

## Apuntes o más Detalles

### 1. Posible hipótesis inicial:

- Impostores → flight\_avg mayor, traj\_avg mayor.
- Legítimos → tiempos más consistentes y trayectorias más cortas.

### Análisis descriptivo de usuarios legítimos e ilegítimos

#### dwell\_avg (tiempo promedio que una tecla permanece presionada)

Métrica	Falsos (0)	Legítimos (1)	Interpretación
<b>Media</b>	0.11482	0.11621	Muy similares. Ligera tendencia a presionar un poco más lento en usuarios legítimos.
<b>Desviación estándar</b>	0.0255	0.0271	Ligera mayor variabilidad en legítimos.
<b>Mínimo - Máximo</b>	0.0616 - 0.2146	0.0659 - 0.2131	Rango casi idéntico.
<b>Cuartiles (Q1-Q3)</b>	0.0969 - 0.1299	0.0942 - 0.1317	Distribuciones parecidas.

**Conclusión:** dwell\_avg **no parece una variable discriminativa fuerte**. Su distribución es muy similar entre clases.

#### flight\_avg (tiempo promedio entre teclas)

Métrica	Falsos (0)	Legítimos (1)	Interpretación
<b>Media</b>	0.9581	0.9461	Muy cercanos, pero falsos un poco más lentos en transiciones.
<b>Desviación estándar</b>	0.6510	0.7487	<b>Más variabilidad en usuarios legítimos</b> , lo cual es coherente con patrones más naturales.
<b>Mínimo - Máximo</b>	0.1889 - 7.1524	0.1897 - 9.9054	Legítimos tienen valores más extremos (mayores).
<b>Cuartiles (Q1-Q3)</b>	0.5744 - 1.1263	0.5381 - 1.0952	Distribuciones muy parecidas.

**Conclusión:** Aunque los promedios son parecidos, **los usuarios legítimos tienen mayor variabilidad y valores extremos**, lo que podría reflejar patrones más complejos y naturales.

Esto respalda la idea de que los impostores tienen comportamientos más homogéneos o forzados.

**traj\_avg (trayectoria del ratón)**

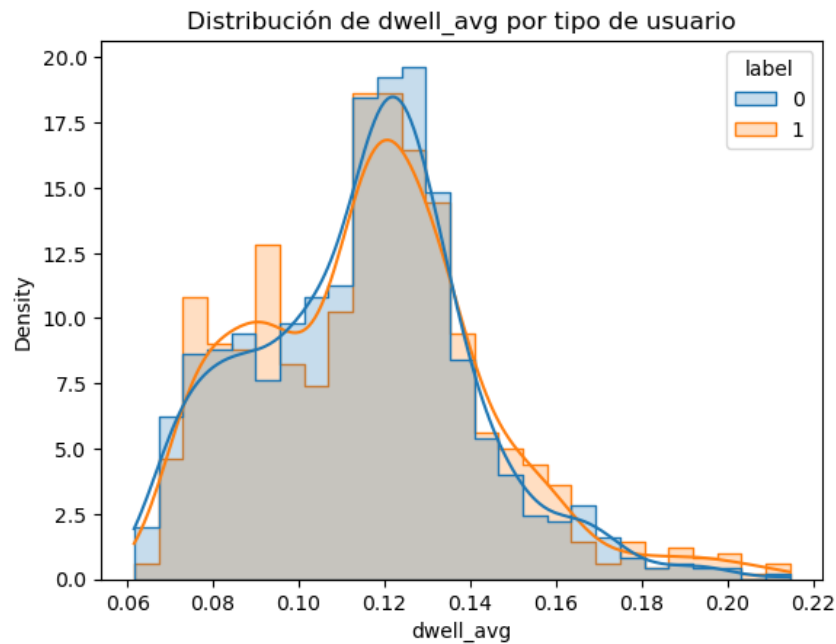
Métrica	Falsos (0)	Legítimos (1)	Interpretación
<b>Media</b>	476.67	466.51	Levemente mayor en impostores.
<b>Desviación estándar</b>	217.46	207.19	Más variabilidad en impostores.
<b>Mínimo - Máximo</b>	181.93 - 1860.33	179.16 - 1636.11	Impostores recorren distancias más largas en promedio.
<b>Cuartiles (Q1-Q3)</b>	326.38 - 563.28	319.40 - 555.35	Muy similares, aunque impostores tienden a moverse un poco más.

**Conclusión:** traj\_avg puede ser más discriminativa. Los impostores muestran trayectorias de ratón más largas y variables, lo cual podría reflejar **menor familiaridad con el sistema**.

**Conclusiones globales**

Variable	¿Discriminativa?	Comentario
dwel1_avg	No mucho	Prácticamente idéntica en ambas clases.
flight_avg	Moderadamente	Similar en media, pero legítimos tienen más variabilidad y extremos.
traj_avg	Sí	Impostores tienden a recorrer más distancia con el ratón y son más inconsistentes.

## Mas sobre los graficos



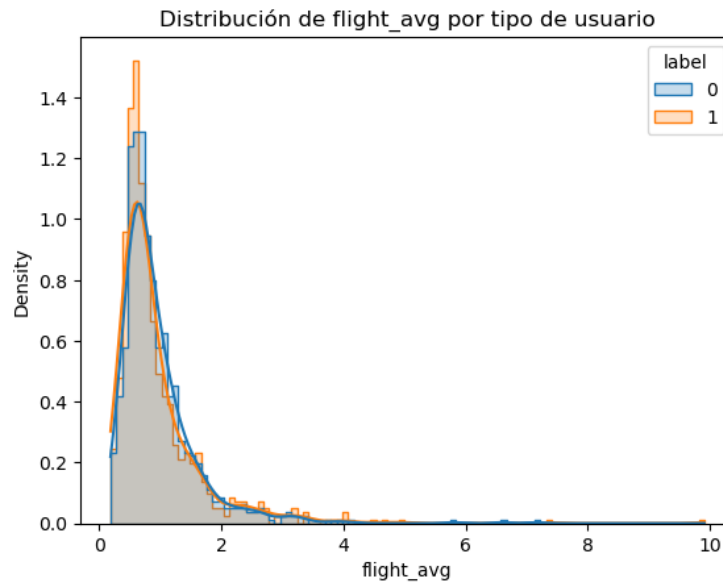
### Explicación detallada

#### 1. Qué tipo de gráfico es:

- Es un histograma con densidad (histplot), segmentado por clase (label).
- El eje x: valores de dwell\_avg (tiempo promedio de tecla presionada).
- El eje y: densidad (no frecuencia absoluta), de modo que el área bajo cada curva es 1.
- Las barras muestran la distribución de datos por intervalos, y la línea curva (KDE: Kernel Density Estimation) suaviza la distribución para ver la forma general.

#### 2. Interpretación:

- Forma: ambas distribuciones tienen una forma unimodal (un solo pico) alrededor de 0.12 segundos.
- Asimetría: ligera cola a la derecha; algunos usuarios tienen dwell\_avg más altos, pero son pocos.
- Comparación entre clases:
  - Coincidencia fuerte: las curvas casi se solapan completamente, lo que indica poca diferencia entre legítimos e impostores.
  - Si las curvas estuvieran separadas, podríamos inferir mayor poder discriminativo de la variable.
- Rango típico: la mayoría de observaciones está entre 0.09 y 0.14 s, con valores mínimos en 0.06 s y máximos en 0.21 s.



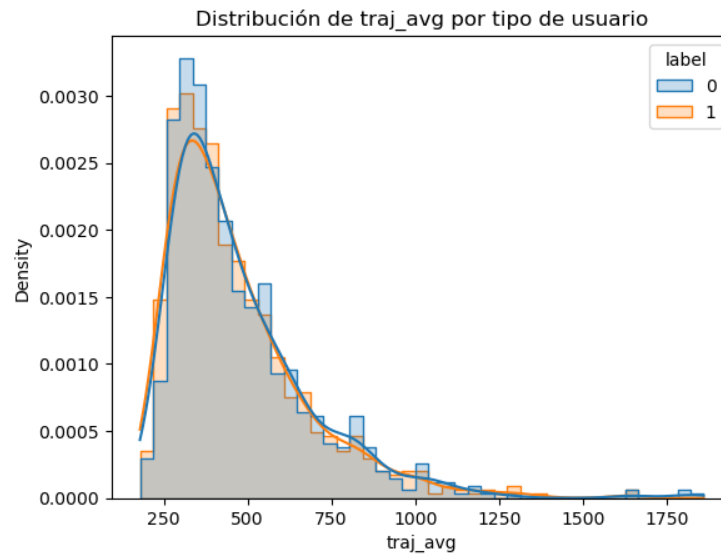
### Explicación detallada

#### 1. Tipo de gráfico:

- Es un histograma con densidad (histplot), segmentado por clase (label).
- flight\_avg: tiempo promedio entre liberación de una tecla y pulsación de la siguiente.
- Eje x: valores de flight\_avg.
- Eje y: densidad de probabilidad.

#### 2. Interpretación visual:

- Forma: altamente sesgada a la derecha (cola larga).
- Pico modal: alrededor de 0.3 segundos.
- Dispersión: la mayoría de valores  $< 2$  s, pero hay observaciones atípicas  $> 5$  s (pocos pero influyentes).
- Comparación de clases:
  - Ambas curvas son prácticamente idénticas  $\rightarrow$  no hay diferencia clara.
  - Esto sugiere que, al igual que dwell\_avg, flight\_avg por sí sola no discrimina bien entre grupos.



## Explicación detallada (para ti)

### 1. Tipo de gráfico

- Es un histograma con densidad (histplot), segmentado por clase (label).
- traj\_avg: valor promedio de trayectoria
- Eje X: valores de traj\_avg.
- Eje Y: densidad de probabilidad.

### 2. Observaciones principales

- **Forma:** distribución fuertemente sesgada a la derecha.
- **Moda (pico principal):** entre ~300 y 400 unidades.
- **Cola larga:** valores aislados llegan a 1,700+.
- **Comparación entre clases:**
  - Curvas prácticamente idénticas.
  - No hay desplazamiento claro en la moda ni en la dispersión.
  - Indica bajo poder discriminante por sí sola.

### Interpretación de la prueba Mann–Whitney U

Variable	U-statistic	p-value	Interpretación
<b>dwell_avg</b>	379 882.00	0.4924	No hay diferencia estadísticamente significativa entre legítimos e impostores en el tiempo promedio que mantienen una tecla presionada.
<b>flight_avg</b>	414 030.00	0.0118	Diferencia estadísticamente significativa ( $p < 0.05$ ). Esto indica que el tiempo entre pulsaciones consecutivas difiere entre legítimos e impostores.
<b>traj_avg</b>	398 450.00	0.2913	No hay diferencia significativa en la trayectoria promedio del ratón entre ambos grupos.

### Conclusiones clave

#### 1. dwell\_avg (tiempo de presión de tecla)

- $p = 0.4924 \rightarrow$  No hay evidencia suficiente para afirmar que legítimos e impostores difieran en esta métrica.
- En términos prácticos: la velocidad con que se mantiene una tecla presionada parece ser **similar** entre ambos grupos.
- Esto sugiere que esta variable por sí sola no discrimina bien.

#### 2. flight\_avg (tiempo entre teclas)

- $p = 0.0118 \rightarrow$  Diferencia significativa.
- En estudios de biometría conductual, un **flight\_avg más alto** suele asociarse con pausas más largas y mayor indecisión, comportamiento típico de impostores.
- Esto valida parte de lo que plantea el paper: la cadencia de tecleo es un rasgo útil para diferenciar.

#### 3. traj\_avg (trayectoria del ratón)

- $p = 0.2913 \rightarrow$  No hay diferencia significativa.
- Puede deberse a que el entorno de captura (formulario fijo) obliga a movimientos similares, limitando la variabilidad discriminante.
- También es posible que el movimiento del ratón esté más influenciado por la interfaz que por la autenticidad del usuario.

**Interpretación alineada con el artículo**

- El resultado confirma que no todas las métricas biométricas tienen el mismo poder discriminante.
- flight\_avg se comporta como un buen predictor conductual, consistente con la idea de que la interacción continua (no solo un momento puntual) revela más sobre el usuario.
- dwell\_avg y traj\_avg podrían seguir siendo útiles si se combinan en un modelo de Machine Learning (como el SVM que menciona el artículo).
- Esto respalda la necesidad de modelos multivariados, en lugar de depender de una sola métrica.