

Data Frames en Python. Tarea 02

Alberto Simón

23/04/2025

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

```
import pandas as pd
import numpy as np
df = pd.read_csv("../data/run.csv", delimiter = ",", decimal = ".", encoding="utf-8")
df.head()
```

```
##      Unnamed: 0  pulso.antes  pulso.despues  ...  altura peso tipo.actividad
## 0              1           48           54  ...   68.0  150             nula
## 1              2           54           56  ...   69.0  145          moderada
## 2              3           54           50  ...   69.0  160          moderada
## 3              4           58           70  ...   72.0  145          moderada
## 4              5           58           58  ...   66.0  135            intensa
##
## [5 rows x 9 columns]
```

Pregunta 1

Indica cuántos estudiantes formaron parte del estudio de deporte.

```
students_n = df.shape[0]
students_n
```

```
## 92
```

Pregunta 2

Indica cuántos individuos son hombres y cuántos son mujeres.

```
gender_counts = df["genero"].value_counts()
gender_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?

```
def pulso_relativo(x):
    return (x['pulso.despues']-x['pulso.antes'])/x['pulso.antes']

df["variacion_pulso"] = 100 * (df["pulso.despues"] - df["pulso.antes"]) / df["pulso.antes"]
```

```
average = df["variacion_pulso"].mean()
average_sport = df[df["hace.deporte"] == "si"]["variacion_pulso"].mean()
average_no_sport = df[df["hace.deporte"] == "no"]["variacion_pulso"].mean()
average, average_sport, average_no_sport
```

```
## (np.float64(10.1030387884714), np.float64(26.215407303688497), np.float64(0.2094791738644114))
```

Bastante diferencia entre los que hacen deporte y los que no.

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?

```
deporte_df = df[df["hace.deporte"] == "si"]

deporte_df.groupby('genero').aggregate({
    'variacion_pulso' : 'mean'
})
```

```
##          variacion_pulso
## genero
## H          19.176886
## M          41.572181
```

Hay diferencia

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?

```
no_deporte_df = df[df["hace.deporte"] == "no"]

no_deporte_df.groupby('fuma').aggregate({
    'variacion_pulso' : 'mean'
})
```

```
##          variacion_pulso
## fuma
## no          -0.064761
## si           0.912219
```

Diferencia muy acusada. ## 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?

```
no_deporte_df.groupby('tipo.actividad').aggregate({
    'variacion_pulso' : 'mean'
})
```

```
##          variacion_pulso
## tipo.actividad
## intensa          -1.412238
## moderada           0.600729
## nula             12.500000
```

suave -0.402434

Es la categoría donde menos variación existe.