



Tecnológico Nacional de México  
Instituto Tecnológico de Zacatepec  
Ingeniería en Sistemas Computacionales



## **TAREA 13**

### *Ejercicio integral.*

**Materia:** Fundamentos de Análisis de Datos con Python

**Grupo:** XA

**Unidad 2**

**Alumno:** Peralta Vargas Ricardo Emmanuel  
**Número de control:** 19091402

**Profesor:** Tiburcio Zúñiga Mario Humberto

## Objetivo

El objetivo de esta práctica en Python utilizando Anaconda Jupyter es realizar diversas operaciones de procesamiento y análisis de datos sobre el conjunto proporcionado en el archivo 'datosx'. Se llevarán a cabo tareas clave, incluyendo la limpieza de datos, filtrado para mostrar únicamente ciertos datos, cálculos estadísticos como la media aritmética y la creación de una nueva columna llamada 'coeficiente alfa' basada en la relación entre la edad y el peso. Además, se realizarán operaciones de ordenamiento alfabético y descendente según criterios específicos, contribuyendo así a un análisis completo y estructurado de los datos proporcionados.

## Desarrollo

Primero se importa el archivo a utilizar (se utilizó la biblioteca "pandas"):

```
import pandas as pd
df = pd.read_csv("C:/Users/ricar/Desktop/Noveno Semestre/Python XA/datosx.csv")
df
```

	nc	nombre	carrera	cooperacion	sexo	procedencia	Edad	Peso
0	18090907	Angel Axel Cuaxilo Rodriguez	ISC	100.0	Masculino	Amacuzac	21.0	80.0
1	17091058	Everest Daniel Gutierrez Andrade	ISC	100.0	Masculino	Ocuituco	21.0	78.0
2	18091045	Aurora de Jesus Linares Jacobo	ISC	100.0	Femenino	Temixco	22.0	65.0
3	18091054	Mario Martinez Arteaga	ISC	NaN	Masculino	Amacuzac	21.0	78.0
4	18091065	Yeel Montes Camacho	ISC	50.0	Masculino	Zacatepec	22.0	76.0
5	19090356	Georgina Itzel Alvarez	ISC	150.0	Femenino	Jojutla	21.0	75.0
6	19090333	Britney Spears	Hollywood	80.0	Femenino	New York	NaN	NaN
7	18090233	Adrian Flores Herreram	ISC	100.0	Masculino	Zacatepec	22.0	78.0
8	18091049	Miriam Luna Aparicio	ISC	100.0	Femenino	Cuetzala	21.0	75.0
9	18090961	Salvador Jesus Barbina Cabrebra	ISC	100.0	Masculino	Cuernava	22.0	78.0

### a) Limpieza de datos

```
dflimpio = df.dropna()
dflimpio
```

	nc	nombre	carrera	cooperacion	sexo	procedencia	Edad	Peso
0	18090907	Angel Axel Cuaxilo Rodriguez	ISC	100.0	Masculino	Amacuzac	21.0	80.0
1	17091058	Everest Daniel Gutierrez Andrade	ISC	100.0	Masculino	Ocuituco	21.0	78.0
2	18091045	Aurora de Jesus Linares Jacobo	ISC	100.0	Femenino	Temixco	22.0	65.0
4	18091065	Yeel Montes Camacho	ISC	50.0	Masculino	Zacatepec	22.0	76.0
5	19090356	Georgina Itzel Alvarez	ISC	150.0	Femenino	Jojutla	21.0	75.0
7	18090233	Adrian Flores Herreram	ISC	100.0	Masculino	Zacatepec	22.0	78.0
8	18091049	Miriam Luna Aparicio	ISC	100.0	Femenino	Cuetzala	21.0	75.0
9	18090961	Salvador Jesus Barbina Cabrebra	ISC	100.0	Masculino	Cuernava	22.0	78.0

- b) Realizar un filtro en donde solamente aparezca el nombre de la persona, su cooperación y su sexo

```
dflimpio[["nombre", "cooperacion", "sexo"]]
```

	nombre	cooperacion	sexo
0	Angel Axel Cuaxilo Rodirguez	100.0	Masculino
1	Everest Daniel Gutierrez Andrade	100.0	Masculino
2	Aurora de Jesus Linares Jacobo	100.0	Femenino
4	Yeel Montes Camacho	50.0	Masculino
5	Georgina Itzel Alvarez	150.0	Femenino
7	Adrian Flores Herreram	100.0	Masculino
8	Miriam Luna Aparicio	100.0	Femenino
9	Salvador Jesus Barbina Cabrebra	100.0	Masculino

- c) Realizar el cálculo de la media aritmética agrupando por sexo

```
mediaSexo = dflimpio.groupby("sexo").mean()
mediaSexo
```

	nc	cooperacion	Edad	Peso
<b>sexo</b>				
<b>Femenino</b>	18424150.0	116.666667	21.333333	71.666667
<b>Masculino</b>	17890844.8	90.000000	21.600000	78.000000

- d) Trabajando con todos los campos calcular la media aritmética, valor máximo y valor mínimo de las columnas peso y edad agrupando por procedencia

```
desc = dflimpio.groupby('procedencia')[['Peso', 'Edad']].describe()
desc
```

procedencia	Peso								Edad							
	count	mean	std	min	25%	50%	75%	max	count	mean	std	min	25%	50%	75%	max
<b>Amacuzac</b>	1.0	80.0	NaN	80.0	80.0	80.0	80.0	80.0	1.0	21.0	NaN	21.0	21.0	21.0	21.0	21.0
<b>Ocuituco</b>	1.0	78.0	NaN	78.0	78.0	78.0	78.0	78.0	1.0	21.0	NaN	21.0	21.0	21.0	21.0	21.0
<b>Cuernava</b>	1.0	78.0	NaN	78.0	78.0	78.0	78.0	78.0	1.0	22.0	NaN	22.0	22.0	22.0	22.0	22.0
<b>Cuetzala</b>	1.0	75.0	NaN	75.0	75.0	75.0	75.0	75.0	1.0	21.0	NaN	21.0	21.0	21.0	21.0	21.0
<b>Jojutla</b>	1.0	75.0	NaN	75.0	75.0	75.0	75.0	75.0	1.0	21.0	NaN	21.0	21.0	21.0	21.0	21.0
<b>Temixco</b>	1.0	65.0	NaN	65.0	65.0	65.0	65.0	65.0	1.0	22.0	NaN	22.0	22.0	22.0	22.0	22.0
<b>Zacatepec</b>	2.0	77.0	1.414214	76.0	76.5	77.0	77.5	78.0	2.0	22.0	0.0	22.0	22.0	22.0	22.0	22.0

- e) Calcular el número de hombres  
f) Calcular el número de mujeres

```
genero = dflimpio['sexo'].value_counts()
genero
```

```
Masculino    5
Femenino     3
Name: sexo, dtype: int64
```

- g) Mostrar los nombres y la procedencia de aquellas personas cuyo peso sea mayor a 77

```
personas_peso_mayor_77 = dflimpio[dflimpio['Peso'] > 77]
resultados = personas_peso_mayor_77[['nombre', 'procedencia']]
resultados
```

	nombre	procedencia
0	Angel Axel Cuaxilo Rodirguez	Amacuzac
1	Everest Daniel Gutierrez Andrade	Ocuituco
7	Adrian Flores Herreram	Zacatepec
9	Salvador Jesus Barbina Cabrebra	Cuernava

- h) Mostrar los nombres de aquellos hombres cuya edad sea mayor a 21

```
hombres_mayor_21 = dflimpio[(dflimpio['sexo'] == 'Masculino') & (dflimpio['Edad'] > 21)]
nombres_hombres_mayor_21 = nombres_mayor_21['nombre']
nombres_hombres_mayor_21
4          Yeel Montes Camacho
7          Adrian Flores Herreram
9  Salvador Jesus Barbina Cabrebra
Name: nombre, dtype: object
```

- i) Calcular una nueva columna llamada "coeficiente\_alfa" que equivalga a un valor vinculado a la edad y al peso (el coeficiente se calcula dividiendo el peso entre la edad)

```
dflimpio["coeficiente_alfa"] = dflimpio['Peso']/dflimpio['Edad']
dflimpio
```

	nc	nombre	carrera	cooperacion	sexo	procedencia	Edad	Peso	coeficiente_alfa
0	18090907	Angel Axel Cuaxilo Rodirguez	ISC	100.0	Masculino	Amacuzac	21.0	80.0	3.809524
1	17091058	Everest Daniel Gutierrez Andrade	ISC	100.0	Masculino	Ocuituco	21.0	78.0	3.714286
2	18091045	Aurora de Jesus Linares Jacobo	ISC	100.0	Femenino	Temixco	22.0	65.0	2.954545
4	18091065	Yeel Montes Camacho	ISC	50.0	Masculino	Zacatepec	22.0	76.0	3.454545
5	19090356	Georgina Itzel Alvarez	ISC	150.0	Femenino	Jojutla	21.0	75.0	3.571429
7	18090233	Adrian Flores Herreram	ISC	100.0	Masculino	Zacatepec	22.0	78.0	3.545455
8	18091049	Miriam Luna Aparicio	ISC	100.0	Femenino	Cuetzala	21.0	75.0	3.571429
9	18090961	Salvador Jesus Barbina Cabrebra	ISC	100.0	Masculino	Cuernava	22.0	78.0	3.545455

j) Ordenar alfabéticamente el dataframe original de acuerdo al nombre

```
dfordenado = dflimpio.sort_values(by='nombre')
dfordenado
```

	nc	nombre	carrera	cooperacion	sexo	procedencia	Edad	Peso	coeficiente_alfa
7	18090233	Adrian Flores Herreram	ISC	100.0	Masculino	Zacatepec	22.0	78.0	3.545455
0	18090907	Angel Axel Cuaxilo Rodriguez	ISC	100.0	Masculino	Amacuzac	21.0	80.0	3.809524
2	18091045	Aurora de Jesus Linares Jacobo	ISC	100.0	Femenino	Temixco	22.0	65.0	2.954545
1	17091058	Everest Daniel Gutierrez Andrade	ISC	100.0	Masculino	Ocuituco	21.0	78.0	3.714286
5	19090356	Georgina Itzel Alvarez	ISC	150.0	Femenino	Jojutla	21.0	75.0	3.571429
8	18091049	Miriam Luna Aparicio	ISC	100.0	Femenino	Cuetzala	21.0	75.0	3.571429
9	18090961	Salvador Jesus Barbina Cabrebra	ISC	100.0	Masculino	Cuernava	22.0	78.0	3.545455
4	18091065	Yeel Montes Camacho	ISC	50.0	Masculino	Zacatepec	22.0	76.0	3.454545

k) Ordenar el dataframe en orden descendente de acuerdo a: cooperación, edad y peso

```
dfordenado2 = dflimpio.sort_values(by=['cooperacion', 'Edad', 'Peso'], ascending=[False, False, False])
dfordenado2
```

	nc	nombre	carrera	cooperacion	sexo	procedencia	Edad	Peso	coeficiente_alfa
5	19090356	Georgina Itzel Alvarez	ISC	150.0	Femenino	Jojutla	21.0	75.0	3.571429
7	18090233	Adrian Flores Herreram	ISC	100.0	Masculino	Zacatepec	22.0	78.0	3.545455
9	18090961	Salvador Jesus Barbina Cabrebra	ISC	100.0	Masculino	Cuernava	22.0	78.0	3.545455
2	18091045	Aurora de Jesus Linares Jacobo	ISC	100.0	Femenino	Temixco	22.0	65.0	2.954545
0	18090907	Angel Axel Cuaxilo Rodriguez	ISC	100.0	Masculino	Amacuzac	21.0	80.0	3.809524
1	17091058	Everest Daniel Gutierrez Andrade	ISC	100.0	Masculino	Ocuituco	21.0	78.0	3.714286
8	18091049	Miriam Luna Aparicio	ISC	100.0	Femenino	Cuetzala	21.0	75.0	3.571429
4	18091065	Yeel Montes Camacho	ISC	50.0	Masculino	Zacatepec	22.0	76.0	3.454545

l) ordenar el dataframe en orden ascendente por sexo y luego por edad

```
dfordenado = dflimpio.sort_values(by=['nombre', 'Edad'])
dfordenado
```

	nc	nombre	carrera	cooperacion	sexo	procedencia	Edad	Peso	coeficiente_alfa
7	18090233	Adrian Flores Herreram	ISC	100.0	Masculino	Zacatepec	22.0	78.0	3.545455
0	18090907	Angel Axel Cuaxilo Rodriguez	ISC	100.0	Masculino	Amacuzac	21.0	80.0	3.809524
2	18091045	Aurora de Jesus Linares Jacobo	ISC	100.0	Femenino	Temixco	22.0	65.0	2.954545
1	17091058	Everest Daniel Gutierrez Andrade	ISC	100.0	Masculino	Ocuituco	21.0	78.0	3.714286
5	19090356	Georgina Itzel Alvarez	ISC	150.0	Femenino	Jojutla	21.0	75.0	3.571429
8	18091049	Miriam Luna Aparicio	ISC	100.0	Femenino	Cuetzala	21.0	75.0	3.571429
9	18090961	Salvador Jesus Barbina Cabrebra	ISC	100.0	Masculino	Cuernava	22.0	78.0	3.545455
4	18091065	Yeel Montes Camacho	ISC	50.0	Masculino	Zacatepec	22.0	76.0	3.454545

## **Conclusión**

Esta práctica en Python ha permitido realizar un análisis exhaustivo del conjunto de datos contenido en el archivo 'datosx'. A través de la aplicación de técnicas de limpieza, filtrado y cálculos estadísticos, se obtuvieron valores valiosos sobre la distribución de la información, destacando tendencias y patrones específicos. Las operaciones de agrupación por sexo, procedencia y la creación de la columna 'coeficiente alfa' han proporcionado una comprensión más profunda del tema. Además, los resultados obtenidos al identificar el número de hombres y mujeres, así como las personas con características específicas, enriquecen la comprensión general del conjunto de datos. El ordenamiento alfabético y descendente según criterios relevantes ha facilitado la interpretación de los resultados y resaltado aspectos clave del conjunto de datos para un análisis más efectivo. En conjunto, esta práctica demuestra la utilidad de las herramientas de Python en el procesamiento y análisis de datos, proporcionando una base sólida para futuras prácticas.