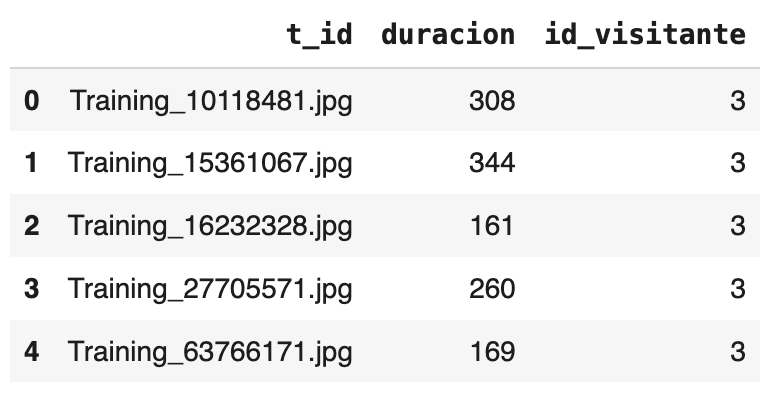
# 🧭 Exploración de id\_visitante\_duracion.json

### ⌗ Sobre los datos

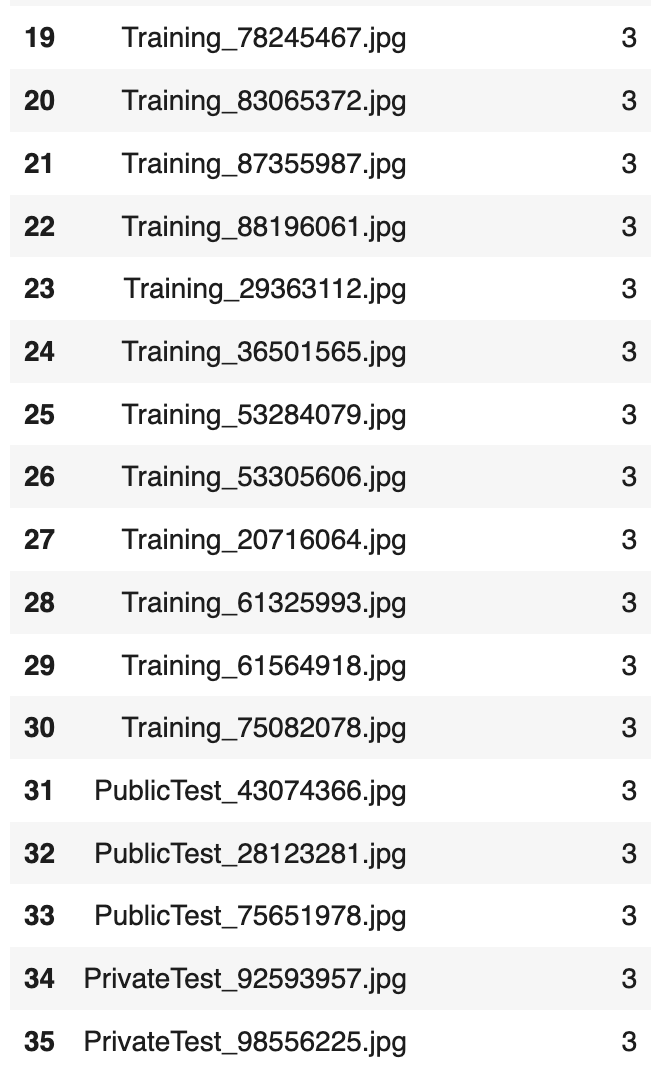
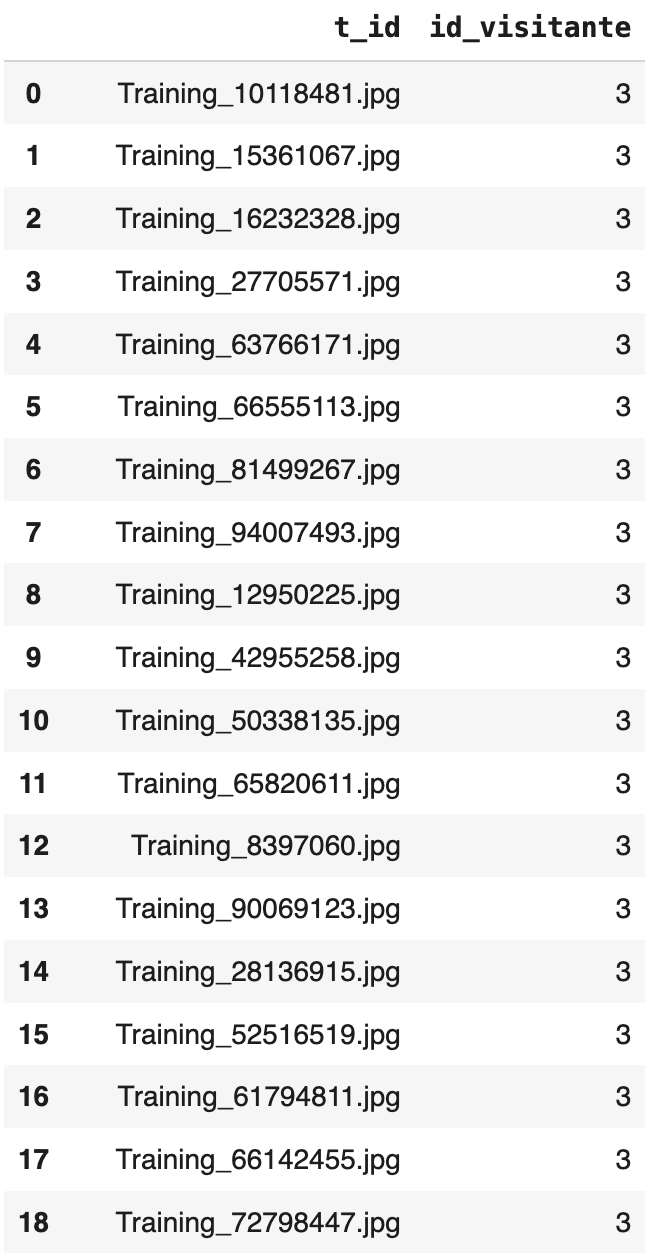
Haciendo un head vemos que la duración está asociada a una imagen y un id\_visitante.



### ❌ Problema

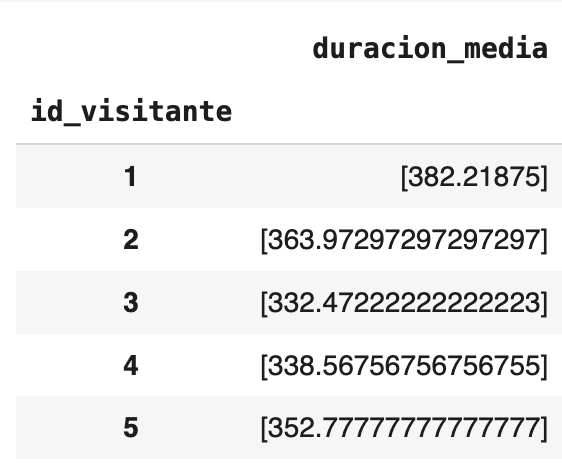
* ¿Por qué hay una duración asociada a cada imagen?
* ¿Por qué los id\_visitante repetidos no tienen la misma duración siempre?

Podemos entender que cada visitante tenga varias fotos, pero no varias duraciones, ya que estas no coinciden. Y no se pueden asociar a un día concreto porque hay visitantes que tendrían más de 30 días. Vamos a ver el ejemplo del visitante 3:



### ✅Posibles soluciones

* Hacer una media de duración por visitante:



* ¿Se puede tener en cuenta que ‘duración’ sea otra cosa y esté mal categorizado?
* ¿Puede ser el id\_visitante erróneo?

### ❌ Problema

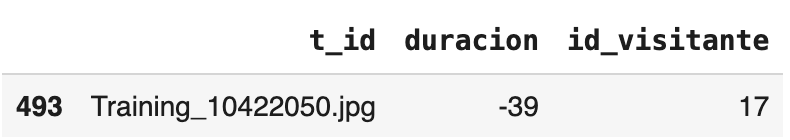
Hay 1279 datos que superan las 9 horas, que supone el 3.56% del df\_duracion

### ✅Posibles soluciones

* Borrar esas filas.
* Dejar aquellos valores que duran menos de 10h y considerarlos como errores del sistema.

### ❌ Problema

Hay un valor negativo:

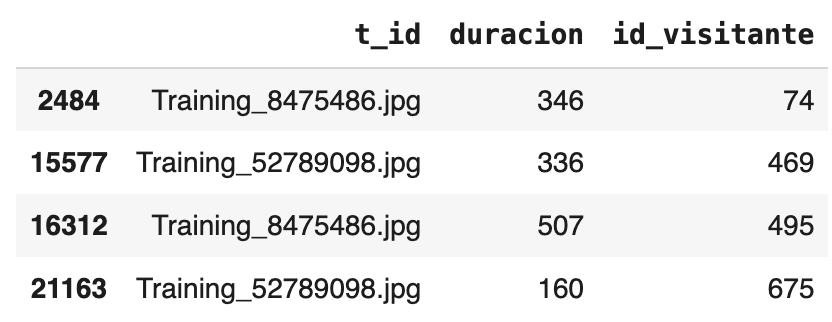


### ✅Posibles soluciones

* Borrar esa fila.
* Sustituir ese valor con la media de duración de ese id\_visitante.

### ❌ Problema

Hay dos duplicados, que son los mismos que en el resto de json:



### ✅Posibles soluciones

* Borrar esas filas.
* Como son dos identificadores diferentes, podemos añadirles el id\_visitante detrás del nombre para que sean únicos.