el caso de los datos numéricos, los valores faltantes fueron reemplazados por el promedio de cada columna, lo cual es una práctica común cuando se busca mantener una tendencia central sin distorsionar los datos. Por otro lado, los datos categóricos se completaron utilizando métodos de propagación hacia adelante y hacia atrás, lo que es útil cuando se espera que la categoría de un registro sea similar a la anterior o la siguiente, como puede ocurrir con sesiones de un mismo usuario. Después de este proceso, el conjunto de datos quedó libre de valores nulos y listo para el análisis.

```
data_cuanti=data.select_dtypes(include=["float64","int64","float","int"])
data_cuali=data.select_dtypes(include=["object","datetime","category"])

cuantias_mean=data_cuanti.fillna(round(data_cuanti.mean(),1))
cualis_bfill=data_cuali.fillna(method="bfill")
cualis_ffill=data_cuali.fillna(method="ffill")

data_sin_nulos = pd.concat([cuantias_mean,cualis_ffill],axis=1)

data_sin_nulos
```

### Código utilizado para la limpieza

	botón correcto	tiempo de interacción	número de interacción	auto push	tiempo de lectión	tiempo de sesión	botón correcto	tiempo de interacción	número de interacción	auto push	tiempo de lectión	tiempo de sesión
0	1.0	5.399169	1.0	0.0	0.000000	0.000000	1.0	5.399169	1.0	0.0	0.000000	0.000000
1	0.0	1.283400	2.0	0.0	0.000000	0.000000	0.0	1.283400	2.0	0.0	0.000000	0.000000
2	1.0	2.700226	3.0	0.0	0.000000	0.000000	1.0	2.700226	3.0	0.0	0.000000	0.000000
3	0.0	3.050262	4.0	0.0	0.000000	0.000000	0.0	3.050262	4.0	0.0	0.000000	0.000000
4	0.0	4.760256	5.0	0.0	0.000000	0.000000	0.0	4.760256	5.0	0.0	0.000000	0.000000
...	...	...	...	...	...	...	...	...	...	...	...	...
5860	0.5	10.000000	13.3	0.0	6.300000	332.240000	1.0	7.099429	4.0	0.0	7.461668	332.240000
5861	0.0	2.135419	1.0	0.0	0.000000	0.000000	0.0	2.135419	1.0	0.0	0.000000	0.000000
5862	0.5	10.000000	13.3	0.0	2.271806	12.400000	0.0	2.135419	1.0	0.0	2.271806	0.000000
5863	0.5	10.000000	13.3	0.0	6.300000	6.478299	0.0	2.135419	1.0	0.0	2.271806	6.478299
5864	0.5	10.000000	13.3	0.0	6.300000	12.400000	0.0	2.135419	1.0	0.0	2.271806	6.478299

### Resultados

# Tratamiento de “outliers”

Una parte importante del código fue la eliminación de valores atípicos (outliers) usando el método del rango intercuartílico (IQR). Esta técnica identifica los valores extremos que se alejan demasiado del comportamiento general de la muestra.

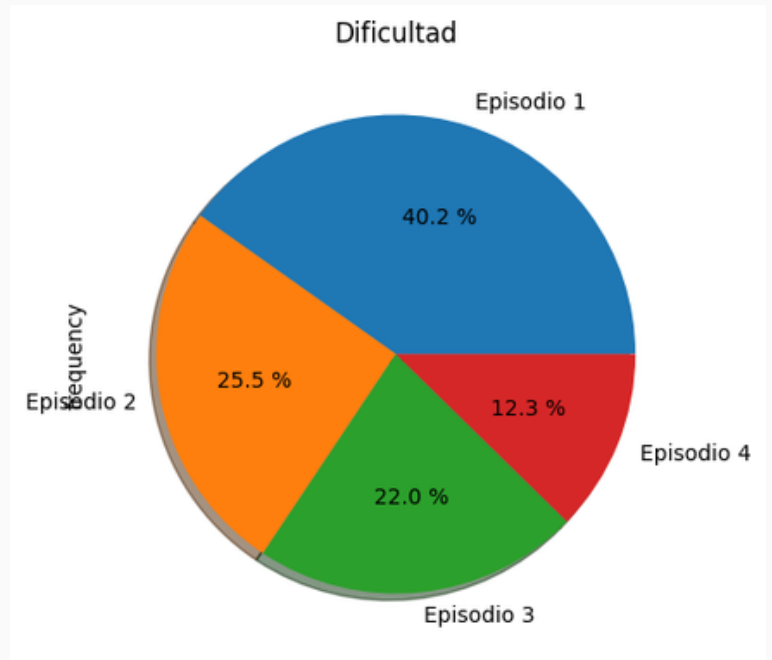
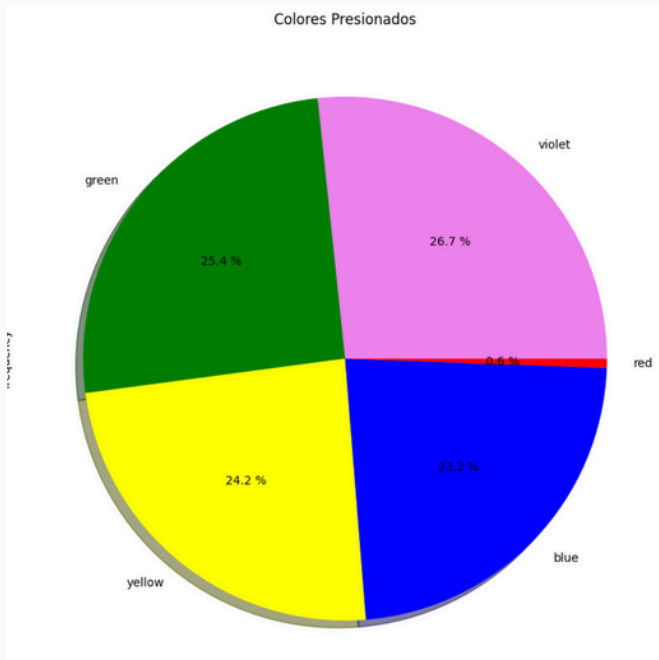
	Administrador	Usuario	mini juego	color presionado	dificultad	fecha	Juego	botón correcto	tiempo de interacción	número de interacción	auto push	tiempo de lección	tiempo de sesión
0	nicolas	nicolas	Despegue	blue	Episodio 1	25/01/2024 09:26:42 a. m.	Astro	1.0	5.399169	1.0	0.0	0.000000	0.000000
1	nicolas	nicolas	Despegue	violet	Episodio 1	25/01/2024 09:26:46 a. m.	Astro	0.0	1.283400	2.0	0.0	0.000000	0.000000
2	nicolas	nicolas	Despegue	green	Episodio 1	25/01/2024 09:26:48 a. m.	Astro	1.0	2.700226	3.0	0.0	0.000000	0.000000
3	nicolas	nicolas	Despegue	green	Episodio 1	25/01/2024 09:26:57 a. m.	Astro	0.0	3.050262	4.0	0.0	0.000000	0.000000
4	nicolas	nicolas	Despegue	green	Episodio 1	25/01/2024 09:26:58 a. m.	Astro	0.0	4.750256	5.0	0.0	0.000000	0.000000
...	...	...	...	...	...	...	...	...	...	...	...	...	...
5860	ALEIDA	ESMERALDA	NaN	NaN	Episodio 1	28/05/2024 04:15:49 p. m.	Astro	0.5	10.000000	13.3	0.0	6.300000	332.240000
5861	ALEIDA	JOSE JAVIER	Asteroides	green	Episodio 3	04/06/2024 11:09:54 a. m.	Astro	0.0	2.135419	1.0	0.0	0.000000	0.000000
5862	ALEIDA	JOSE JAVIER	Asteroides	NaN	Episodio 3	04/06/2024 11:09:58 a. m.	Astro	0.5	10.000000	13.3	0.0	2.271806	12.400000
5863	ALEIDA	JOSE JAVIER	NaN	NaN	Episodio 3	04/06/2024 11:09:58 a. m.	Astro	0.5	10.000000	13.3	0.0	6.300000	6.478299
5864	ALEIDA	JOSE JAVIER	MiniGame_0	NaN	Episodio 3	04/06/2024 11:09:59 a. m.	Astro	0.5	10.000000	13.3	0.0	6.300000	12.400000

Aplicado sobre las variables como tiempo de interacción o tiempo de sesión, este filtrado permite mantener una base de datos más representativa, descartando sesiones inusualmente largas o registros con errores extremos.

El resultado fue un conjunto de datos más limpio y centrado en comportamientos típicos.

# Análisis Univariado

El análisis univariado del código sirve para revisar una columna a la vez y entender qué tipo de datos tiene. Ayuda a ver si hay datos vacíos, si los valores son muy altos o bajos, y cómo están distribuidos. Todo esto se hace sin compararla con otras columnas.



***Estas son algunas de las gráficas presentadas en el código***

Por ejemplo, en estos dos casos, el color presionado y la dificultad, indican cuales son los colores que mas le puede llamar la atención al paciente; al igual puedes comparar el progreso de los pacientes viendo el tiempo que llevan en el programa y la dificultad. Estas son variables importantes para la toma de decisiones, debido a que con esto podemos hacer una base mas solida y llamativa para el tratamiento de los pacientes.

Algunas otras variables que llaman la atención es, el botón correcto y número de interacción indican la precisión y frecuencia con la que el usuario responde.

El tiempo de interacción y el tiempo de lección permiten medir el nivel de concentración o el grado de permanencia en una actividad. El campo llamado auto push parece estar relacionado con alguna funcionalidad automatizada del sistema, y aunque no se tiene contexto completo, también se trató como una métrica numérica. En conjunto, estas variables sirven como indicadores clave para evaluar el rendimiento, el compromiso y posiblemente el aprendizaje del usuario.

# Categorización de variables

En el código se realizó una separación de las variables en dos grupos: cuantitativas y cualitativas. Las variables cuantitativas son aquellas que contienen números, como el tiempo de sesión, número de interacción o auto push, ya que se pueden medir y hacer operaciones matemáticas con ellas.

Por otro lado, las variables cualitativas son las que tienen texto o categorías, como dificultad, color presionado, mini juego o administrador, y sirven para clasificar o describir ciertos aspectos, pero no se pueden sumar ni promediar. Esta clasificación es muy útil porque permite aplicar distintos métodos de análisis según el tipo de dato, lo cual ayuda a entender mejor la información y a tomar decisiones más acertadas durante el procesamiento de los datos.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5865 entries, 0 to 5864
Data columns (total 13 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Administrador                        5865 non-null   object
1   Usuario                             5865 non-null   object
2   botón correcto                    5103 non-null   float64
3   tiempo de interacción               5103 non-null   float64
4   mini juego                          5709 non-null   object
5   número de interacción               5103 non-null   float64
6   color presionado                    5103 non-null   object
7   dificultad                          5865 non-null   object
8   fecha                              5865 non-null   object
9   Juego                              5865 non-null   object
10  auto push                           5103 non-null   float64
11  tiempo de lección                   5688 non-null   float64
12  tiempo de sesión                    5259 non-null   float64
dtypes: float64(6), object(7)
memory usage: 595.8+ KB
```

# Conclusión

Después de limpiar la base de datos, eliminar los espacios vacíos y quitar los valores que no seguían el comportamiento normal, se logró tener un conjunto de datos más confiable y ordenado. Esto significa que ahora sí se pueden hacer análisis y sacar conclusiones sin preocuparse por errores o información incompleta.

Por ejemplo, ahora se pueden comparar los resultados entre diferentes niveles de dificultad, analizar qué juegos son más efectivos o ver cuánto tiempo dedican en promedio los usuarios. También se pueden hacer gráficas, resúmenes o modelos que nos digan quién está aprendiendo más, qué partes del sistema son más atractivas o dónde podrían estar los errores. En resumen, este proceso de preparación hizo que los datos sean mucho más útiles y valiosos para tomar decisiones.

