

WUUPI AVANCE 3

EQUIPO 4 | Máquinas de chambeo



INICIO DE CÓDIGO

- Importamos las bibliotecas esenciales para el análisis
- Realizamos la carga de los datos por el archivo csv
- Realizamos la exploración de datos, y la búsqueda de valores nulos por columna

NULOS Y OUTLIERS

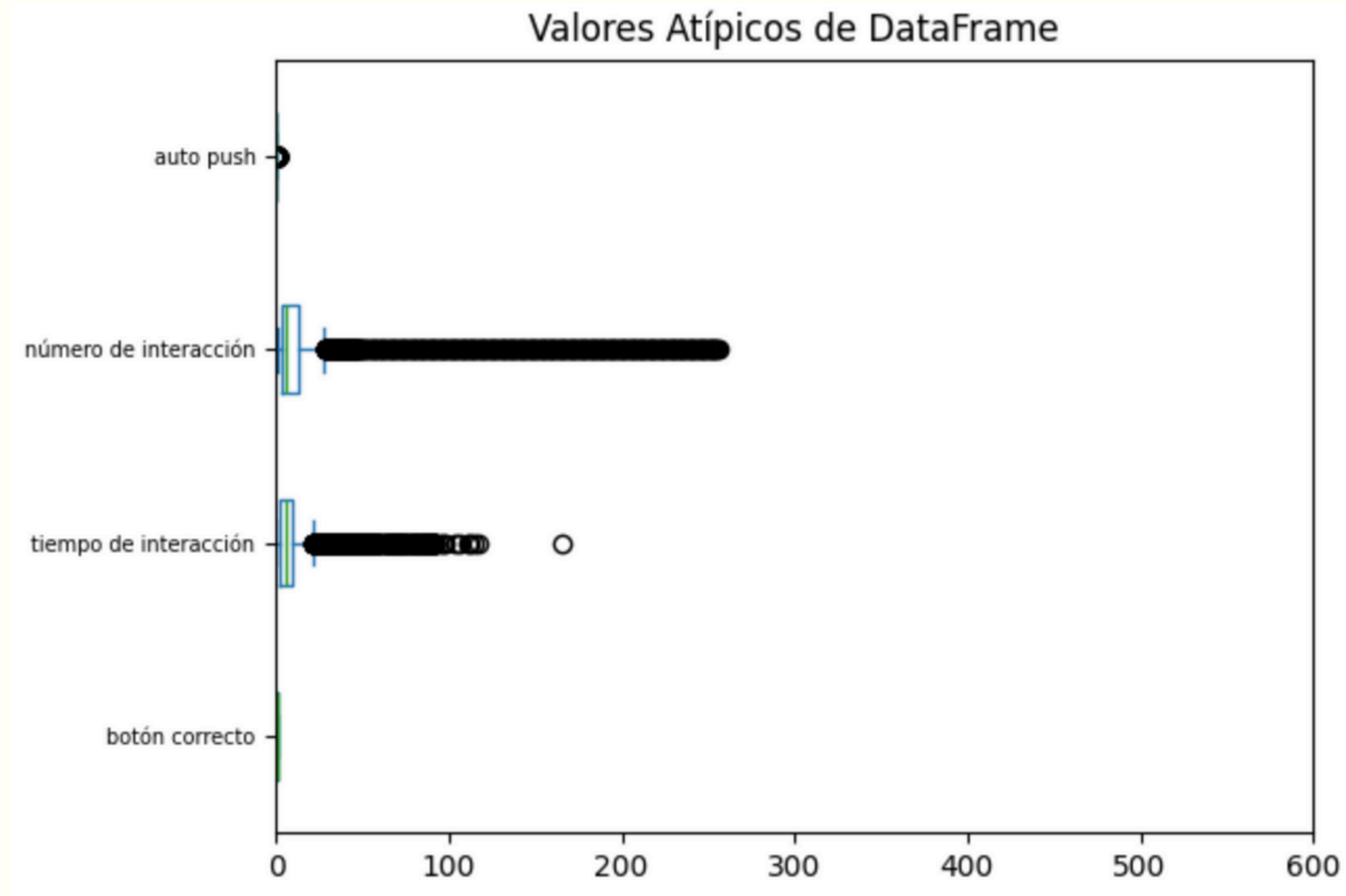
- Varias columnas tienen **valores nulos**, especialmente:
- "botón correcto", "tiempo de interacción", "número de interacción", "color presionado" y "auto push" con 762 valores nulos cada una.

SOLUCIÓN:

- Se **verifican** los tipos de **datos** de cada columna
- **Separamos** las variables por **numéricas y cuantitativas**:
 - Cuantitativas ('int64' & 'float64'): media
 - Cualitativas ('object'): "Sin dato"

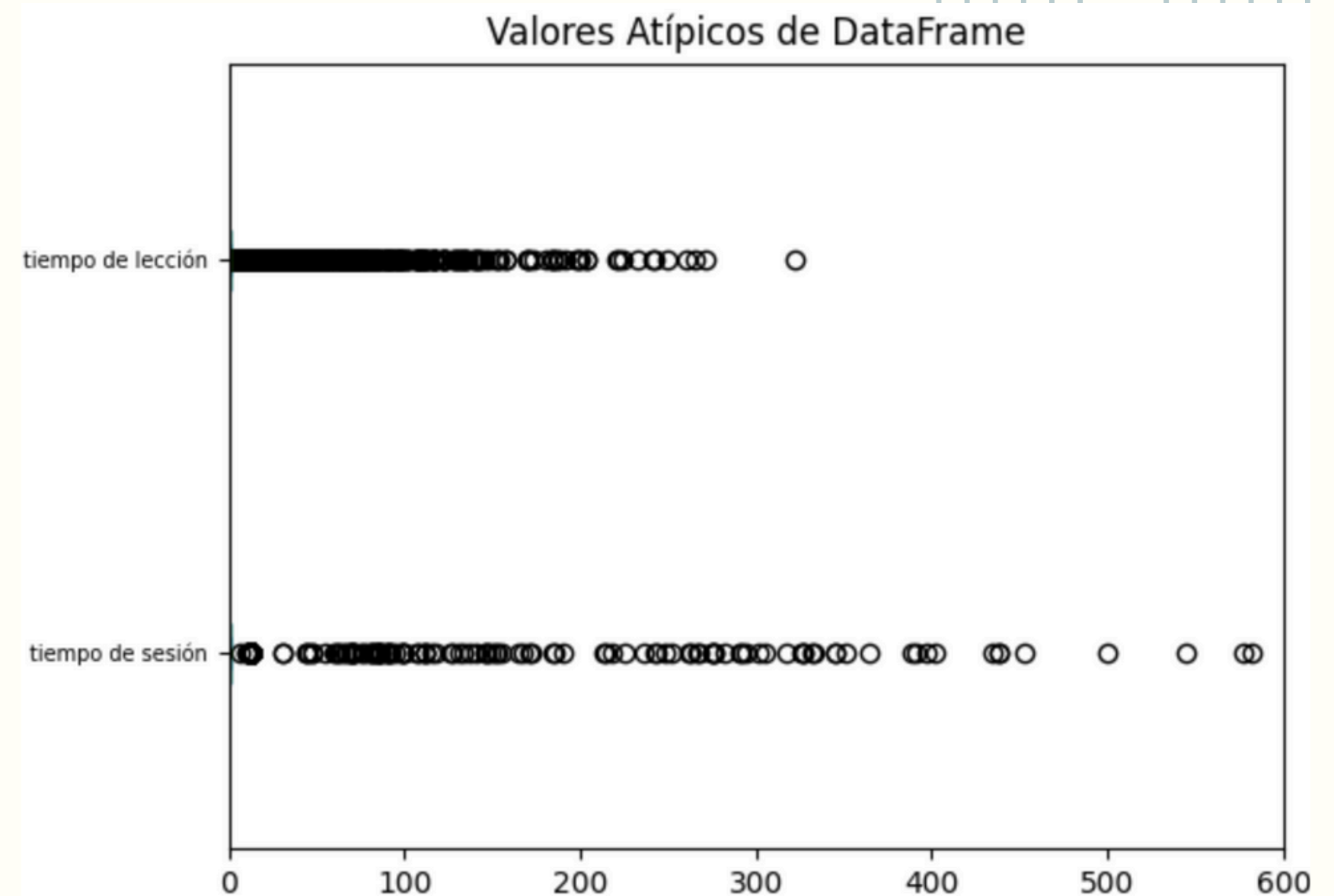


Valores Atípicos del DataFrame



Variables: auto push, número de interacción, tiempo de interacción y botón correcto

Método: rango intercuartílico (IQR)



Variables: tiempo de lección y tiempo de sesión (por la gran cantidad de registros "0")

Método: percentil 1 y 99

CONVERSIÓN DE VARIABLES

categorías —————> **numéricas**

método Frecuencia

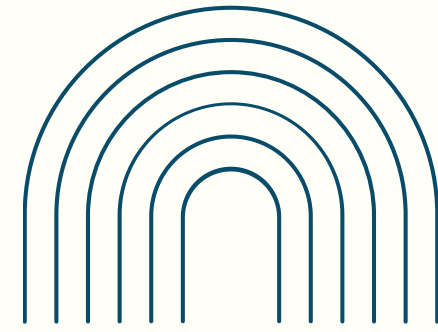
```
Administrador  
ALEIDA          3260  
nicolas         440  
LEONARDO        371  
DENISSE         302  
SERGIO ANGEL    243  
CARLOS ENRIQUE  228  
Yael DAVID      224  
AUSTIN          199  
VALENTIN        163  
erick           158  
IKER BENJAMIN   128  
KYTZIA          98  
BENJAMIN        51  
Name: Administrador, dtype: int64
```

Aleida	1
Nicolas	2
Leonardo	3
Denisse	4
Sergio Angel	5

ESTRATEGIA DE SELECCIÓN

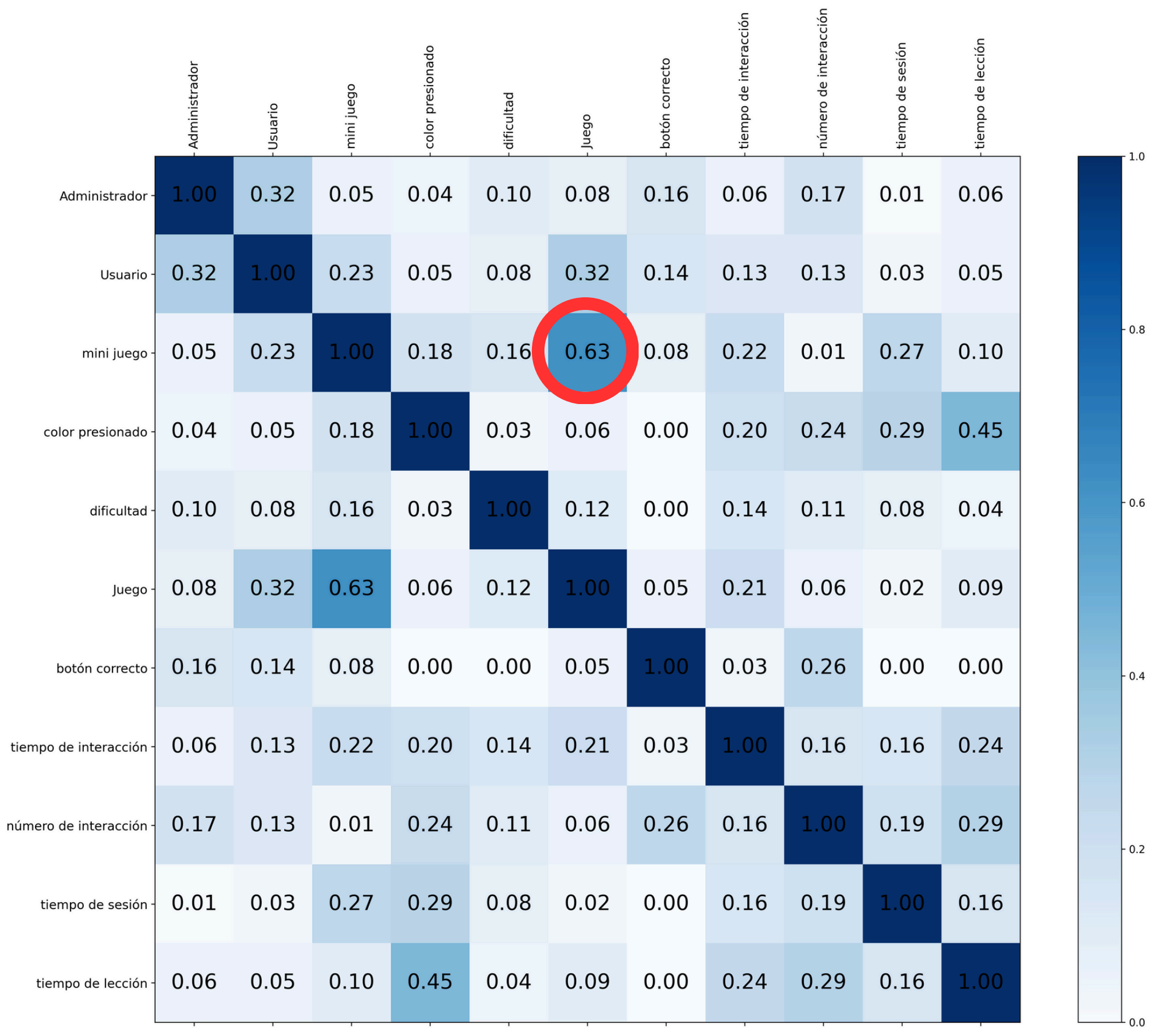
de variables a evaluar

- Revisamos el heatmap para encontrar las variables más relevantes
- Escogimos las variables objetivo y variables independientes que tuvieran los coeficientes más altos dentro de los heatmaps
- Aplicamos diferentes modelos no lineales para evaluar si era posible aumentar los coeficientes
- Escogimos aquellos modelos que elevaron más los coeficientes
- Comparamos los resultados



RESULTADOS





Heatmap General

Variable x	mini juego
Variable y	Juego
Modelo	Exponencial

COEFICIENTES

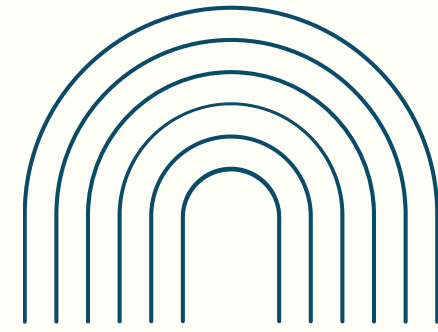
**REGRESIÓN
LINEAR SIMPLE**

0.64



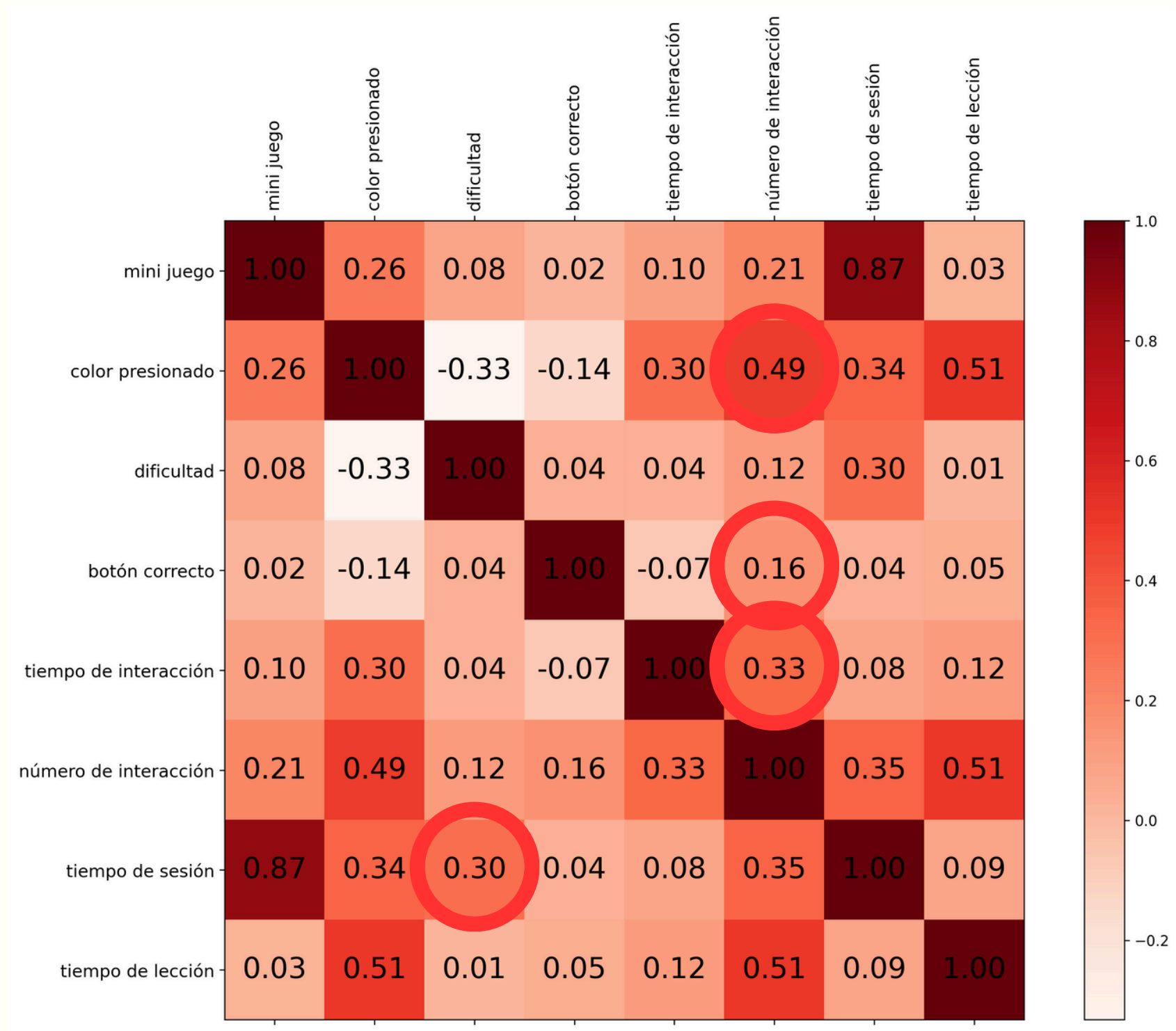
**MODELO NO
LINEAR**

0.68



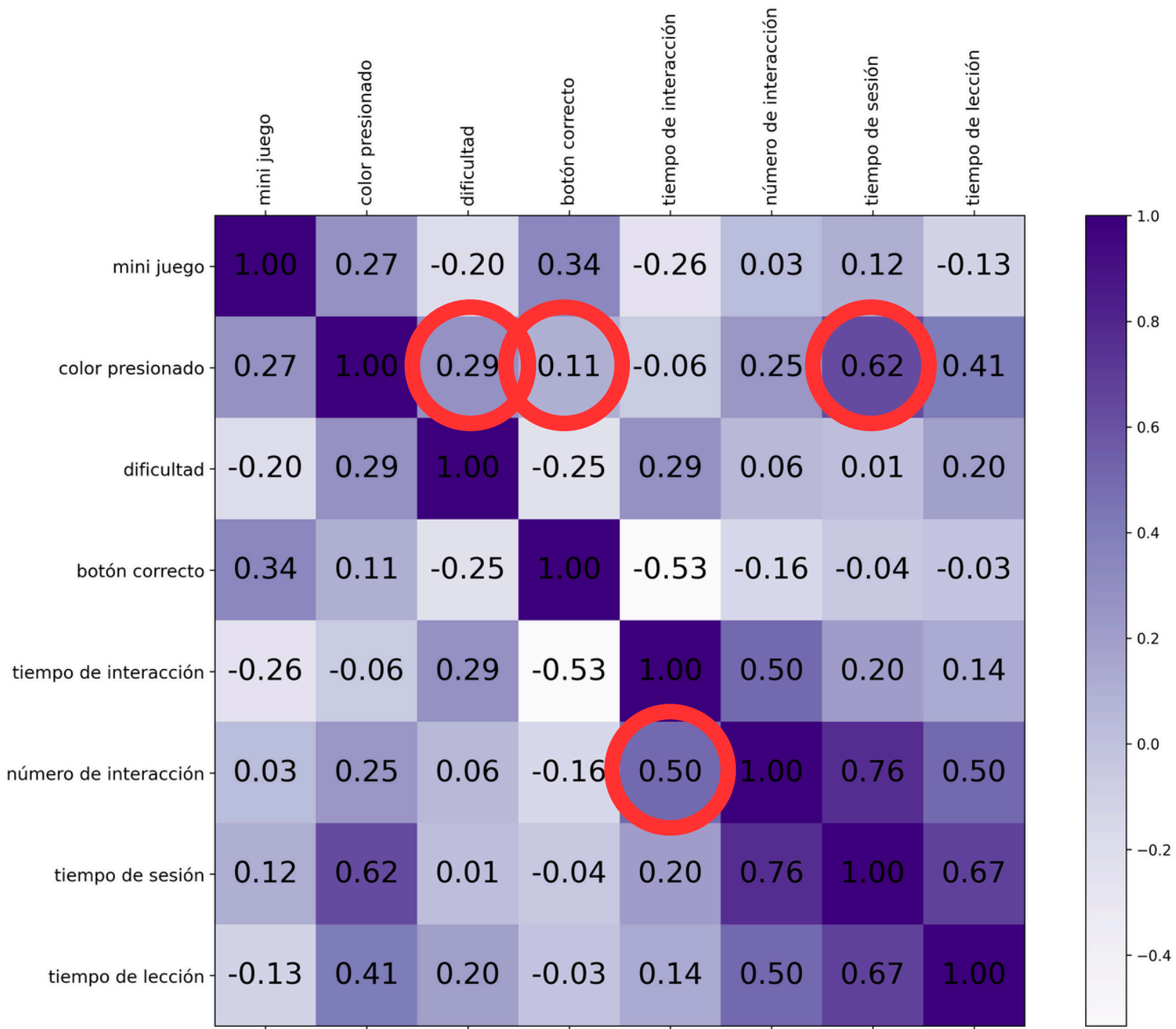
RESULTADOS POR USUARIO

Benjamín



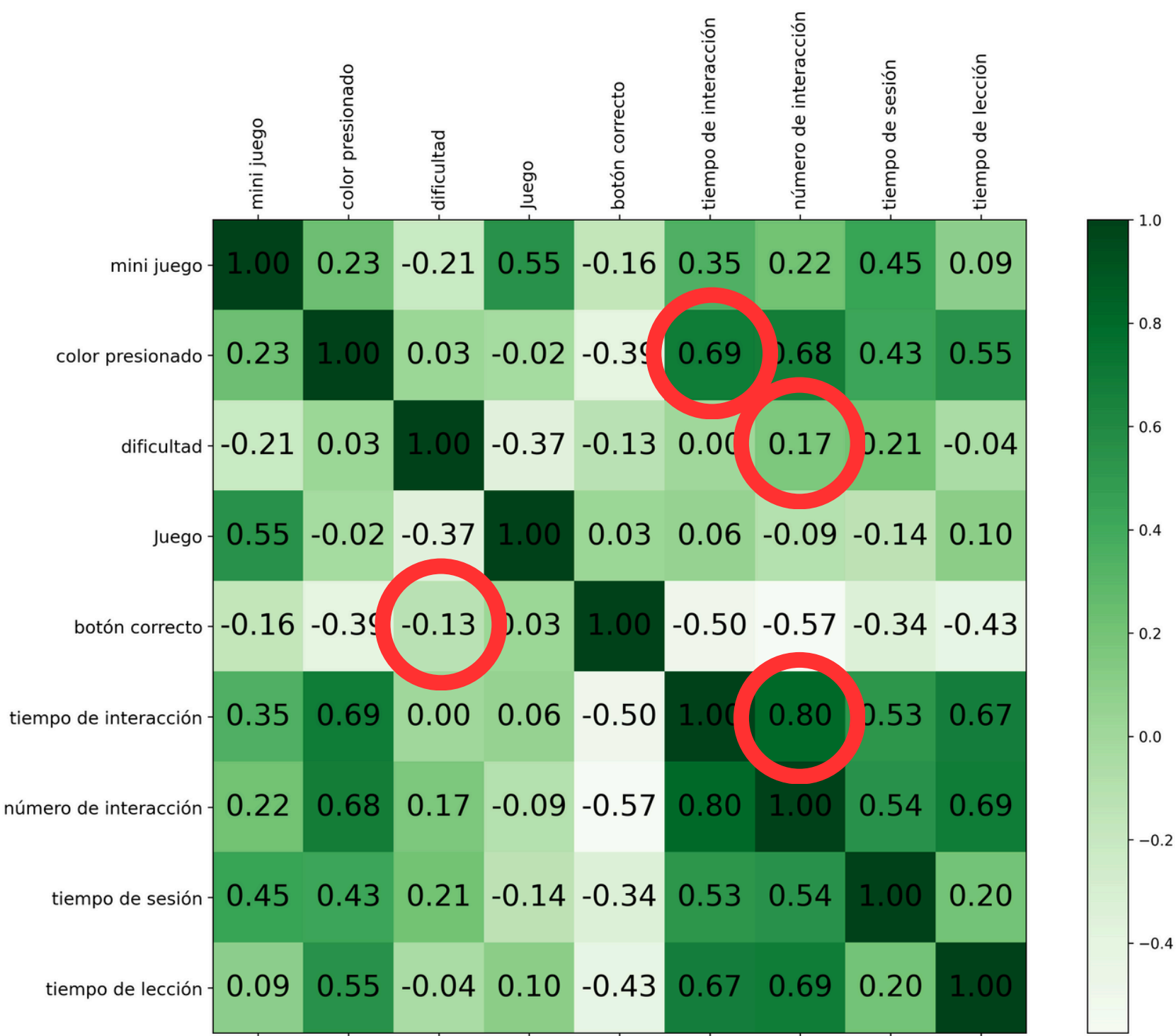
Variables	Función	Coef. lineal	Coef. no lineal
x = número de interacción y = tiempo de interacción	cuadrática	0.33	0.59
x = número de interacción y = botón correcto	exponencial	0.16	0.32
x = número de interacción y = color presionado	cuadrática	0.49	0.54
x = tiempo de sesión y = dificultad	cuadrática	0.30	0.34

Carlos Abel



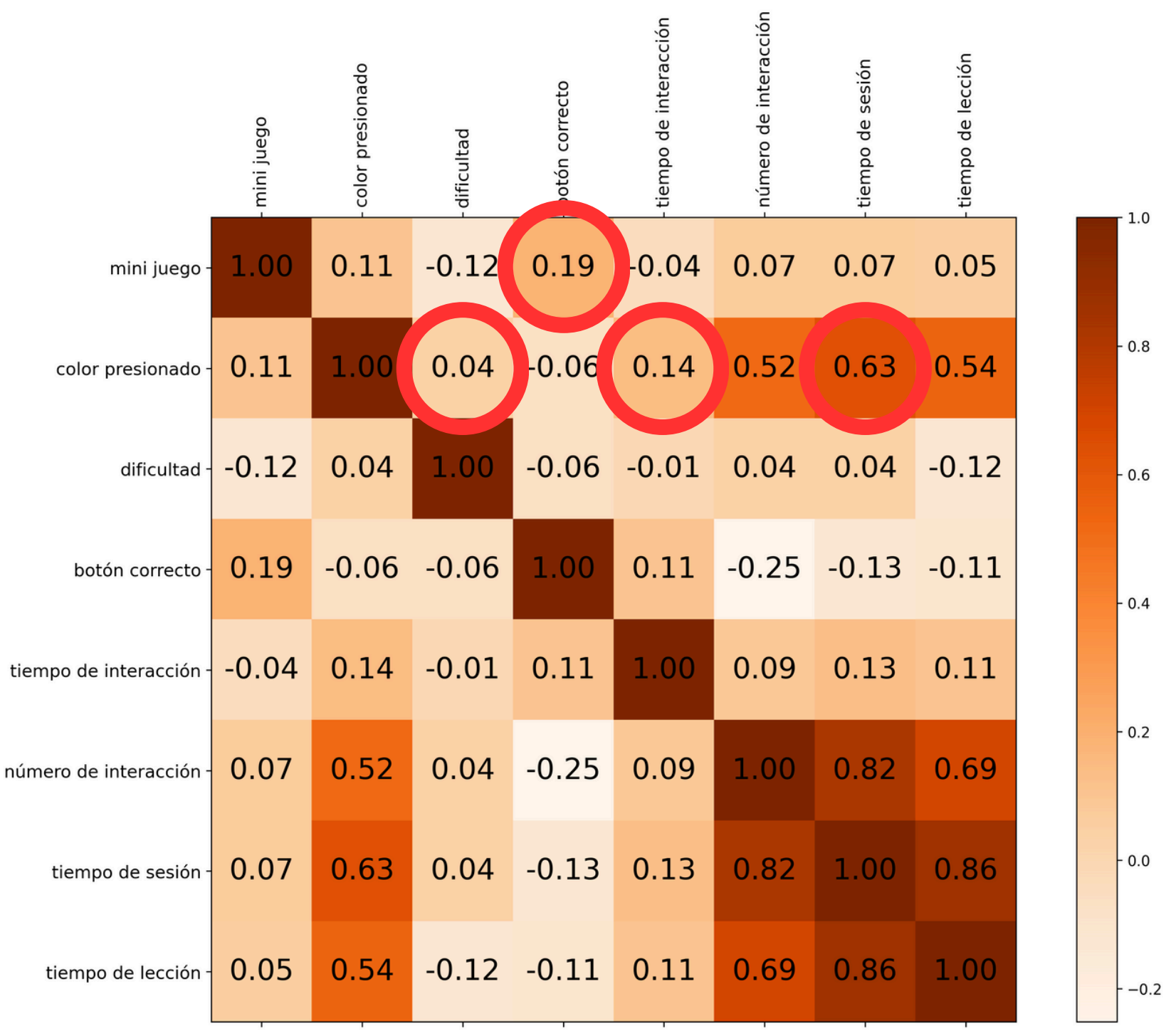
Variables	Función	Coef. lineal	Coef. no lineal
x = número de interacción y = tiempo de interacción	cuadrática	0.50	0.61
x = color presionado y = botón correcto	exponencial	0.11	0.23
x = tiempo de sesión y = color presionado	cuadrática	0.62	0.65
x = color presionado y = dificultad	exponencial	0.29	0.36

Carlos Enrique



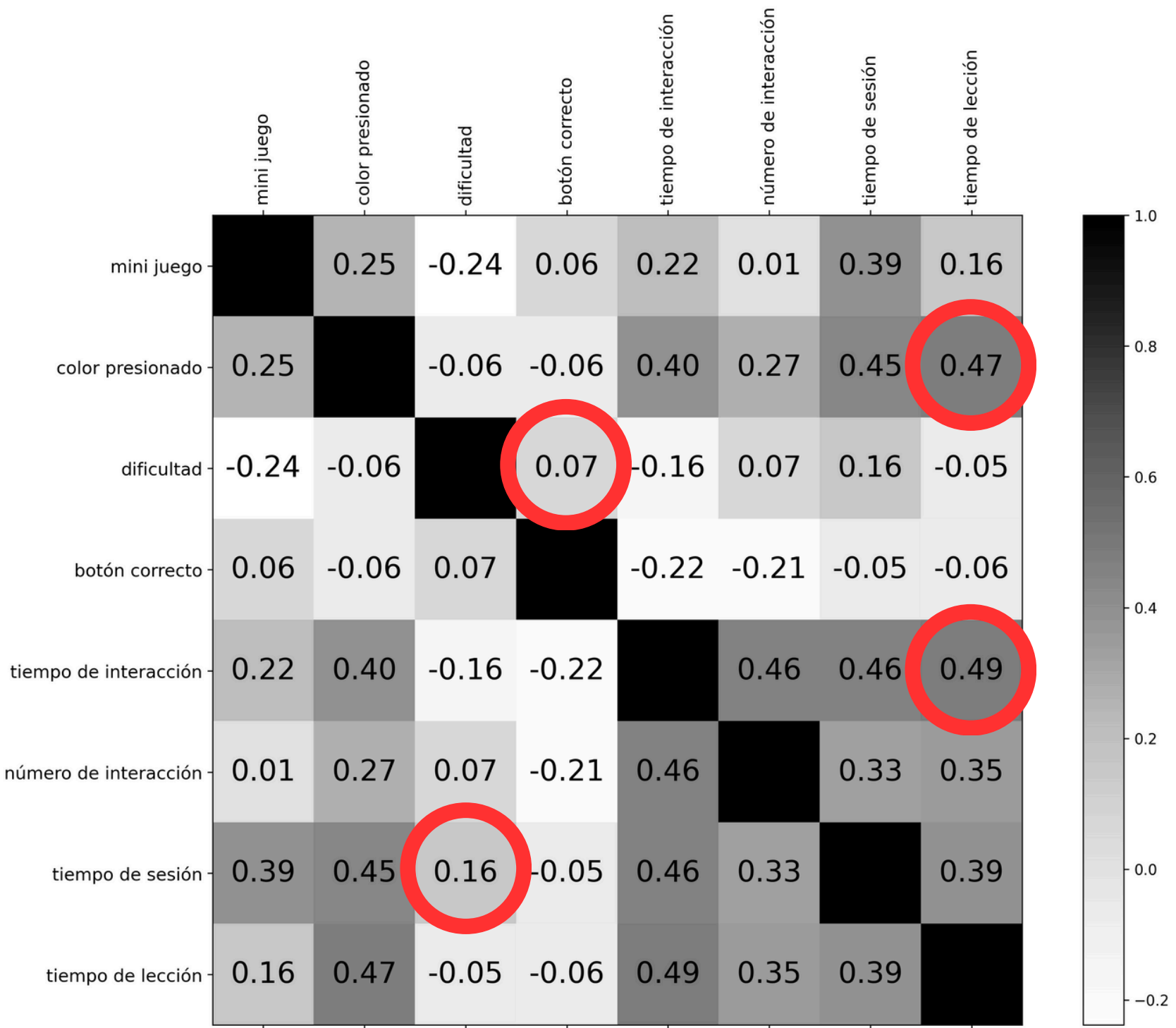
Variables	Función	Coef. lineal	Coef. no lineal
x = número de interacción y = tiempo de interacción	cuadrática	0.80	0.88
x = dificultad y = botón correcto	cuadrática	-0.13	0.14
x = tiempo de interacción y = color presionado	cuadrática	0.69	0.71
x = número de interacción y = dificultad	cuadrática	0.17	0.23

Concepción



Variables	Función	Coef. lineal	Coef. no lineal
x = color presionado y = tiempo de interacción	cuadrática	0.14	0.24
x = mini juego y = botón correcto	senoidal	0.19	0.23
x = tiempo de sesión y = color presionado	cuadrática	0.63	0.66
x = color presionado y = dificultad	cuadrática	0.04	0.08

Denisse



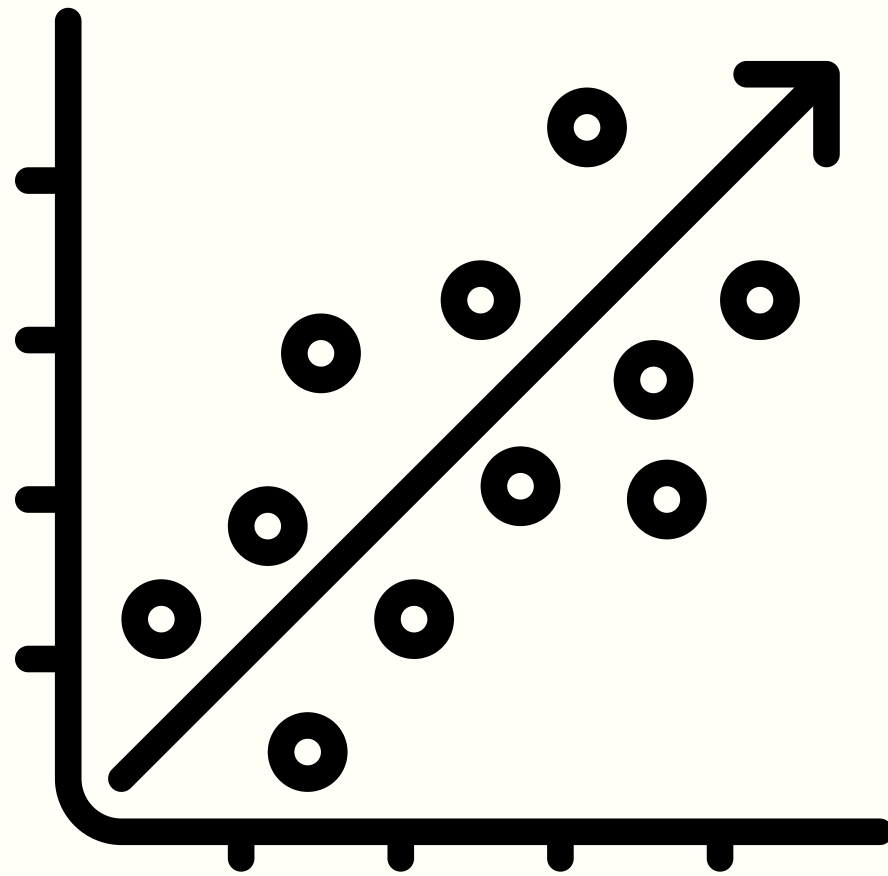
Variables	Función	Coef. lineal	Coef. no lineal
x = tiempo de lección y = tiempo de interacción	cuadrática	0.49	0.60
x = dificultad y = botón correcto	cuadrática	0.07	0.10
x = tiempo de lección y = color presionado	cuadrática	0.47	0.58
x = tiempo de sesión y = dificultad	cuadrática	0.16	0.18

Usuario	Variable	Modelo no lienal	Coef. Lineal	Coef. No Lineal	Aumento
	x número de interacción	cuadrática	0.33	0.59	79%
	y tiempo de interacción				
	x número de interacción	exponencial	0.16	0.32	100%
	y botón correcto				
	x número de interacción	cuadrática	0.49	0.54	10%
	y color presionado				
	x tiempo de sesión	cuadrática	0.3	0.34	13%
	y dificultad				
	x número de interacción	cuadrática	0.5	0.61	22%
	y tiempo de interacción				
	x color presionado	exponencial	0.11	0.23	109%
	y botón correcto				
	x tiempo de sesión	cuadrática	0.62	0.65	5%
	y color presionado				
	x color presionado	exponencial	0.29	0.36	24%
	y dificultad				
	x número de interacción	cuadrática	0.8	0.88	10%
	y tiempo de interacción				
	x dificultad	cuadrática	-0.13	0.14	114%
	y botón correcto				
	x tiempo de interacción	cuadrática	0.69	0.71	3%
	y color presionado				
	x número de interacción	cuadrática	0.17	0.23	35%
	y dificultad				
	x color presionado	cuadrática	0.14	0.24	71%
	y tiempo de interacción				
	x mini juego	senoidal	0.19	0.23	21%
	y botón correcto				
	x tiempo de sesión	xua	0.63	0.66	5%
	y color presionado				
	x color presionado	cuadrática	0.04	0.08	100%
	y dificultad				
	x tiempo de lección	cuadrática	0.49	0.6	22%
	y tiempo de interacción				
	x dificultad	cuadrática	0.07	0.1	43%
	y botón correcto				
	x tiempo de lección	cuadrática	0.47	0.58	23%
	y color presionado				
	x tiempo de sesión	cuadrática	0.16	0.18	13%
	y dificultad				

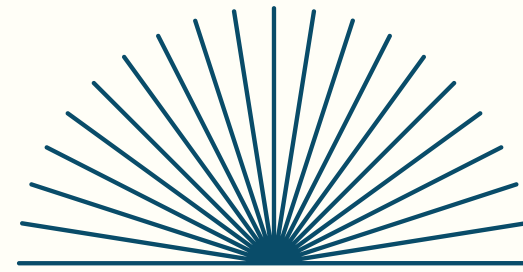
Comparativo

Variables	Aumento % (promedio)
Benjamín	50.58%
Carlos Abel	40.02%
Carlos Enrique	40.55%
Concepción	49.31%
Denisse	25.30%
General	41.15%
Modelo más utilizado	cuadrático

Conclusiones



- **Capturan relaciones complejas:** Los modelos no lineales pueden identificar interacciones y patrones que los modelos lineales no detectan, mejorando la precisión en contextos con estructuras de datos más complejas.
- **Mayor flexibilidad, menor interpretación:** Aunque suelen tener mejor desempeño, los modelos no lineales pueden ser más difíciles de interpretar, lo que representa una diferencia entre precisión y transparencia.



Gracias