

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

Curso de Especialización: Inteligencia Artificial y Big Data  
Sistemas de Big Data

Módulo profesional:

## ¿Qué contenidos o resultados de aprendizaje trabajaremos?

### Resultados de aprendizaje

- ✓ **RA1.** Aplica técnicas de análisis de datos que integran, procesan y analizan la información, adaptando e implementando sistemas que las utilicen.

### Contenidos

- 1.- Conceptos básicos.
  - 1.1.- Lógica algorítmica.
  - 1.2.- Combinatoria.
  - 1.3.- Complejidad computacional.
- 2.- Analítica y búsqueda de respuestas.
  - 2.1.- Análisis de Datos y Analítica de Datos.
  - 2.2.- Niveles de Analítica de Datos.
  - 2.3.- Principales metodologías en Minería de Datos.
    - 2.3.1.- SEMMA.
    - 2.3.2.- CRISP-DM.
- 3.- R para analizar datos.
  - 3.1.- Primeros pasos.
  - 3.2.- Arrays y Matrices.
  - 3.3.- Listas y Dataframes.
  - 3.4.- Análisis y Visualización.
- 4.- Python para analizar datos.
  - 4.1.- Primeros pasos.
  - 4.2.- Arrays con Numpy.
  - 4.3.- Análisis con Pandas.
  - 4.4.- Visualización con Matplotlib.

Siguiente »

# 1.- Descripción de la tarea.



## Caso práctico

**Enrique y Ana** se conocieron hace 3 años y han decidido irse a vivir juntos.

Les encantan los perros, así que además están pensando en adoptar uno.

Enrique, que es programador, creó hace unos días un sencillo programa en Python para poder dar un vistazo rápido a las razas que más le gustan, a modo de DataFrame de Pandas en el que contempla tanto el nombre de la raza como una puntuación de 0 a 10 indicando cuánto le gusta esa raza.



[Chevanon Photography](#) (Dominio público)

Ana está de acuerdo en adoptar una de las razas que Enrique tiene contempladas, así que le ha dicho cuánto le gusta cada una de ellas (de 0 a 100) para que Enrique lo integre en su DataFrame. Así podrán calcular medias para decidir qué raza les gusta más a los dos.

## ¿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], ['Labrador', 9.3]]

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

Utiliza la información que encontrarás en los contenidos de la unidad para crear un documento en el que escribirás el código fuente que Enrique ha necesitado seguir escribiendo para llegar a las soluciones correspondientes a lo que se te dice en los siguientes apartados.

Al final de los apartados te damos toda la salida por pantalla que deberá tener el programa, para que sepas con claridad si estás consiguiendo lo que se te solicita. Tendrás que usar constantemente la función `print` para ello.

### ✓ 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.

### ✓ 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.

### ✓ 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.

### ✓ 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.

### ✓ Apartado 5:

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

### ✓ Apartado 6:

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

### ✓ 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 incluida-).

### ✓ Apartado 8:

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

A continuación tienes la salida esperada del programa:

```
-----
--- 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      --
-----

      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      --
-----

      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

[« Anterior](#) [Siguiente »](#)

## 2.- Información de interés.



### Recursos necesarios y recomendaciones

#### Recursos necesarios

- ✓ Ordenador personal capaz de ejecutar un editor de textos y conectarse a Internet.
- ✓ Editor de textos.

#### Recomendaciones

- ✓ Antes de abordar la tarea:
  - Lee con detenimiento la unidad, consulta los enlaces para saber más, examina el material proporcionado por el profesor y aclara las dudas que te surjan con él.
  - Realiza el examen online de la unidad, y consulta nuevamente las dudas que te surjan. Sólo cuando lo tengas todo claro, debes abordar la realización de la tarea.
- ✓ No olvides elaborar el documento explicativo.



### Indicaciones de entrega

Una vez realizada la tarea, el envío se realizará a través de la plataforma. El archivo se nombrará siguiendo las siguientes pautas:

**Apellido1\_Apellido2\_Nombre\_SBD04\_Tarea**

Asegúrate que el nombre no contenga la letra ñ, tildes ni caracteres especiales extraños. Así por ejemplo la alumna **Begoña Sánchez Mañas para la cuarta unidad del MP de SBD**, debería nombrar esta tarea como...

**sanchez\_manas\_begona\_SBD04\_Tarea**

[« Anterior](#) [Siguiente »](#)

### 3.- Evaluación de la tarea.



#### Criterios de evaluación implicados

##### Criterios de evaluación RA1

- ✔ d. Se ha construido un conjunto de datos complejos y se han relacionado entre sí.

#### ¿Cómo valoramos y puntuamos tu tarea?

Rúbrica de la tarea	
<b>Apartado 1:</b> Código fuente correcto.	1,25 puntos
<b>Apartado 2:</b> Código fuente correcto.	1,25 puntos
<b>Apartado 3:</b> Código fuente correcto.	1,25 puntos
<b>Apartado 4:</b> Código fuente correcto.	1,25 puntos
<b>Apartado 5:</b> Código fuente correcto.	1,25 puntos
<b>Apartado 6:</b> Código fuente correcto.	1,25 puntos
<b>Apartado 7:</b> Código fuente correcto.	1,25 puntos
<b>Apartado 8:</b> Código fuente correcto.	1,25 puntos

[« Anterior](#)

# Sistemas de Big Data

[Página Principal](#)[Mis cursos](#)[CIDEAD 2022/23](#)[Formación Profesional](#)[Cursos de Especialización](#)[Inteligencia Artificial y Big Data](#)[Sistemas de Big Data](#)[UT04.- Análisis y Búsqueda de Respuestas en Datos.](#)[Foro para SBD04.](#)[Formato entrega Tarea SBD04](#)[Buscar en los foros](#)

Foro para SBD04.

Formato entrega Tarea SBD04

[Configuraciones](#)[examen](#)

Mostrar respuestas anidadas



Formato entrega Tarea SBD04

de [ADRIAN PALENZUELA LOPEZ \(M\)](#) - jueves, 9 de marzo de 2023, 13:37

Buenas

¿Como entregamos la Tarea SBD04? ¿En que formato? (zip, txt, .py...)

En la descripción no pone nada sobre como entregarla.

Un saludo

[Enlace permanente](#) [Responder](#)

Re: Formato entrega Tarea SBD04

de [José Manuel Fernández Díaz](#) - jueves, 9 de marzo de 2023, 20:17

Buenas tardes.

Tienes razón, no se especifica claramente. Deberéis entregarla **principalmente en PDF**, aunque es recomendable también que entreguéis también el código bien en un .py o en .zip con todos las funciones.

De manera similar a la del tema 2, el de MongoDB.

Saludos.

[Enlace permanente](#) [Mostrar mensaje anterior](#) [Responder](#)[examen](#)

Actividad previa

[SBD04.- Análisis y Búsqueda de Respuestas en Datos.](#)

Ir a...

Siguiente actividad