

# **Regresión simple y multiple**



**Tecnológico  
de Monterrey**

A 06 de abril de 2025

## **Regresión simple y multiple para datos de “Wuupi”**

**Equipo 3:**

Braulio González Esquivel || A01425274

Lucero Jannete López García || A01736938

Pilar Méndez Briones || A01736843

Rubén Uriel Flores Bello || A01737098

## **Analítica de datos y herramientas de inteligencia artificial II (Gpo 501)**

**Profesores: Juan Manuel Ahuactzin Larios,  
Candy Yuridiana Alemán Muñoz,  
Rigoberto Cerino Jiménez y  
Alfredo García Suárez**

**CC: Sexto semestre profesional – Escuela de Negocios  
– Licenciatura en Administración y Transformación de  
Negocios.**

# Regresión simple y multiple para datos de



# Índice

## Contenido

Índice .....	2
Introducción .....	3
Correlaciones lineales simples (mapa de calor) .....	4
Comparativa de las principales correlaciones.....	5
Correlación de la variable usuario .....	6
Modelo de regresión múltiple .....	7
Dataframe por “usuario” .....	8
Bibliografía.....	16

## *Introducción*

Wuupi es un startup mexicano que desarrolla soluciones tecnológicas innovadoras para mejorar la calidad de vida de personas con discapacidad y sedentarismo. Su línea de productos combina videojuegos terapéuticos con controles adaptativos, permitiendo realizar terapias físicas y conductuales.

La empresa nació de las experiencias personales de sus fundadores con familiares que enfrentan dificultades para acceder a terapias efectivas y atractivas. Su tecnología se basa en algoritmos pedagógicos y herramientas de medición de progreso, lo que permite realizar un seguimiento detallado del progreso de los usuarios.

Principales actividades (Wuupi. (s.f.):

- Se especializa en el desarrollo de algoritmos pedagógicos para personalizar las experiencias de juego.
- Incorpora mediciones de avances de los usuarios en términos de tiempo de reacción, eficiencia motriz y otros indicadores clave.
- Su plataforma integra un tablero interactivo, que permite a terapeutas y administradores visualizar métricas y generar informes detallados.
- Utiliza métricas de uso en cada sesión de juego, analizando la interacción del usuario en términos de aciertos, errores y patrones de respuesta.
- Su enfoque es proporcionar herramientas que facilitan la interpretación de datos y que permiten adaptar las terapias a cada usuario.

Correlaciones lineales simples  
(mapa de calor)



Elaboración propia, (con base en la base de datos principal de Wuupi): gráfica de pastel que muestra la distribución de los tiempos de consulta más frecuentes, [marzo de 2025].

## *Comparativa de las principales correlaciones*

Variables relacionadas	correlación	Interpretación
Juego - mini juego	0.6257	La alta correlación sugiere que los mini juegos son una parte fundamental del compromiso de los niños con el juego completo. En contextos de discapacidad, esto podría significar que los mini juegos ofrecen dinámicas más accesibles, claras o motivadoras que capturan mejor su atención o se ajustan a sus capacidades específicas.
Botón correcto - número de interacción	0.4670	Esta relación indica que, en la medida en que los niños interactúan más veces, su probabilidad de presionar el botón correcto aumenta. Esto puede reflejar un proceso de aprendizaje progresivo, ensayo-error o mejora en la familiarización con la dinámica del juego. Importante: no necesariamente más interacciones implican mayor habilidad, sino un proceso adaptativo que debe valorarse.
Botón correcto - color presionado	0.4568	Esta correlación sugiere que el color presionado guarda una relación importante con el botón correcto. Para niños con discapacidades visuales, cognitivas o atencionales, los colores pueden servir como guías visuales o estímulos clave para la toma de decisiones. Esto puede ser aprovechado para diseñar interfaces más intuitivas.
Tiempo de interacción - botón correcto	0.2797	A mayor tiempo de interacción, mayor precisión. Esto puede indicar que los niños necesitan más tiempo para procesar información y tomar decisiones. La variable "tiempo" puede ser un factor crítico a adaptar según el tipo de discapacidad
Mini juego - botón correcto	0.2322	Los mini juegos tienen una influencia moderada en la precisión. Es posible que estos juegos sirvan como entrenamiento cognitivo o psicomotor, o bien que brinden motivación adicional para desempeños más precisos. Su diseño puede estar ayudando a mejorar la comprensión o la concentración.
Color presionado - número de interacción	0.2724	Esta relación puede reflejar una exploración de opciones por parte del niño. Es importante considerar si la diversidad de colores estimula o distrae. En algunos casos, puede estar relacionado con la preferencia sensorial o visual del usuario más que con una lógica de juego en sí.

## *Correlación de la variable usuario*

Variables relacionadas	Correlación	Interpretación
Usuario-Juego	0.3365	Los usuarios muestran una relación moderada con su participación en el juego. Esto podría estar vinculado a factores como la motivación individual, el nivel de accesibilidad percibida o el grado de familiarización con la tecnología. La personalización puede mejorar el compromiso
Usuario - mini juego	0.2442	Una parte importante de los usuarios interactúa con mini juegos, lo cual puede sugerir que estos representan un espacio más seguro o comprensible para explorar. Pueden ser útiles como evaluadores indirectos de atención, habilidades motoras o comprensión.
Usuario - botón correcto	0.1737	La baja correlación indica que la precisión (aciertos) no depende fuertemente del perfil de usuario. Esto puede deberse a una amplia diversidad de habilidades entre los niños, lo que refuerza la importancia de ajustes individualizados para medir desempeño.
Usuario- Número de interacción	0.1288	Aunque débil, hay una tendencia de que algunos usuarios interactúan más que otros. Es relevante considerar que una mayor interacción no necesariamente refleja mayor competencia, sino quizá necesidad de repetición, confusión o perseverancia.
Usuario-tiempo de interacción	0.0699	Muy baja relación. Esto puede indicar que cada niño gestiona su tiempo de forma muy particular. Hay que evitar asumir que "más tiempo" o "menos tiempo" implica mejor o peor desempeño: puede estar influenciado por factores motores, sensoriales, emocionales, etc.
Usuario-color presionado	0.0681	El color que presionan los niños no está relacionado de forma fuerte con su identidad. Esto puede ser una señal de que no hay una preferencia marcada por colores, o que el diseño no hace del color un factor determinante para el usuario. También puede indicar que los estímulos visuales deben adaptarse mejor a las necesidades individuales.

## *Modelo de regresión múltiple*

Es por ello que, en base a las correlaciones del usuario, tomamos las correlaciones más altas para genera un modelo de regresión lineal múltiple con mayor correlación de acuerdo al usuario y sus variables.

En este caso, las que se utilizaron como variable independiente fueron: Bóton correcto, Juego y minijuego.

```
#Declaramos las variables dependientes e independientes para la regresión lineal
Vars_Indep= df[['botón correcto', 'Juego','mini juego']]
Var_Dep= df['Usuario']
```

*Elaboración propia, (con base en la base de datos principal de Wuupi): gráfica de pastel que muestra la distribución de los tiempos de consulta más frecuentes, [marzo de 2025].*

Esto generó que el modelo diera una correlación del 0.37 que es mayor que todas las variables independientes de acuerdo a la variable dependiente que es usuario, sin embargo, esta correlación del 37% es muy baja y sigue sin explicar algo a gran detalle, de igual forma el modelo da un score del 0.1373, así que , en un contexto educativo o conductual, un R del 13% puede considerarse útil, especialmente cuando se trabaja con poblaciones diversas o con discapacidad, donde muchos otros factores no medidos influyen en los resultados.

Desde una perspectiva de modelo predictivo, este valor no es muy alto, por lo que no sería suficiente para predecir con precisión el comportamiento del usuario solo con las variables analizadas. En términos de diseño de intervención, el resultado sugiere que las variables "Juego" y "Mini juego" tienen cierto peso en la experiencia del usuario, por lo que pueden ser útiles para adaptar las estrategias, aunque por sí solas no son determinantes en el resultado final.

```
#Corroboramos cual es el coeficiente de Determinación de nuestro modelo
coef_Deter=model.score(X=Vars_Indep, y=Var_Dep)
coef_Deter
✓ 0.4s
```

0.13731646502302441

*Elaboración propia, (con base en la base de datos principal de Wuupi): gráfica de pastel que muestra la distribución de los tiempos de consulta más frecuentes, [marzo de 2025].*



```
#Corroboramos cual es el coeficiente de Correlación de nuestro modelo  
coef_Correl=np.sqrt(coef_Deter)  
coef_Correl
```

✓ 1.4s

0.3705623632035833

*Elaboración propia, (con base en la base de datos principal de Wuupi): gráfica de pastel que muestra la distribución de los tiempos de consulta más frecuentes, [marzo de 2025].*

## ***Dataframe por “usuario”***

**Primer paciente: Valentin**



Hay muchos valores que se repiten por eso salen valores nulos, porque no los cuenta al ser lo mismo., pero podemos observar que el que cuenta con mayor correlación es botón correcto con color presionado, esto debido a que los colores son parte esencial del juego dónde si responde bien o mal se ve relacionado con el las pulsaciones.

Para poder realizar su regresión lineal múltiple, de las variables x se tomó la que tuviera una correlación más alta para que fuera la variable dependiente, de ahí, se tomaron los siguientes 3 valores más altos de Y de acuerdo con X para hacer una correlación más alta.

Las variables que se tomaron en cuenta fueron:

```
#Declaramos las variables dependientes e independientes para la regresión lineal
Vars_IndepV= df1[['tiempo de interacción', 'número de interacción','color presionado']]
Var_DepV= df1[['botón correcto']]
```

✓ 0.9s

Generando que tuviéramos una correlación de 0.60 como se muestra en la siguiente imagen, De igual forma, indica una relación fuerte baja y positiva entre las variables independientes y el acierto en la elección del botón. Esto sugiere que, en general, a mayor tiempo o número de interacciones, y dependiendo del color presionado, hay una mayor probabilidad de que el usuario presione el botón correcto.

En términos prácticos, esta correlación implica que estas variables sí aportan información relevante sobre el comportamiento del usuario y, por tanto, podrían ser consideradas para diseñar adaptaciones en la experiencia o mecánicas del juego.

```
coef_DeterV=modelV.score(X=Vars_IndepV, y=Var_DepV)
coef_DeterV
```

✓ 0.3s

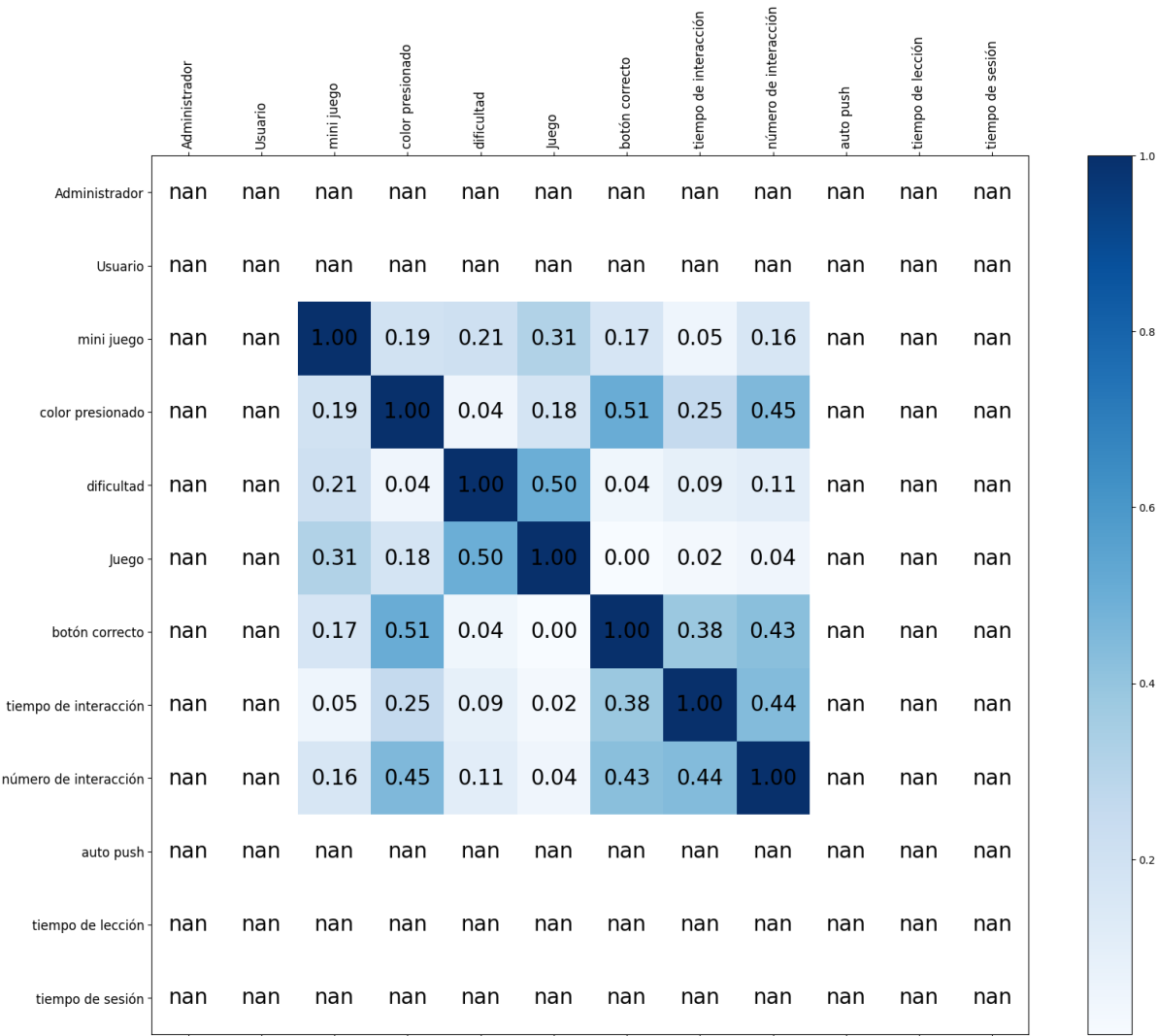
0.3669722632565591

```
#Corroboramos cual es el coeficiente de Correlación de nuestro modelo
coef_CorrelV=np.sqrt(coef_DeterV)
coef_CorrelV
```

✓ 0.2s

0.6057823563430674

Segundo paciente Yael david



De igual forma se pueden observar que hay varios valores nulos, porque son números repetidos y que no tiene correlación alguna, generando que se rellenen con nulos y también se repite el mismo patrón de botón correcto y color presionando.

Asimismo, el procedimiento para generar un modelo de regresión lineal múltiple fue el mismo que el del primer paciente, se tomó en cuenta la variable dependiente de la variable que tuviera mayor correlación, en este caso fue juego y de las siguientes correlaciones en Y de la misma variable dependiente fueron color presiona, dificultad y minijuego.

Generando que la correlación fuera de 0.55, siendo un valor relativamente bajo

```
coef_DeterYa=modelYa.score(X=Vars_IndepYa, y=Vars_DepYa)
coef_DeterYa
```

✓ 0.1s

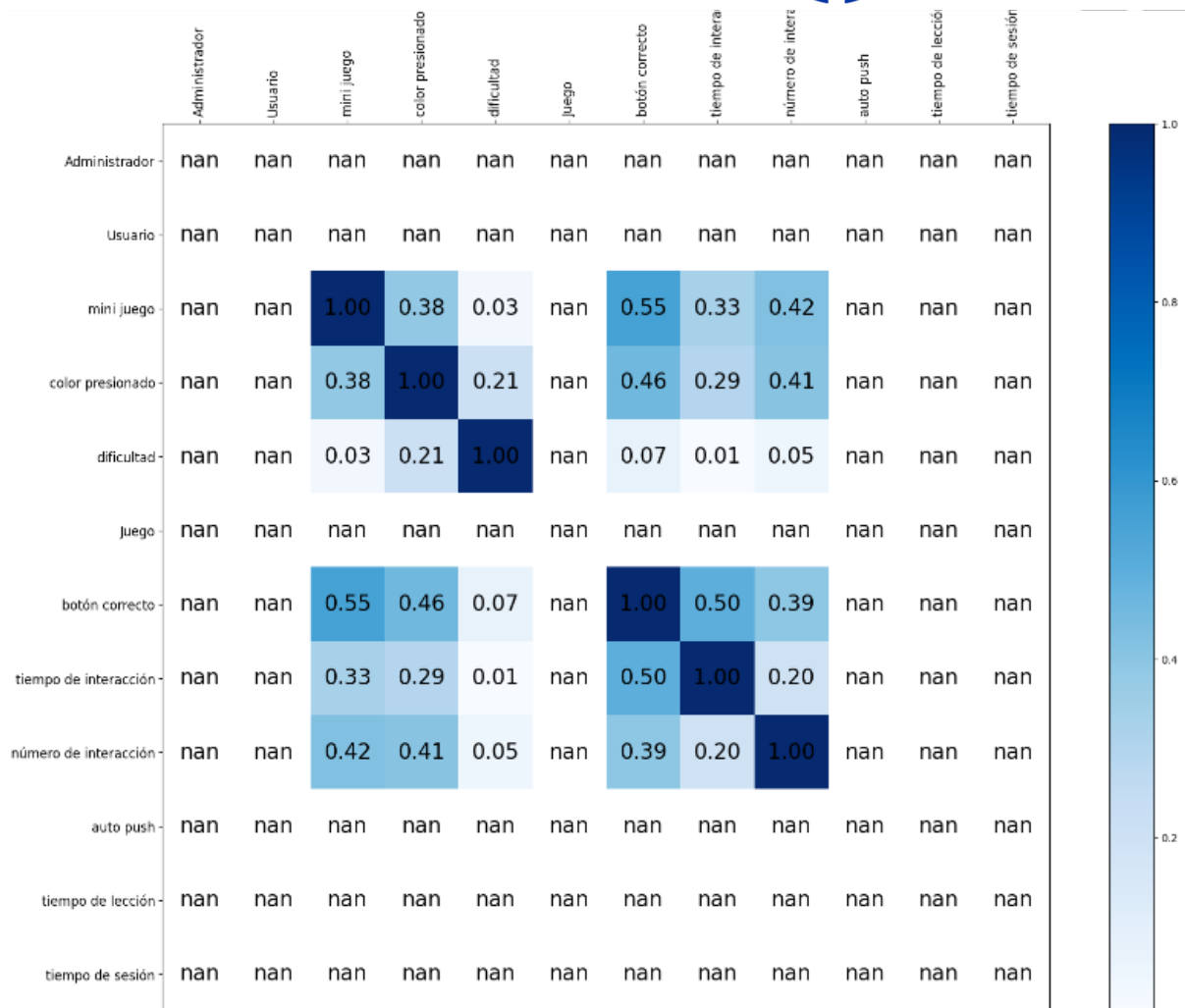
0.31062629060696356

```
#Corroboramos cual es el coeficiente de Correlación de nuestro modelo
coef_CorrelYa=np.sqrt(coef_DeterYa)
coef_CorrelYa
```

✓ 0.2s

0.5573385780716813

Tercer paciente: Yeremi Yazmin



### Proceso

Primero que nada, nos dimos cuenta que menos del 50% de los datos son utilizables, esto se debe a que la mayoría o son repetidos o no son útiles para una correlación.

Para analizar los resultados, elegimos los valores más altos, ya que estos indican mayor correlación entre las variables involucradas.

De acuerdo a esas variables que elegimos, se construyó un modelo matemático para predecir si el paciente va a presionar correctamente el botón, usando como base tres cosas: cuánto tiempo se tarda en responder, qué mini juego está jugando y qué color presionó.

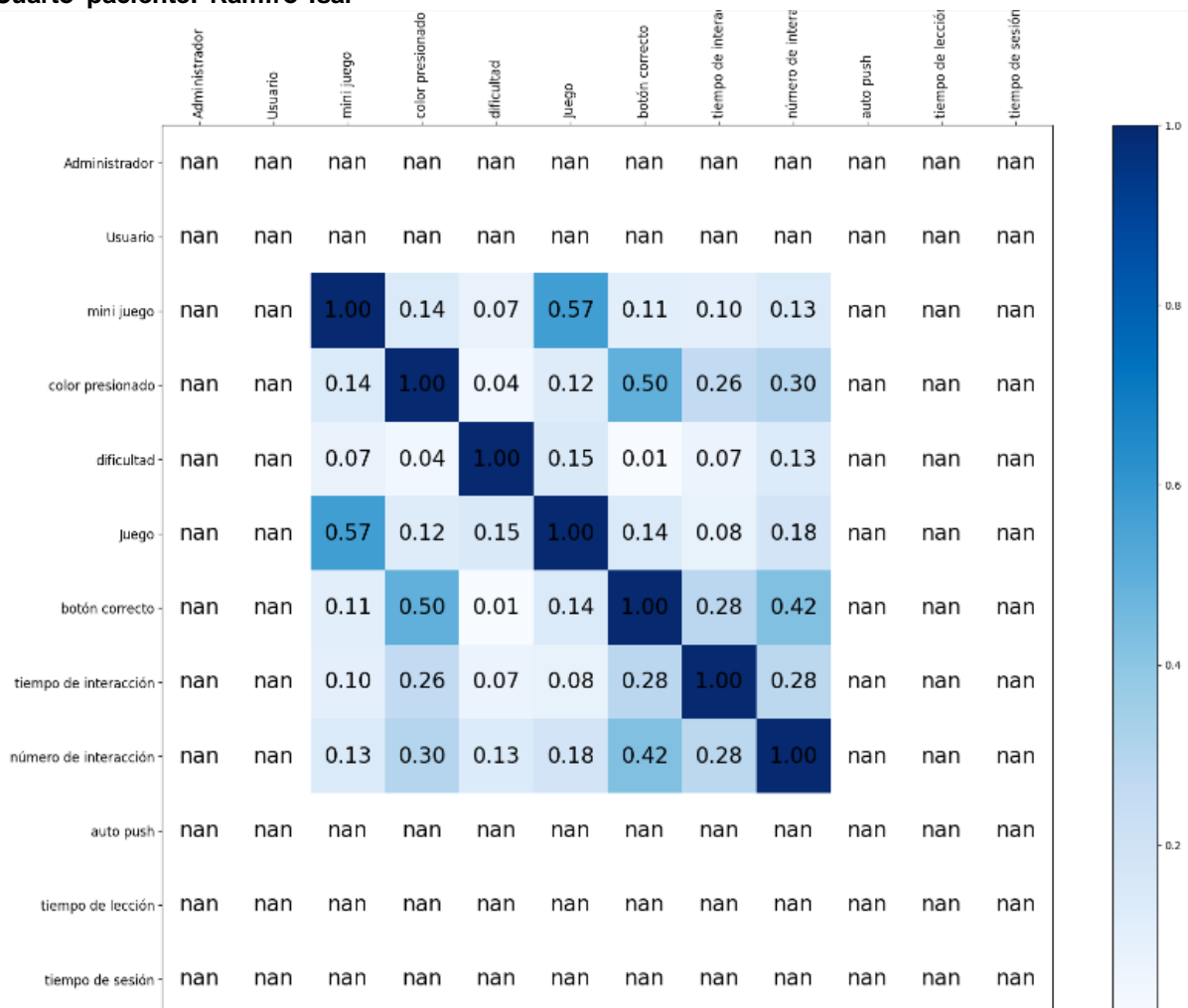
### Modelo

Modelo =  $-0.04 \times \text{tiempo\_interacción} + 0.11 \times \text{mini\_juego} + 0.10 \times \text{color\_presionado} + 0.61$

### Análisis

- A mayor tiempo de interacción, ligeramente baja la probabilidad de acertar (sorprendente, quizá por pocos datos). Mientras más tiempo pasa en el juego, menos acierta, creemos que a lo mejor se debe a que el tiempo de interacción suele ser mucho.
- Algunos juegos y colores aumentan las probabilidades de presionar el botón correctamente. Hay ciertos mini juegos y ciertos colores que están relacionados con un mayor número de aciertos al presionar el botón correcto. Es decir, cuando el paciente juega determinados mini juegos (por ejemplo, uno que ya conoce o que le resulta más intuitivo) o cuando se le presentan ciertos colores (que le son más fáciles de reconocer o recordar), tiene más probabilidades de responder bien.

### Cuarto paciente: Ramiro Isai



### Proceso

Se creó un modelo para intentar predecir qué tipo de juego le aparece al paciente (por ejemplo, "Astro" o "Cadetes"), usando como base tres aspectos de su comportamiento: cuántas veces ha interactuado, si presionó el botón correcto o no, y qué nivel de dificultad, tomando en cuenta que fueron las variables con mayor correlación.

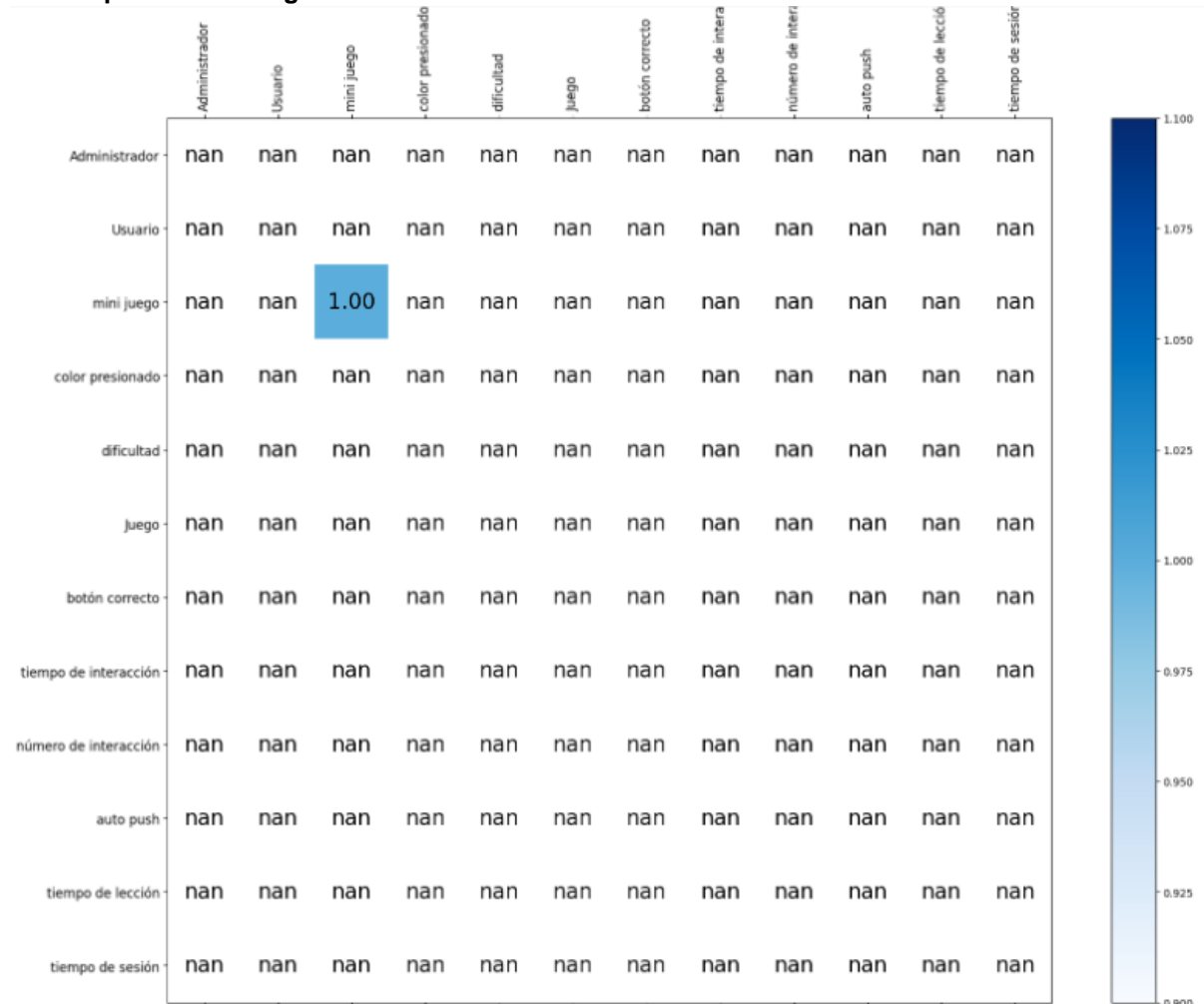
### Modelo

Modelo =  $-0.023 \times \text{numero de interacción} + 0.043 \times \text{boton correcto} + 0.09 \times \text{dificultad} + 0.26$

### Análisis

El modelo no representa una gran correlación

Quinto paciente: Irving



No hay datos válidos numéricos con los que se pueda hacer análisis. En el mapa de calor de Irving, todas las casillas son NaN, lo que significa que no se puede calcular ninguna relación. Esto se debe a que en el DataFrame, solo hay dos interacciones, lo cual es muy poco para poder analizar, además de que esas dos interacciones no tienen los registros (variables de las columnas) completos.



## ***Bibliografía***

Wuupi. (s.f.). Wuupi. Recuperado el 9 de marzo de 2025, de  
<https://wuupi.com.mx/>

El Economista. (2023, 10 de octubre). Conoce los 4 proyectos ganadores de los UP Innovation Awards 2023. Recuperado el 9 de marzo de 2025, de  
<https://www.eleconomista.com.mx/el-empresario/Conoce-los-4-proyectos-ganadores-de-los-UP-Innovation-Awards-2023-20231010-0113.html>

Valor Compartido. (2023, 4 de diciembre). Finalizan mentorías a emprendimientos sociales tecnológicos de T-Systems. Recuperado el 9 de marzo de 2025, de  
<https://valor-compartido.com/finalizan-mentorias-a-emprendimientos-sociales-tecnologicos-de-t-systems/>