

10_Funciones_agrupacion

November 28, 2023



0.1 Groupby: Funciones Agregación Avanzadas

Hasta ahora hemos visto funciones de agregación sencillas (la función que se pone al final del groupby para que "haga algo" con las columnas indicadas), en esta sesión vamos a aumentar esa capacidad. ¿Cómo? Pues aplicando las funciones por columna, es decir no siempre la misma función a cada columna, y además pudiendo hacer que sea una función de usuario (al estilo apply, del que tanto te acuerdas)

Y, ¿cómo no?, comencemos creando nuestro DataFrame a partir de los datos de vuelos:

```
[ ]: 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 Agg: Funciones de usuario

Supongamos que ahora queremos obtener por Compañía y tipo de Avión, el número de vuelos y además el consumo medio y el destino más visitado. Hala ahí es nada. Descompongámoslo en tres pequeños problemas y luego veremos como juntarlos con la ayuda del método `agg`

1. Por compañía y tipo de avión, el número de vuelos. Ese es fácil, lo hicimos en la píldora anterior:

```
[ ]:
```

2. Por compañía y tipo de avión, el consumo medio. También lo tengo:

[]:

Para terminar:

3. Por compañía y tipo de avion el destino más visitado. Hmm, está... ah, ya

Necesito calcular un estadístico la "moda", del que hablamos en algún momento y que viene a decirnos de una serie de valores (numeros o cadenas) aquel que más se repite. Las series tienen el metodo mode:

[]:

Pero desgraciadamente las series que devuelve un groupby....:

[]:

Vale, he hecho trampa para enseñarte primero como aplicar una función definida por el usuario. Aquí lo que tienes que tener en cuenta es que cómo parámetro o argumento la función va a recibir una serie por agrupación...

Si, como hemos venido haciendo, agrupamos por Compañía y vuelo y aplicamos una función (como ahora mismo veremos) esa función va a recibir una serie con todos los valores para "Airnar", "Airbur A380" de una vez, y luego otra serie con todos los valores para "Airna", "Airbus 320" y luego se llamará con todos los valores de "Airnar", "Boeing 737" y así tantas llamadas como combinaciones de valores de "Aircompany" y "avion" (porque son las columnas que hemos elegidos) y en cada llamada todos los valores. Y qué valores pues los de las columnas que hayamos escogido después del groupby. Vale, que necesitas un ejemplo... Primero la función:

[]:

Ahora la forma de aplicarlo, sí eso... con apply:

[]:

Vale ya tenemos las tres cosas que queríamos.... por separado, ¿Cómo lo hago a la vez? Con agg o aggregate

0.1.2 Agg, aggregate: aplicando funciones y operaciones diferentes por columnas

Muy sencillo, vamos a usar el método agg o el aggregate al que se le pasa un diccionario:

```
{nombre_de_la_columna: operacion_sobre_esa_columna,  
nombre_de_la_columna2: operacion_sobre_esa_columna,...  
nombre_de_la_columnan: operacion_sobre_esa_columna}
```

Venga, en formato diccionario:

[]:

Para más info [aquí](#) (donde puedes ver también no más formatos para pasar las funciones), y ahora:

[]:

Por supuesto, podemos usar `as_index`:

[]: