Curso de Especialización de
Inteligencia Artificial y Big Data (IABD)

Sistemas de Big Data

UD04. Manejo de Dataframes con
Python/Pandas.
Tarea Online.

JUAN ANTONIO GARCÍA MUELAS

INDICE

Tarea para SBD04

Título de la tarea: Manejo de DataFrames con Python/Pandas

Curso de Especialización: Inteligencia Artificial y Big Data

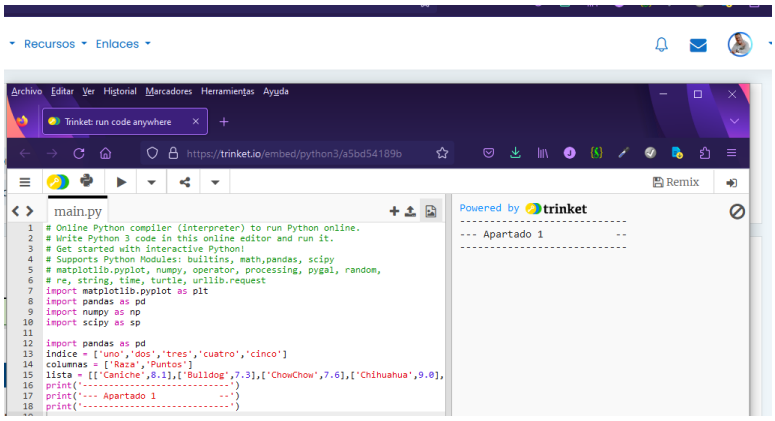
Módulo profesional: Sistemas de Big Data

¿Qué te pedimos que hagas?

A continuación, tienes la primera parte del programa de Enrique.

Accede al intérprete de Python en trinket.io, pega el código fuente que te entregamos y ejecútalo (con el botón del triángulo negro de la parte superior).

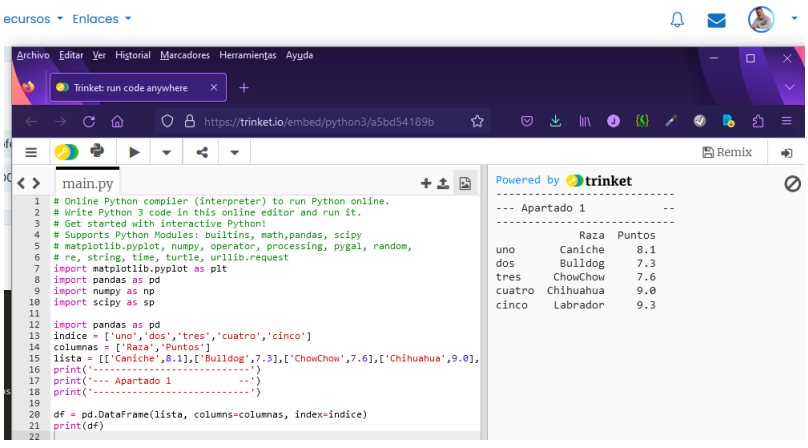
```
import pandas as pd

indice = ['uno','dos','tres','cuatro','cinco']
columnas = ['Raza','Puntos']
lista =
[['Caniche',8.1],['Bulldog',7.3],['ChowChow',7.6],['Chihuahua',9.0],['Labrado
r',9.3]]

print('----------------------------')
print('--- Apartado 1            --')
print('----------------------------')
```



✓ **Apartado 1:**

Crea un DataFrame llamado **df** a partir de la lista de razas y puntuaciones otorgadas por Enrique, usando el índice, las columnas que se te proporcionan.

```
df = pd.DataFrame(lista, columns=columnas, index=indice)
print(df)
```



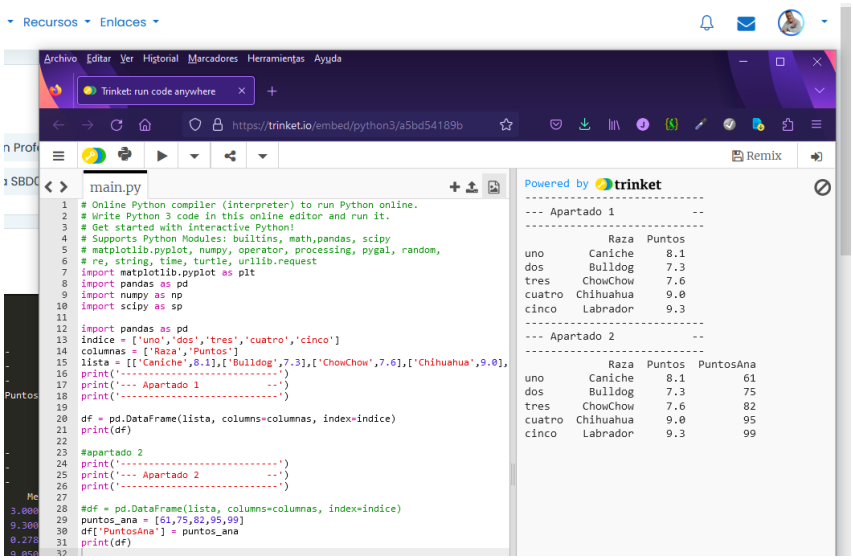
✓ **Apartado 2:**

Ana le entrega sus puntuaciones a Enrique y éste las transcribe en forma de la siguiente lista:

`puntos_ana = [61,75,82,95,99]`

Añade la columna con etiqueta `'PuntosAna'` al DataFrame.

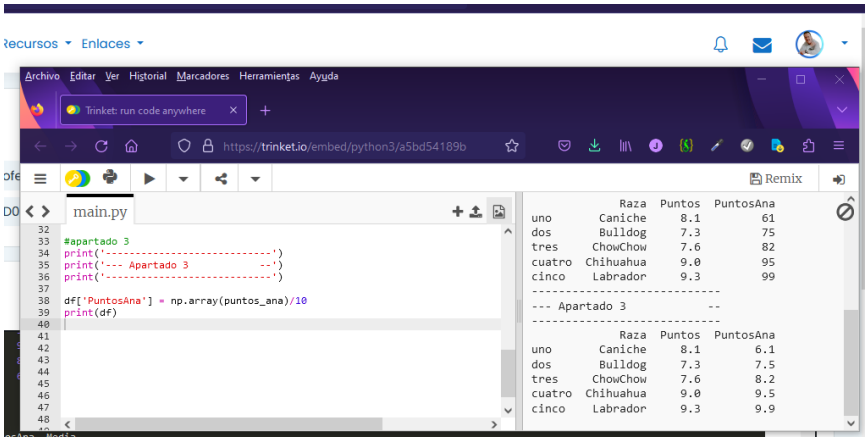
```
puntos_ana = [61,75,82,95,99]
df['PuntosAna'] = puntos_ana
print(df)
```



✓ **Apartado 3:**

Como las puntuaciones de Ana han entrado en valores de 0 a 100 y las de Enrique estaban de 0 a 10, divide la columna `'PuntosAna'` entre 10 de modo que todas las puntuaciones queden en la misma escala.

```
df['PuntosAna'] = np.array(puntos_ana)/10
print(df)
```



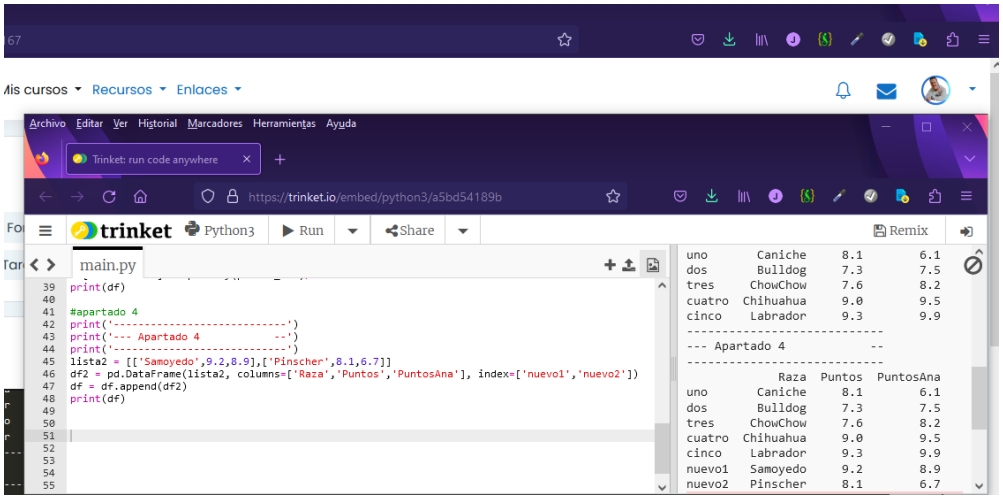
✓ **Apartado 4:**

Enrique y Ana han estado viendo una revista de animales y han decidido añadir otras dos razas más, por lo que han creado la siguiente lista ya con sus puntuaciones integradas.

`lista2 = [['Samoyedo',9.2,8.9],['Pinscher',8.1,6.7]]`

Añade esas dos nuevas filas al DataFrame. Fíjate en que se han quedado sin ideas para los índices y han decidido usar **'nuevo1'** y **'nuevo2'** para las nuevas filas.

```
lista2 = [['Samoyedo',9.2,8.9],['Pinscher',8.1,6.7]]
df2 = pd.DataFrame(lista2, columns=['Raza','Puntos','PuntosAna'],
index=['nuevo1','nuevo2'])
df = df.append(df2)
print(df)
```

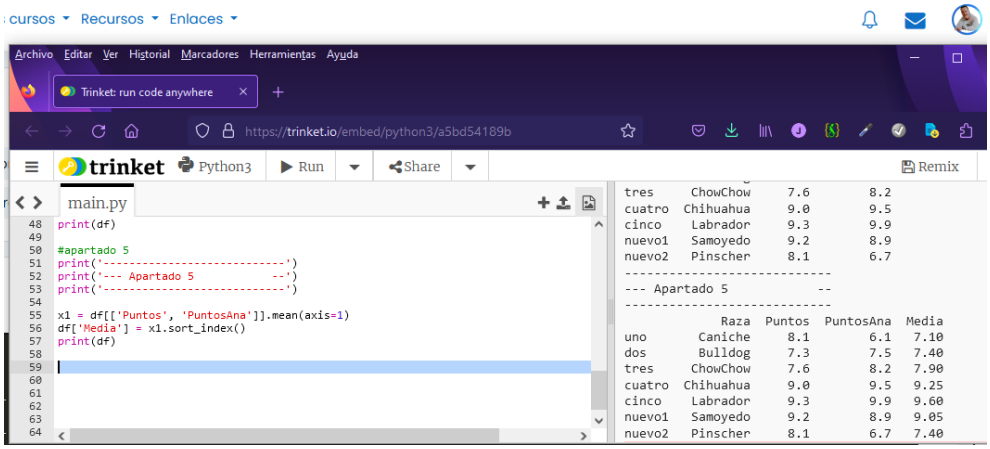


En este ejercicio, el editor avisa de un futuro paso a desuso en la función **append().**

✓ **Apartado 5:**

Crea la columna **'Media'** para poder ver la puntuación media para cada raza.

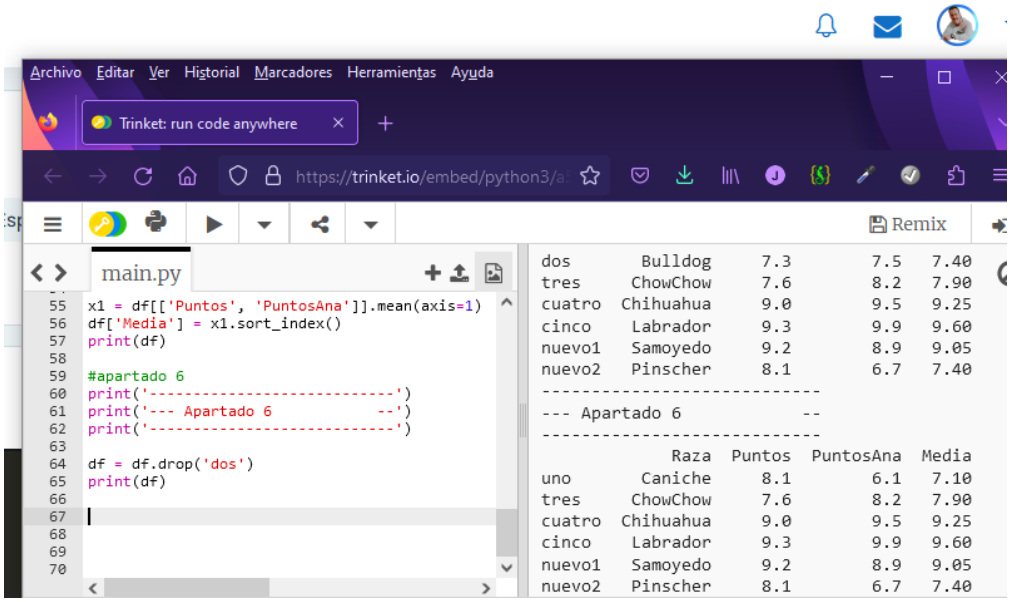
```
x1 = df[['Puntos', 'PuntosAna']].mean(axis=1)
df['Media'] = x1.sort_index()
print(df)
```



✓ **Apartado 6:**

Enrique y Ana deciden que no tendrán un bulldog. Elimina la fila cuyo índice es `'dos'`.

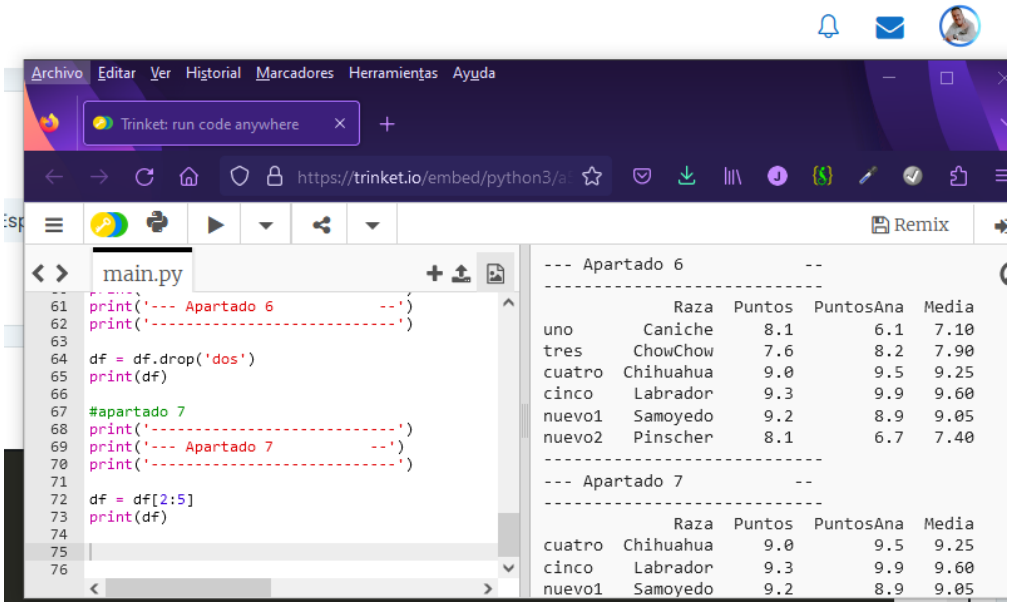
```
df = df.drop('dos')
print(df)
```



✓ **Apartado 7:**

Viendo las puntuaciones medias, Enrique y Ana deciden que la decisión va a estar entre chihuahua, labrador y samoyedo, por lo que deciden obtener sólo esas filas del DataFrame (en concreto desde la posición 2 hasta la 5 -no incluída-).

```
df = df[2:5]

print(df)
```



✓ **Apartado 8:**

Por último, sólo por curiosidad, deciden ver información estadística sobre las filas que les han quedado.

```
print(df.describe())
```