

✓ Ejercicio completo de Data Frames

Carga en Python el CSV llamado run.csv que encontrarás en el repositorio del curso.

```
1 import pandas as pd
2 import numpy as np
3 import duckdb
4
5 url = 'run.csv'
6 data = pd.read_csv(url, sep=',')
7 data.drop(['Unnamed: 0'], axis=1, inplace=True)
8 data
```



	pulso.antes	pulso.despues	hace.deporte	fuma	genero	altura	peso	tipo.actividad
0	48	54	no	si	H	68.0	150	nula
1	54	56	no	si	H	69.0	145	moderada
2	54	50	no	no	H	69.0	160	moderada
3	58	70	si	no	H	72.0	145	moderada
4	58	58	no	no	H	66.0	135	intensa
...
87	92	94	no	si	H	69.0	150	moderada
88	94	92	no	si	M	62.0	131	moderada
89	96	140	si	no	M	61.0	140	moderada
90	96	116	si	no	M	68.0	116	moderada
91	100	115	si	si	M	63.0	121	moderada

92 rows × 8 columns

✓ Pregunta 1

Indica cuántos estudiantes formaron parte del estudio de deporte.

```
1 data.shape[0]

92
```

✓ Pregunta 2

Indica cuántos individuos son hombres y cuántos son mujeres.

```
1 data['genero'].value_counts()

genero
H      57
M      35
Name: count, dtype: int64
```

✓ Pregunta 3

Calcula el porcentaje medio de variación del pulso por minuto entre antes y después de hacer ejercicio y compara el valor de los que hacen ejercicio habitualmente y los que no. ¿Observas mucha diferencia?

```
1 data['pulso_diff'] = (data['pulso.despues'] - data['pulso.antes']) / data['pulso.antes']
2 data.groupby(['hace.deporte'], as_index=False, dropna=False).agg(mean_pulso=('pulso_diff', 'mean'))
```

	hace.deporte	mean_pulso
0	no	0.002095
1	si	0.262154

La variación entre los que si hacen con respecto a los que no hacen es realmente alta la diferencia

▼ Pregunta 4

Calcula el porcentaje medio de variación del pulso por minuto entre antes y después de hacer ejercicio para los estudiantes que hacen ejercicio habitualmente y compara el valor de los hombres con el de las mujeres. ¿Observas mucha diferencia?

```
1 data1 = data[data['hace.deporte'] == 'si'].copy()
2 data1.groupby(['genero'], as_index=False, dropna=False).agg(mean_pulso=('pulso_diff', 'mean'))
```

	genero	mean_pulso
0	H	0.191769
1	M	0.415722

Como podemos observar, entre las personas que si hacen deporte la diferencia de hombres y mujeres es relativamente más alta en comparación con sus homologos que no hacen deporte.

▼ Pregunta 5

Calcula el porcentaje medio de variación del pulso por minuto entre antes y después de hacer ejercicio para los estudiantes que no hacen ejercicio habitualmente y compara el valor de los fumadores con los no fumadores. ¿Observas mucha diferencia?

```
1 data2 = data[data['hace.deporte'] == 'no'].copy()
2 data2 = data2.groupby(['fuma'], as_index=False, dropna=False).agg(mean_pulso=('pulso_diff', 'mean'))
3 data2
```

	fuma	mean_pulso
0	no	-0.000648
1	si	0.009122

```
1 np.diff(data2['mean_pulso'])
array([0.00976979])
```

▼ Pregunta 6

Calcula el porcentaje medio de variación del pulso por minuto entre antes y después de hacer ejercicio de todos los estudiantes según el tipo de actividad física que realizan. ¿Observas alguna diferencia?

```
1 data3 = data.copy()
2 data3 = data3.groupby(['tipo.actividad'], as_index=False, dropna=False).agg(mean_pulso=('pulso_diff', 'mean'))
3 data3
```

	tipo.actividad	mean_pulso
0	intensa	0.013475
1	moderada	0.140157
2	nula	0.125000
3	suave	0.037472

