

# 11\_Transform

November 28, 2023



## 0.1 Groupby: Transform

Vamos a dedicar esta sesión al metodo transform como cierre de las sesiones dedicadas a las agrupaciones hechas con groupby. Pero como tónica habitual de esta unidad, carga datos en nuestro DataFrame guía:

```
[ ]: import numpy as np
import pandas as pd

df_aviones = pd.read_csv("../data/dataset_inicial_aviones.csv", index_col = "Id_vuelo")
```

### 0.1.1 Transform fuera del *groupby*

Aunque lo vamos a ver en el contexto de las agrupaciones hechas con "groupby", **transform** es un método que también se aplica a **Series** y **DataFrame** sin que haya un **groupby** de por medio.

Es similar a **apply**, y sus principales diferencias son: \* Puede aplicar una o varias funciones (sí varias funciones a la vez, introducidas como lista o diccionario) \* Sólo se puede aplicar a una serie (o columna) a la vez (no vale el método de varias columnas)

Supongamos que queremos pasar los Origenes y los Destinos a todo mayúsculas y además generar una abreviatura con las tres primeras letras, en vez de hacerlo en dos pasadas podemos:

```
[ ]: 
```

```
[ ]: 
```

```
[ ]:
```

```
[ ]:
```

```
[ ]:
```

### 0.1.2 Transform para agrupaciones

La diferencia más importante respecto a `apply` o `agg` es que `transform` devuelve un serie con tantos elementos como la serie que se le pasa como input (recuerda que `transform` solo puede manejar un "columna" o serie a la vez). En concreto para los elementos de una misma agrupación devuelve el valor calculado por la función, veámoslo con un ejemplo.

```
[ ]: ### Sin transform
```

```
[ ]: ### Con transform
```

Fíjate en el número de elementos y los índices de la salida de la línea de código con `transform`

¿Y esto para que sirve? Típicamente para pegar a cada elemento individual su valor de agrupación y poder utilizarlo después

Por ejemplo:

```
[ ]:
```

```
[ ]:
```

Y ahora podemos hacer cálculos directos que de otra forma serían más complejos de conseguir. Vamos a crearnos otra columna que recoja para cada vuelo el porcentaje sobre la media de su duración lo que nos permitirá luego por ejemplo hacer control de vuelos que se nos desvíen mucho o poco de la media.

Tal como lo tenemos es fácil hacer esa columna

```
[ ]:
```

```
[ ]:
```

Ahora piensa tú en cómo harías para calcular la columna "desviacion\_duracion" sin usar el `transform` previo... te lo dejo como ejercicio a ver en la sesión en vivo.