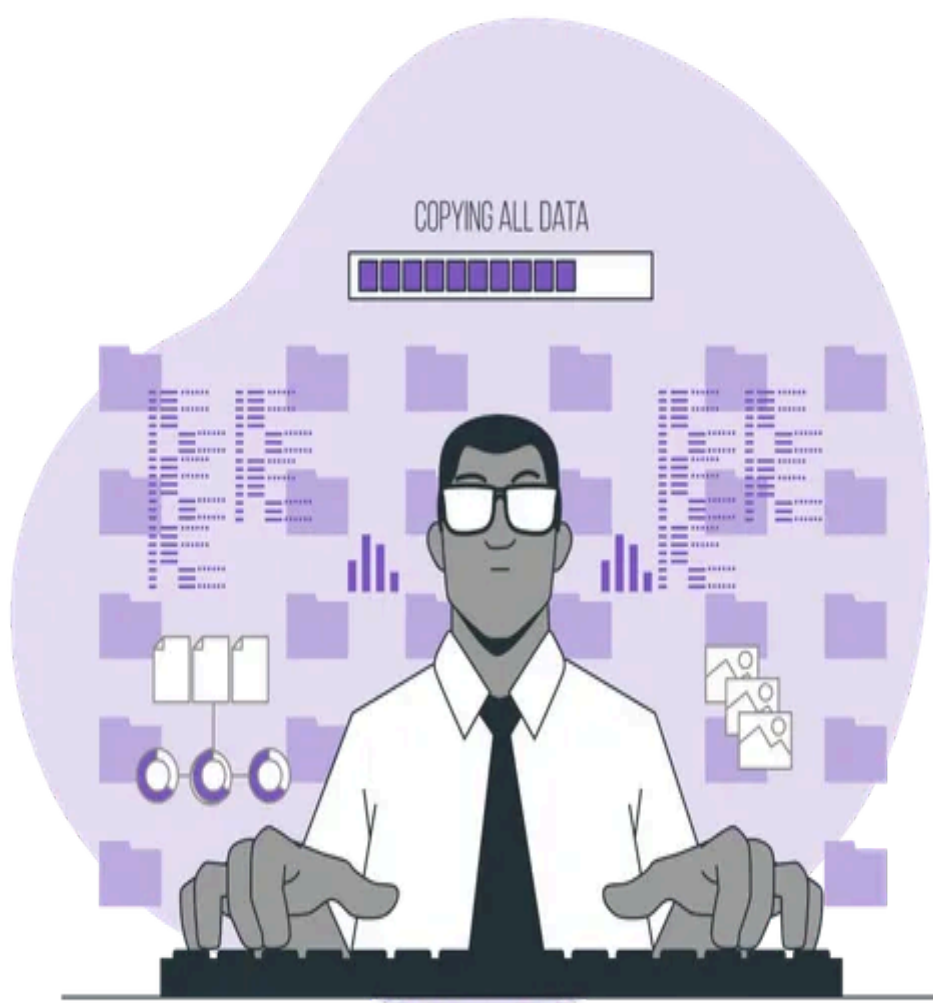


Sistemas de Big Data



Nombre: Victoria Jiménez Martín

Módulo: Sistemas de Big Data

Curso: Especialización de Inteligencia Artificial y Big Data

Índice

Apartado 1	3
Apartado 2	3
Apartado 3	4
Apartado 4	5
Apartado 5	5
Apartado 6	6
Apartado 7	6
Apartado 8	7

Apartado 1

Para este apartado, se declarará el DataFrame que se compondrá por la lista de razas con la puntuación (lista), se asignará la lista para el índice (index=índice) y se asignará las columnas (columns=columnas).

```
main.py
1 # Online Python compiler (interpreter) to run Python online.
2 # Write Python 3 code in this online editor and run it.
3 # Get started with interactive Python!
4 # Supports Python Modules: builtins, math,pandas, scipy
5 # matplotlib.pyplot, numpy, operator, processing, pygal, random,
6 # re, string, time, turtle, urllib.request
7
8 # Autora: Victoria Jiménez Martín
9
10 import pandas as pd
11
12 indice = ['uno','dos','tres','cuatro','cinco']
13 columnas = ['Raza','Puntos']
14 lista = [['Caniche',8.1],['Bulldog',7.3],['ChowChow',7.6],['Chihuahua',9.0],['Labrador',9.3]]
15
16 print('-----')
17 print('--- Apartado 1 ---')
18 print('-----')
19
20 df = pd.DataFrame(lista, index=indice, columns=columnas)
21 print(df)
22
23
24
25
26
```

	Raza	Puntos
uno	Caniche	8.1
dos	Bulldog	7.3
tres	ChowChow	7.6
cuatro	Chihuahua	9.0
cinco	Labrador	9.3

Apartado 2

En este apartado se creará la lista para los puntos de Ana, y se añadirá esta misma columna al DataFrame:



```
main.py
1 # Online Python compiler (interpreter) to run Python online.
2 # Write Python 3 code in this online editor and run it.
3 # Get started with interactive Python!
4 # Supports Python Modules: builtins, math,pandas, scipy
5 # matplotlib.pyplot, numpy, operator, processing, pygal, random,
6 # re, string, time, turtle, urllib.request
7
8 # Autora: Victoria Jiménez Martín
9
10 import pandas as pd
11
12 indice = ['uno','dos','tres','cuatro','cinco']
13 columnas = ['Raza','Puntos']
14 lista = [['Caniche',8.1],['Bulldog',7.3],['ChowChow',7.6],['Chihuahua',9.0],['Labrador',9.3]]
15
16 print('-----')
17 print('--- Apartado 1 ---')
18 print('-----')
19
20 df = pd.DataFrame(lista, index=indice, columns=columnas)
21 print(df)
22
23 print('-----')
24 print('--- Apartado 2 ---')
25 print('-----')
26 puntos_ana = [61, 75, 82, 95, 99]
27 df['PuntosAna'] = puntos_ana
28 print(df)
29
30
31
```





	Raza	Puntos	PuntosAna
uno	Caniche	8.1	61
dos	Bulldog	7.3	75
tres	ChowChow	7.6	82
cuatro	Chihuahua	9.0	95
cinco	Labrador	9.3	99

Apartado 3

En este apartado, se especificará la columna de los puntos de Ana y la dividirá entre 10:

≡







main.py

+

↕



```
1 # Online Python compiler (interpreter) to run Python online.
2 # Write Python 3 code in this online editor and run it.
3 # Get started with interactive Python!
4 # Supports Python Modules: builtins, math,pandas, scipy
5 # matplotlib.pyplot, numpy, operator, processing, pygal, random,
6 # re, string, time, turtle, urllib.request
7
8 # Autora: Victoria Jiménez Martín
9
10 import pandas as pd
11
12 indice = ['uno','dos','tres','cuatro','cinco']
13 columnas = ['Raza','Puntos']
14 lista = [['Caniche',8.1],['Bulldog',7.3],['ChowChow',7.6],['Chihuahua',9.0],
15
16 print('-----')
17 print('--- Apartado 1      --')
18 print('-----')
19
20 df = pd.DataFrame(lista, index=indice, columns=columnas)
21 print(df)
22
23 print('-----')
24 print('--- Apartado 2      --')
25 print('-----')
26 puntos_ana = [61, 75, 82, 95, 99]
27 df['PuntosAna'] = puntos_ana
28 print(df)
29
30 print('-----')
31 print('--- Apartado 3      --')
32 print('-----')
33 df['PuntosAna'] = df['PuntosAna'] / 10
34 print(df)
```

Powered by  trinket

--- Apartado 1 --

	Raza	Puntos
uno	Caniche	8.1
dos	Bulldog	7.3
tres	ChowChow	7.6
cuatro	Chihuahua	9.0
cinco	Labrador	9.3

--- Apartado 2 --

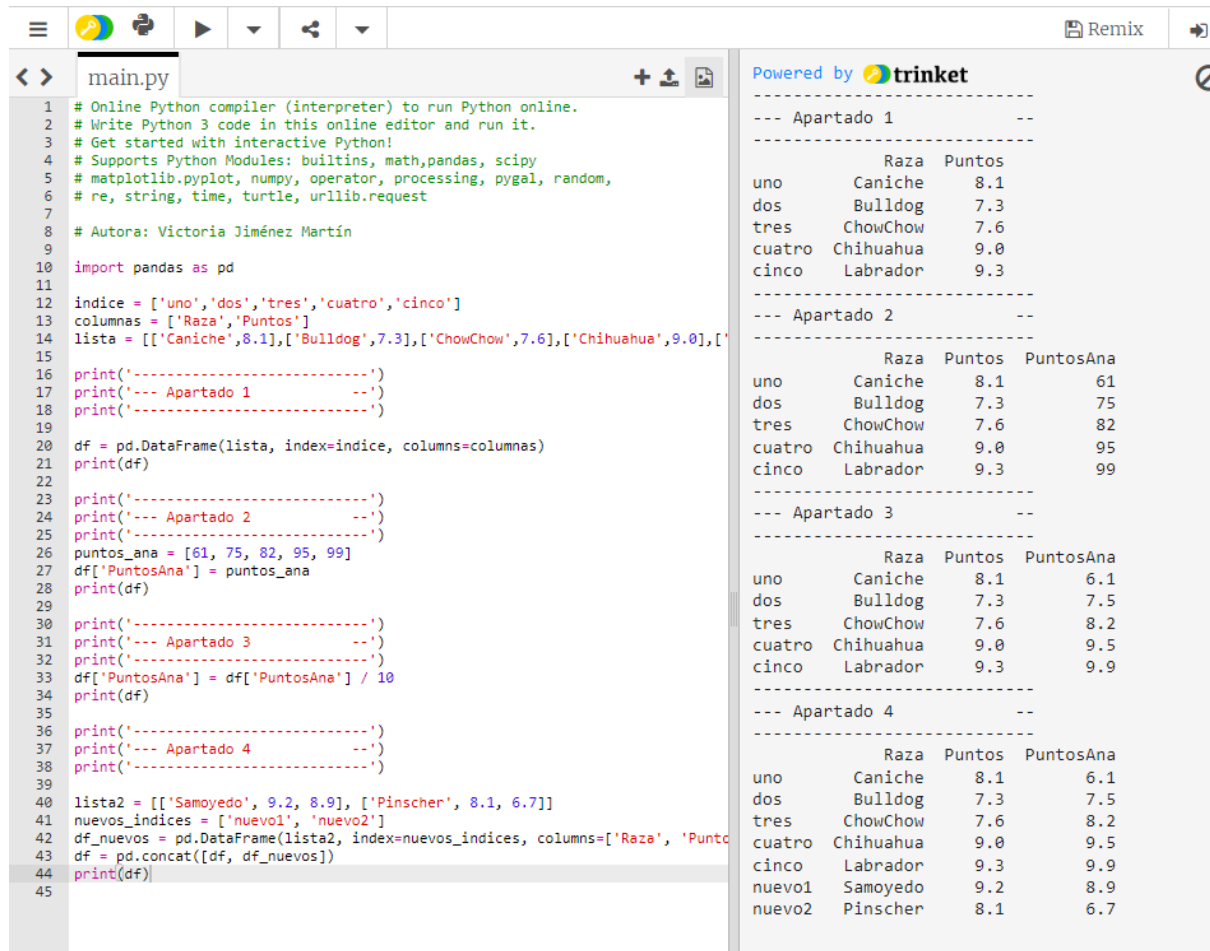
	Raza	Puntos	PuntosAna
uno	Caniche	8.1	61
dos	Bulldog	7.3	75
tres	ChowChow	7.6	82
cuatro	Chihuahua	9.0	95
cinco	Labrador	9.3	99

--- Apartado 3 --


	Raza	Puntos	PuntosAna
uno	Caniche	8.1	6.1
dos	Bulldog	7.3	7.5
tres	ChowChow	7.6	8.2
cuatro	Chihuahua	9.0	9.5
cinco	Labrador	9.3	9.9

Apartado 4

Se creará una nueva lista para los nuevos registros con la raza y la puntuación, se creará dos registros nuevos para añadir al índice. Definimos el nuevo DataFrame en el df_nuevos y con la función .concat unirá el DataFrame antiguo junto con el nuevo:



```
1 # Online Python compiler (interpreter) to run Python online.
2 # Write Python 3 code in this online editor and run it.
3 # Get started with interactive Python!
4 # Supports Python Modules: builtins, math,pandas, scipy
5 # matplotlib.pyplot, numpy, operator, processing, pygal, random,
6 # re, string, time, turtle, urllib.request
7
8 # Autora: Victoria Jiménez Martín
9
10 import pandas as pd
11
12 indice = ['uno','dos','tres','cuatro','cinco']
13 columnas = ['Raza','Puntos']
14 lista = [['Caniche',8.1],['Bulldog',7.3],['ChowChow',7.6],['Chihuahua',9.0],['Labrador',9.3]]
15
16 print('-----')
17 print('--- Apartado 1 ---')
18 print('-----')
19
20 df = pd.DataFrame(lista, index=indice, columns=columnas)
21 print(df)
22
23 print('-----')
24 print('--- Apartado 2 ---')
25 print('-----')
26 puntos_ana = [61, 75, 82, 95, 99]
27 df['PuntosAna'] = puntos_ana
28 print(df)
29
30 print('-----')
31 print('--- Apartado 3 ---')
32 print('-----')
33 df['PuntosAna'] = df['PuntosAna'] / 10
34 print(df)
35
36 print('-----')
37 print('--- Apartado 4 ---')
38 print('-----')
39
40 lista2 = [['Samoyedo', 9.2, 8.9], ['Pinscher', 8.1, 6.7]]
41 nuevos_indices = ['nuevo1', 'nuevo2']
42 df_nuevos = pd.DataFrame(lista2, index=nuevos_indices, columns=['Raza', 'Puntos', 'PuntosAna'])
43 df = pd.concat([df, df_nuevos])
44 print(df)
45
```

Powered by  trinket

--- Apartado 1 ---

	Raza	Puntos
uno	Caniche	8.1
dos	Bulldog	7.3
tres	ChowChow	7.6
cuatro	Chihuahua	9.0
cinco	Labrador	9.3

--- Apartado 2 ---

	Raza	Puntos	PuntosAna
uno	Caniche	8.1	61
dos	Bulldog	7.3	75
tres	ChowChow	7.6	82
cuatro	Chihuahua	9.0	95
cinco	Labrador	9.3	99

--- Apartado 3 ---

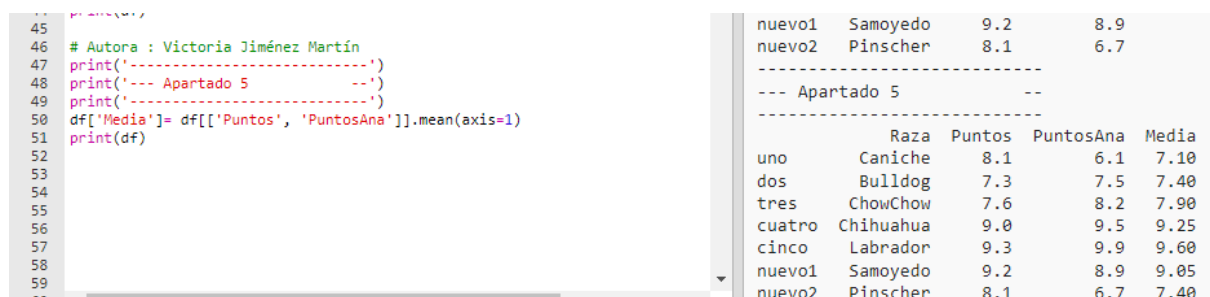
	Raza	Puntos	PuntosAna
uno	Caniche	8.1	6.1
dos	Bulldog	7.3	7.5
tres	ChowChow	7.6	8.2
cuatro	Chihuahua	9.0	9.5
cinco	Labrador	9.3	9.9

--- Apartado 4 ---

	Raza	Puntos	PuntosAna
uno	Caniche	8.1	6.1
dos	Bulldog	7.3	7.5
tres	ChowChow	7.6	8.2
cuatro	Chihuahua	9.0	9.5
cinco	Labrador	9.3	9.9
nuevo1	Samoyedo	9.2	8.9
nuevo2	Pinscher	8.1	6.7

Apartado 5

Para este apartado, se creará una nueva columna que hará la media con la función .mean:



```
45
46 # Autora : Victoria Jiménez Martín
47 print('-----')
48 print('--- Apartado 5 ---')
49 print('-----')
50 df['Media'] = df[['Puntos', 'PuntosAna']].mean(axis=1)
51 print(df)
52
53
54
55
56
57
58
59
60
```

nuevo1 Samoyedo 9.2 8.9
nuevo2 Pinscher 8.1 6.7

--- Apartado 5 ---

	Raza	Puntos	PuntosAna	Media
uno	Caniche	8.1	6.1	7.10
dos	Bulldog	7.3	7.5	7.40
tres	ChowChow	7.6	8.2	7.90
cuatro	Chihuahua	9.0	9.5	9.25
cinco	Labrador	9.3	9.9	9.60
nuevo1	Samoyedo	9.2	8.9	9.05
nuevo2	Pinscher	8.1	6.7	7.40

Apartado 6

Para este apartado, se usará la función `.drop` se especificará mediante un string el registro del índice a eliminar:

45		nuevo1	Samoyedo	9.2	8.9	
46	# Autora : Victoria Jiménez Martín	nuevo2	Pinscher	8.1	6.7	
47	print('-----')					
48	print('--- Apartado 5 ---')					
49	print('-----')					
50	df['Media'] = df[['Puntos', 'PuntosAna']].mean(axis=1)					
51	print(df)					
52						
53	print('-----')					
54	print('--- Apartado 6 ---')					
55	print('-----')					
56	df = df.drop('dos')					
57	print(df)					
58						
59						
60						
61						
62						
63						
64						
65						
66						
67						
68						
69						
70						
71						
72						

Apartado 7

Para este apartado , se usará la función `.iloc` y el rango de las líneas de los registros a seleccionar (excluye los números en los que empiezan y terminan):

45		tres	ChowChow	7.6	8.2	7.90
46	# Autora : Victoria Jiménez Martín	cuatro	Chihuahua	9.0	9.5	9.25
47	print('-----')	cinco	Labrador	9.3	9.9	9.60
48	print('--- Apartado 5 ---')	nuevo1	Samoyedo	9.2	8.9	9.05
49	print('-----')	nuevo2	Pinscher	8.1	6.7	7.40
50	df['Media'] = df[['Puntos', 'PuntosAna']].mean(axis=1)					
51	print(df)					
52						
53	print('-----')					
54	print('--- Apartado 6 ---')					
55	print('-----')					
56	df = df.drop('dos')					
57	print(df)					
58						
59	print('-----')					
60	print('--- Apartado 7 ---')					
61	print('-----')					
62	df_seleccionado = df.iloc[2:5]					
63	print(df_seleccionado)					
64						
65						
66						
67						
68						
69						
70						
71						
72						

Apartado 8

Para este apartado, utilizaremos la función `.describe` para que muestre las estadísticas:

```
45 # Autora : Victoria Jiménez Martín
46 print('-----')
47 print('--- Apartado 5 ---')
48 print('-----')
49 df['Media'] = df[['Puntos', 'PuntosAna']].mean(axis=1)
50 print(df)
51
52 print('-----')
53 print('--- Apartado 6 ---')
54 print('-----')
55 df = df.drop('dos')
56 print(df)
57
58 print('-----')
59 print('--- Apartado 7 ---')
60 print('-----')
61 df_seleccionado = df.iloc[2:5]
62 print(df_seleccionado)
63
64 print('-----')
65 print('--- Apartado 8 ---')
66 print('-----')
67 estadistica = df_seleccionado.describe()
68 print(estadistica)
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
```

	Raza	Puntos	PuntosAna	Media
uno	Caniche	8.1	6.1	7.10
tres	ChowChow	7.6	8.2	7.90
cuatro	Chihuahua	9.0	9.5	9.25
cinco	Labrador	9.3	9.9	9.60
nuevo1	Samoyedo	9.2	8.9	9.05
nuevo2	Pinscher	8.1	6.7	7.40

```
--- Apartado 7 ---
-----
```

	Raza	Puntos	PuntosAna	Media
cuatro	Chihuahua	9.0	9.5	9.25
cinco	Labrador	9.3	9.9	9.60
nuevo1	Samoyedo	9.2	8.9	9.05

```
--- Apartado 8 ---
-----
```

	Puntos	PuntosAna	Media
count	3.000000	3.000000	3.000000
mean	9.166667	9.433333	9.300000
std	0.152753	0.503322	0.278388
min	9.000000	8.900000	9.050000
25%	9.100000	9.200000	9.150000
50%	9.200000	9.500000	9.250000
75%	9.250000	9.700000	9.425000
max	9.300000	9.900000	9.600000