

GROUP BY

Sintaxis:

Para una operación

```
variable = df.groupby(columna)[columna_operacion].operacion
```

Para varias operaciones

```
variable = df.groupby(columna)[columna_operacion].agg(['op1', 'op2'])
```

Operaciones de agregación:

.count(): **número** valores **no nulos**

.describe(): **resumen** de los principales estadísticos

.sum(): **suma** de todos los valores

.mean(): **media** de los valores

.median(): **mediana**. Ordenados de menor a mayor, valor que queda en la mitad

.min(): valor **mínimo**

.max(): valor **máximo**

.std(): **desviación estándar**. Cuánto se desvían los valores de la media.

.var(): **varianza**. Cuánto se desvían los valores de la media al cuadrado.

Métodos:

.reset_index(): nuevo **DataFrame del resultado** con índice en 0

`df.groupby(columna)[columna_operacion].operacion(numeric_only=True)`: aplica la operación a todas columnas numéricas

variable_agrupacion.ngroups: grupos formados después de la agrupación

APPLY

Para aplicar una función a una columna. Se puede añadir a una nueva columna de o aplica a una columna ya existente para modificar el DataFrame

Sintaxis:

```
df['nueva_columna'] = df['columna'].apply(función)
```

```
df['columna_modificar'] = df['columna_modificar'].apply(función)
```

Para aplicar una función que debe recibir dos parámetros (dos columnas del DF) se hace con lambda.

Sintaxis:

```
df['nueva_columna'] = df.apply(lambda x: función(x['col1'], x['col2']), axis=1)
```

OTROS MÉTODOS DE LIMPIEZA

Para aplicar una **transformación** o reemplazo **de los valores** a un Serie o DataFrame

.map(): transformación de cada elemento de una Serie

Pasos:

1. Crear diccionario de mapeo, keys valores actuales, values valores por los que reemplazar

```
diccionario = {0: "No", 1: "Si"}
```

2. Aplicar diccionario a una columna:

```
df['col'] = df['col'].map(diccionario)
```

.replace(): reemplaza valores en un DataFrame o Serie por otros especificados

Sintaxis:

```
df['col'] = df['col'].replace(valor a reemplazar, nuevo valor)
```

Se utiliza para reemplazar un valor concreto de esa columna, no todos