

5 - GESTIONANDO DATOS Y GRÁFICOS



En esta unidad se exponen conceptos básicos de los módulos `numpy` y `matplotlib`, que definen clases y métodos útiles en el análisis estadístico de datos.

Python define tuplas, listas, conjuntos y diccionarios como tipos básicos con los que el desarrollador puede gestionar colecciones de datos de cualquier tipo. Una característica de estas colecciones es que pueden guardar en la misma colección datos de diferentes tipos, incluyendo sub-colecciones. Esta característica da mucha flexibilidad al desarrollador, pero le obliga a tener un mayor control sobre cada uno de los datos que contiene, utilizo métodos genéricos y poco eficientes cuando queremos hacer análisis estadísticos sobre colecciones de datos.

El módulo `numpy` define `ndarray` para almacenar y gestionar eficientemente colecciones de datos del mismo tipo. La gestión de archivos con este tipo de colecciones es elemental, facilitando métodos con los que el desarrollador solo necesita unas líneas de código para guardar los datos y recuperarlos.

El módulo `matplotlib` define las clases `plot` e `hist`, entre otras, para representar gráficamente colecciones de datos del tipo `ndarray`, entre otros.

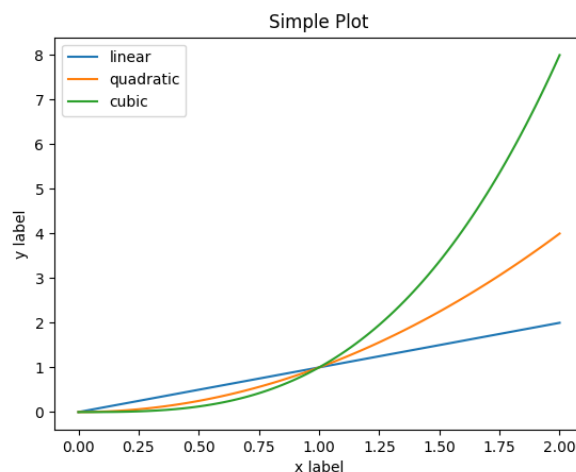
```
import matplotlib.pyplot as plt
import numpy as np

x = np.linspace(0, 2, 100)

plt.plot(x, x, label='linear')
plt.plot(x, x**2, label='quadratic')
plt.plot(x, x**3, label='cubic')

plt.xlabel('x label')
plt.ylabel('y label')
plt.title("Simple Plot")
plt.legend()

plt.show()
```



Los módulos `pandas`, `scipy`, `scikit-learn`, `csv`, `json`, `sqlite3`, `openpyxl`, `xlrd`, `xlutils` o `pyexcel` proporcionan clases y métodos que suplen las carencias de Python en la gestión de datos en formatos o usando diferentes formatos de archivo.