

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



**Tecnológico  
de Monterrey**

**Analítica de datos y herramientas de inteligencia artificial II**

**Actividad 2.2 (Regresión Lineal Simple y Múltiple)**

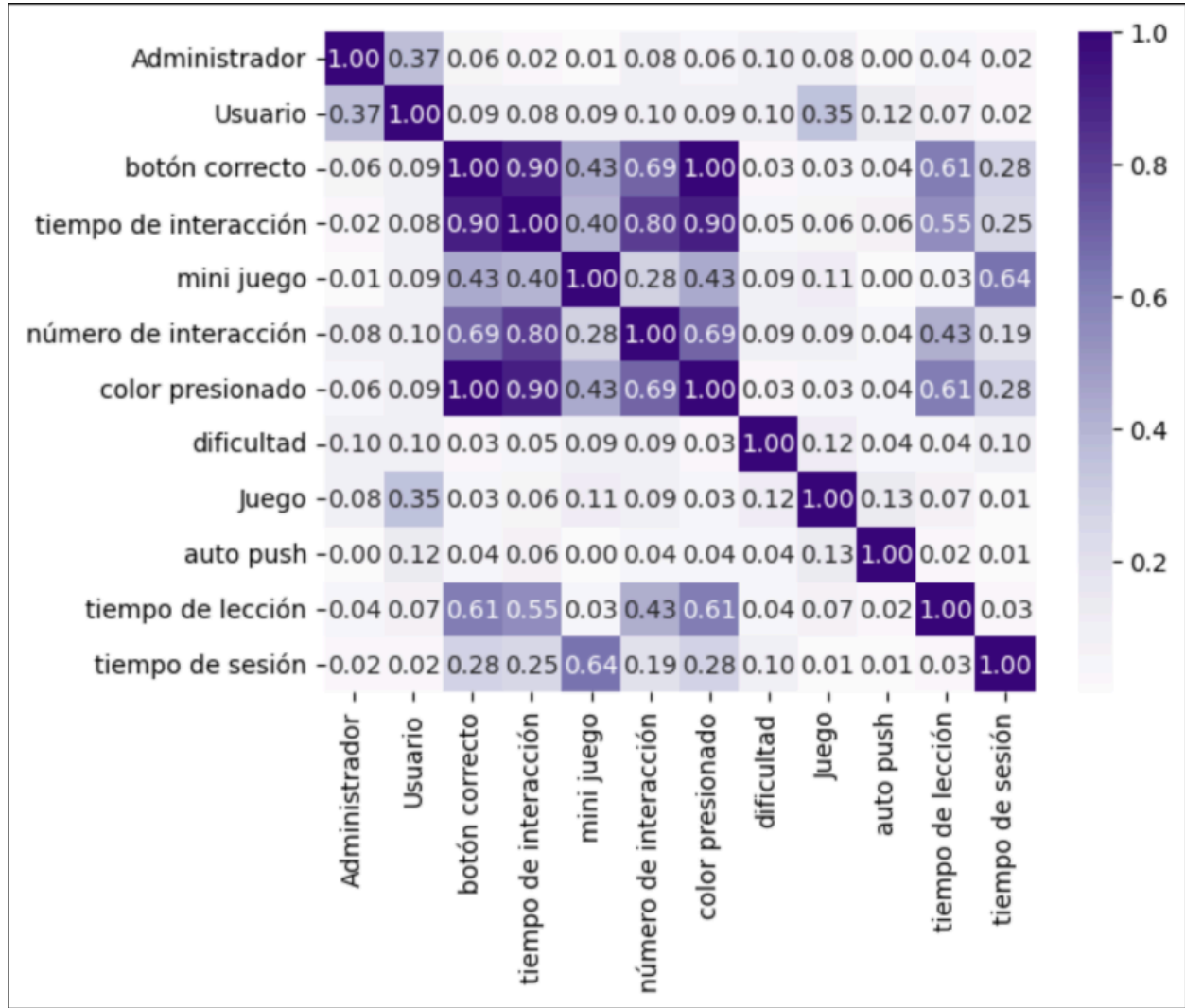
Diego Matias Rossano | A01737509

**(Gpo 501)**

6 abr 2025

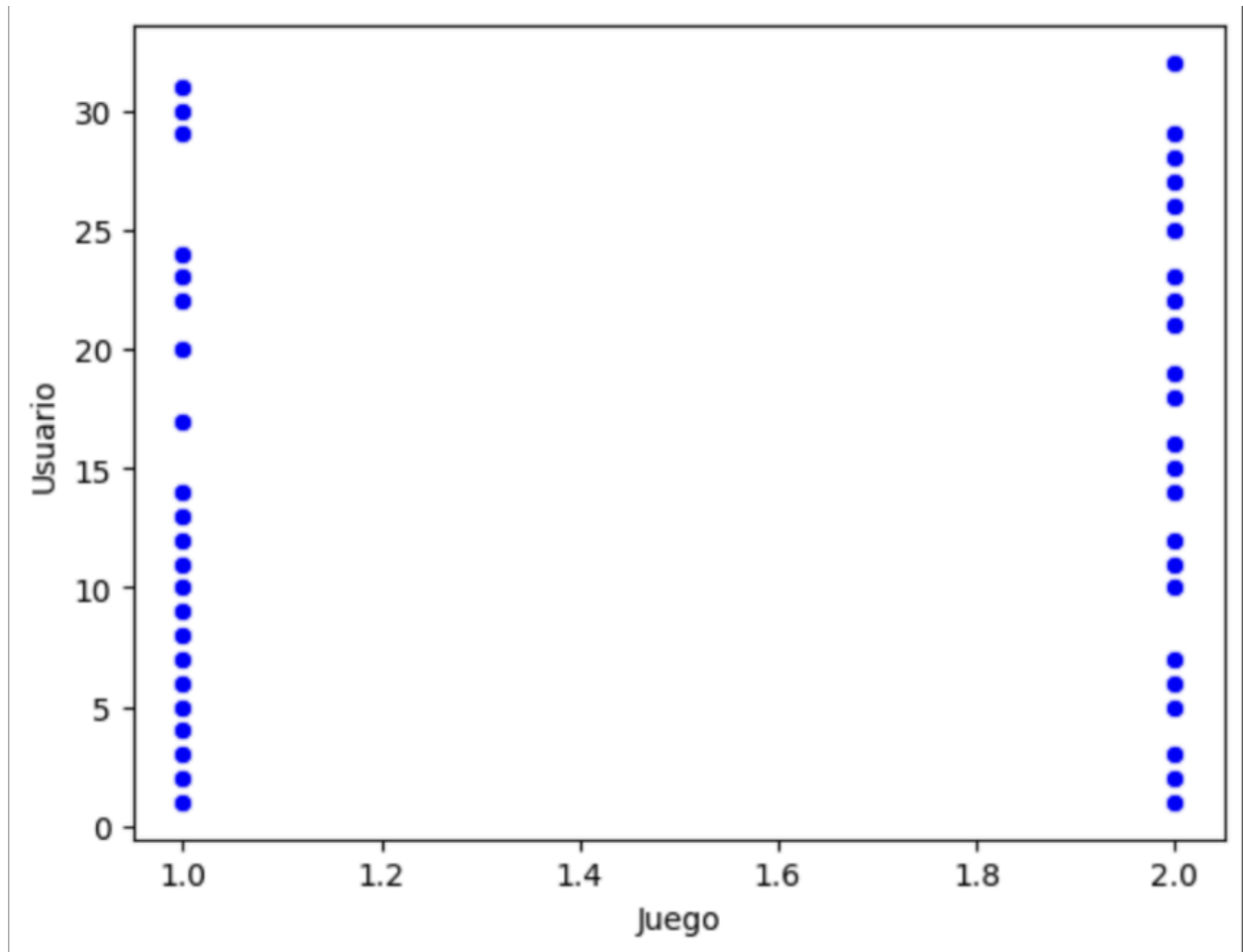
## Reporte de resultados: regresión lineal simple y múltiple

### Heatmap con la variable “usuario” como objetivo



En el heatmap podemos observar que la correlación entre nuestra variable objetivo, “usuario”, y la variable “Juego” es la más alta.

### Regresión Simple



### Interpretación:

**Coefficiente de determinación de 0.12:** Se considera una correlación débil a moderada positiva. Es un indicador de bajo poder predictivo o explicativo.

**Coefficiente de correlación de 0.35:** Un  $R^2$  de 0.12 significa que solo el 12% de la variabilidad en los usuarios se explica por la cantidad de juegos.

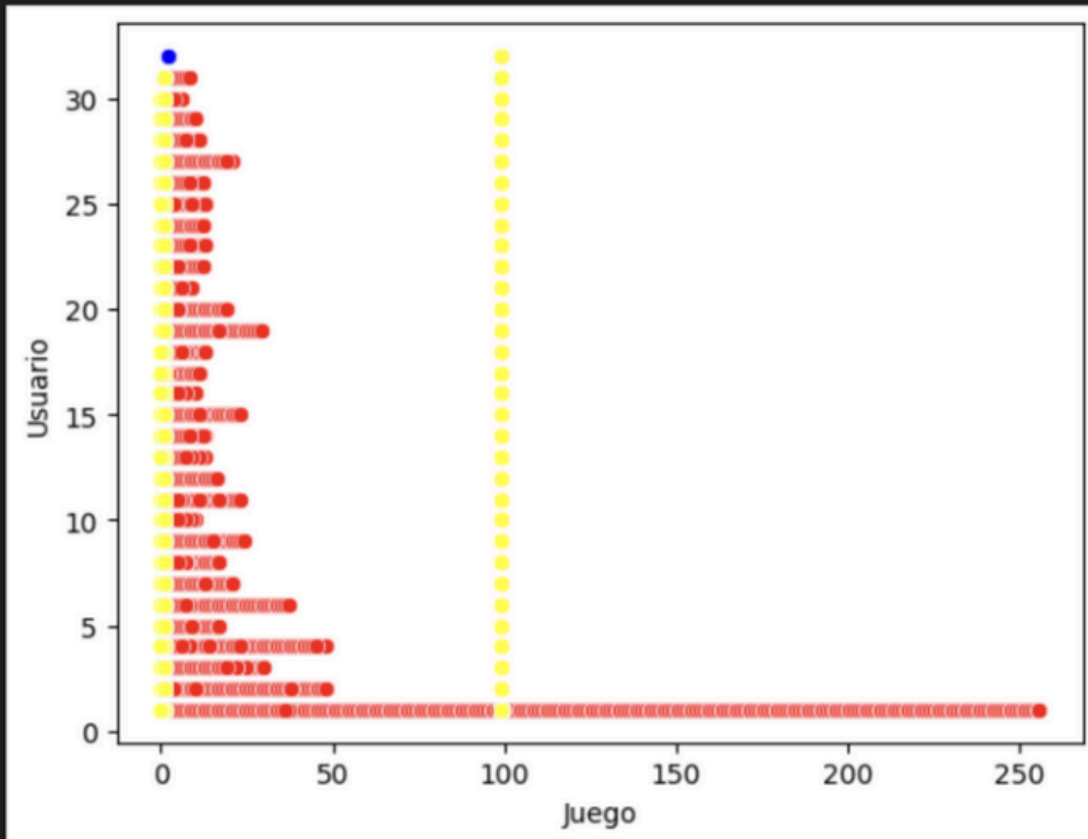
El 88% restante se debe a otros factores no incluidos en el modelo.

### Regresión Múltiple

Seleccionamos las variables con mayor correlación según el heatmap Para aumentar la correlacion y al determinación

```
sns.scatterplot(x='Juego', y='Usuario', color="blue", data=df)
sns.scatterplot(x='número de interacción', y='Usuario', color="red", data=df)
sns.scatterplot(x='botón correcto', y='Usuario', color="yellow", data=df)
```

<Axes: xlabel='Juego', ylabel='Usuario'>



### Interpretación:

**Coefficiente de determinación de 0.15:** Un  $R^2$  de 0.1589 significa que el 15.89% del comportamiento de los usuarios se puede explicar por el juego, interacción y botón correcto.

El 84.11% restante se debe a otros factores que el modelo no está capturando.

**Coefficiente de correlación de 0.39:** Un  $r$  de 0.39 indica una correlación débil a moderada positiva.

Es decir, cuando aumentan los valores de juego, interacción o botón correcto, también tiende a aumentar el número de usuarios, aunque la relación no es muy fuerte.

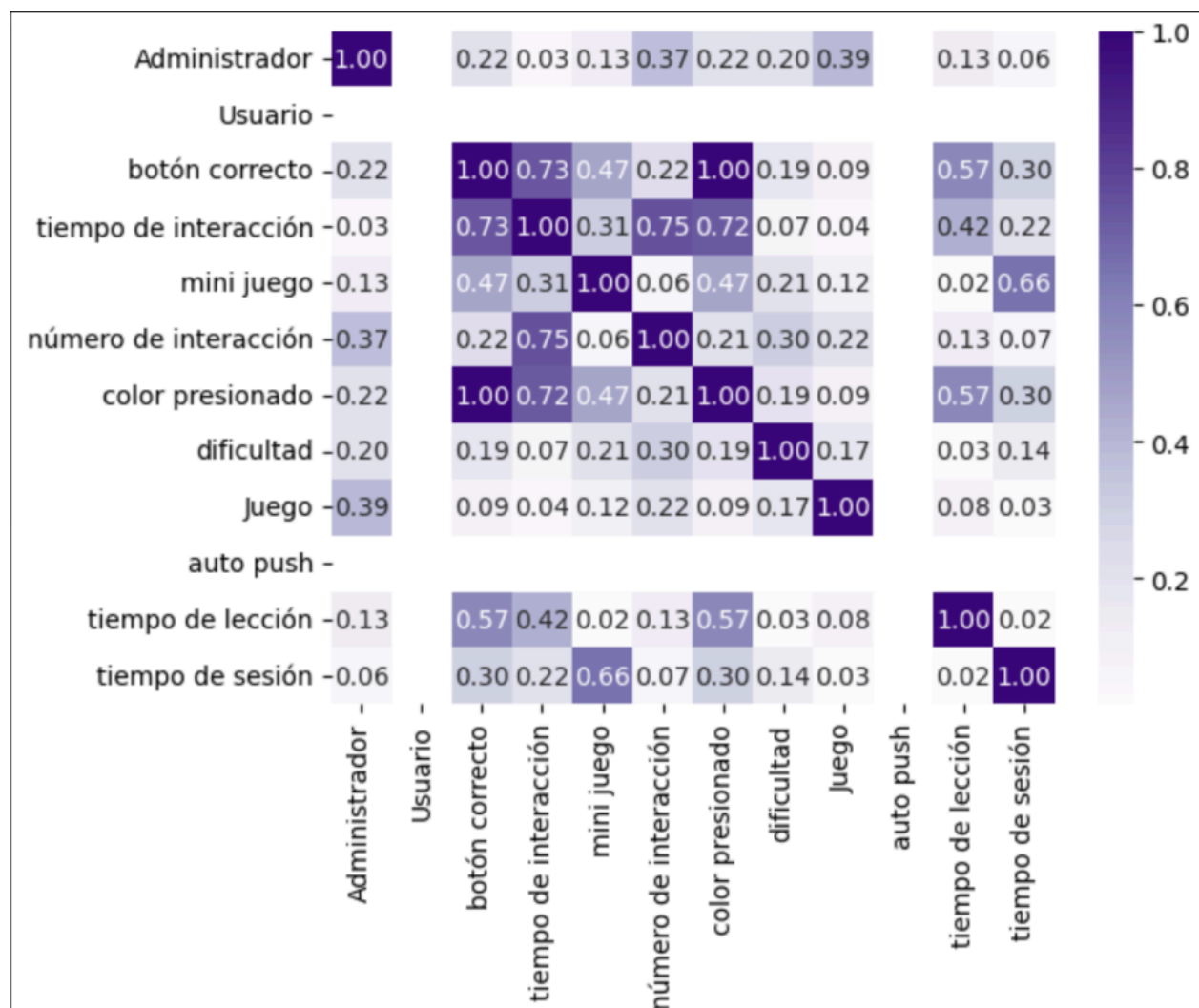
## Conclusiones generales

- Hay una mejora ligera respecto al modelo anterior (que solo incluía juego), pero la relación sigue siendo débil.
- Aunque las nuevas variables (interacción y botón correcto) ayudan un poco, la mayoría de la variación en los usuarios depende de otras cosas que aún no se están considerando.

En el siguiente paso se filtraron a 5 pacientes específicos para encontrar información que pudiese ser relevante.

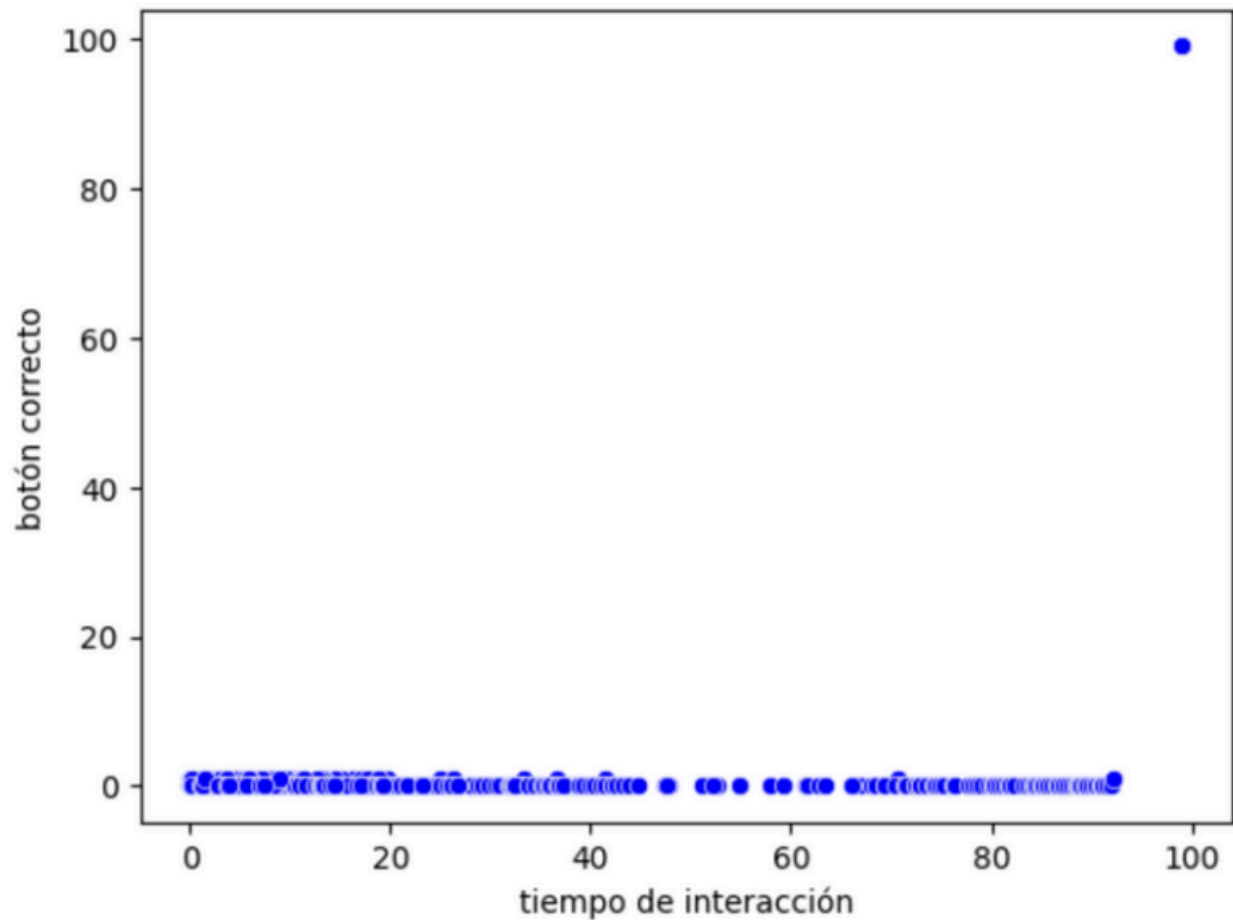
## Leonardo

### Heat Map



La variable tiempo de interacción tuvo una correlación alta en todos los pacientes con número de interacción. Por lo que se creó un modelo de regresión simple usando esas variables.

## Regresión Simple



### Interpretación

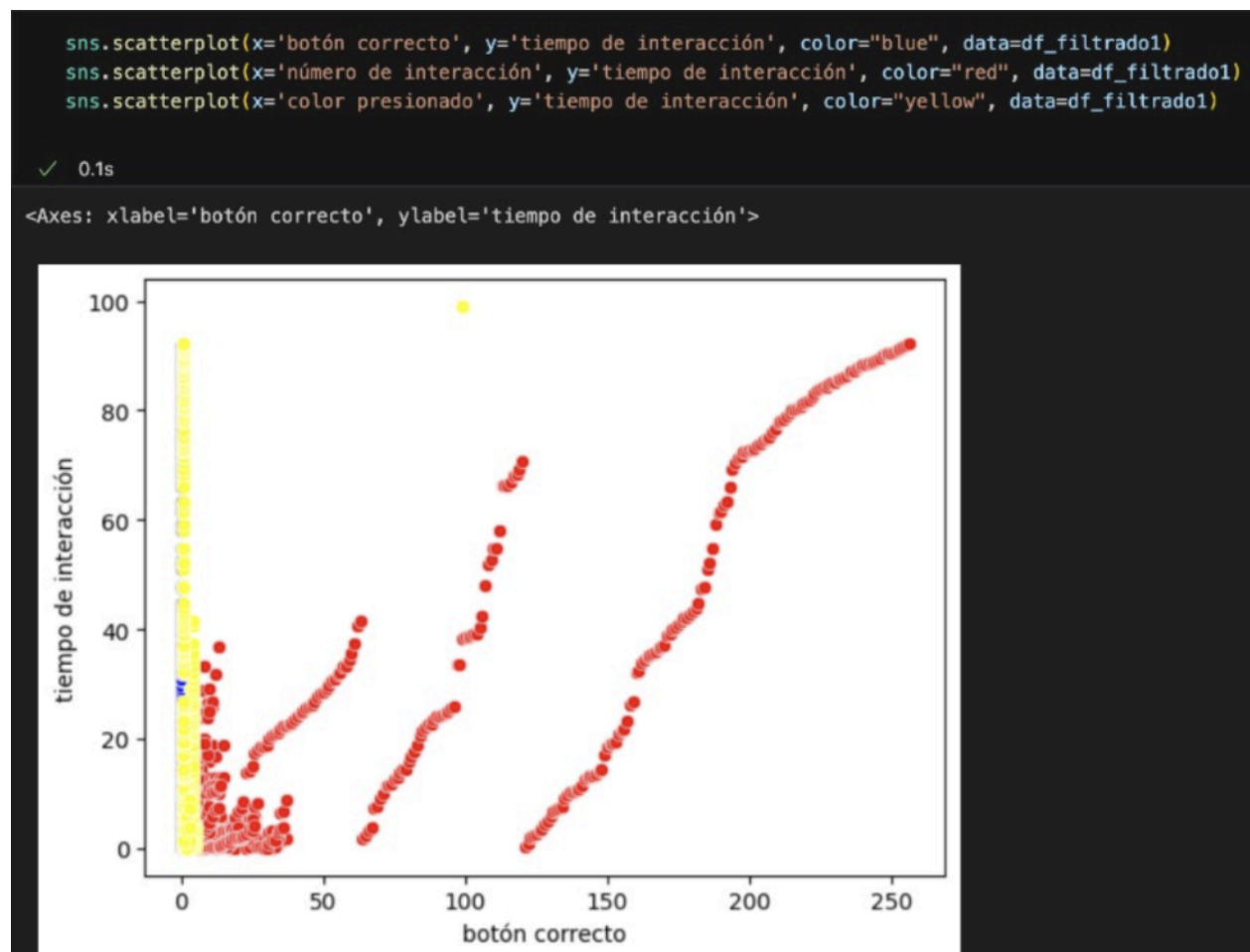
**Coefficiente de determinación:** 0.52 Es un valor moderado a alto, lo cual sugiere que el color influye bastante en cómo reaccionan los usuarios.

El otro 48% depende de otros factores, como atención, fatiga, edad, contexto visual, entre otros.

**Coefficiente de correlación:** 0.72 Eso significa que el color que se presiona sí tiene un impacto importante en el tiempo que tarda una persona en reaccionar.

A medida que cambia el color, también cambia el tiempo de reacción de manera bastante consistente.

## Regresión Múltiple



### Interpretación

**Coefficiente de determinación:** 0.89 indica una relación muy fuerte y positiva entre las variables independientes (en conjunto) y el tiempo de interacción.

**Coefficiente de Correlación:** 0.94 quiere decir que hay una relación muy clara y fuerte entre las variables (botón correcto, número de interacción y color presionado) y el tiempo de interacción.

### Hallazgos Generales

La variable tiempo de interacción tuvo una correlación alta en todos los pacientes con color presionado. Por lo que se creó un modelo de regresión simple usando esas variables.

## **Conclusión**

Después de realizar el análisis y generar las gráficas correspondientes utilizando distintas variables como objetivo, se observaron ciertas inconsistencias en los resultados. Estas incongruencias sugieren que el modelo actual no tiene la suficiente información para identificar patrones confiables o relaciones importantes entre las variables. Una de las principales limitaciones detectadas fue la cantidad reducida de datos disponibles, lo cual perjudica directamente el modelo en cuestión, la precisión de los cálculos y la calidad de las visualizaciones. Un conjunto de datos pequeño genera resultados sesgados.

Por lo tanto, concluimos que es importante tener una base de datos más amplia y equilibrada para mejorar el modelo, obtener análisis más sólidos y generar visualizaciones que reflejen el comportamiento real de los usuarios estudiados. Con más datos, no solo se mejoraría la capacidad predictiva del modelo, sino que también se podrían identificar con mayor claridad los patrones de interacción o la respuesta de los participantes analizados.