



Máster Universitario en Ciencia de Datos

# **Herramientas de Programación: Python**

**Guía de Estudio**  
Curso 2023/24

# Información práctica

Asignatura: Herramientas de Programación: Python  
Profesor: Dr. Alejandro Baldominos Gómez  
Duración: 4 semanas (15 sesiones)  
Créditos: 3 ECTS

## Descripción del curso

La asignatura busca proporcionar al estudiante una base introductoria pero sólida a la programación informática, con un marcado enfoque a las aplicaciones de ciencias de datos. Esta asignatura pretende por tanto asentar los fundamentos necesarios para que el estudiante pueda enfrentarse a asignaturas como aprendizaje automático, técnicas de visualización, redes neuronales y aprendizaje profundo, etc.

Para esta parte de la asignatura se ha escogido uno de los lenguajes de programación con uso más extendido en la disciplina de la ciencia de datos: Python. Este es un lenguaje con una orientación general, pero con un importante ecosistema de librerías eficientes para el cálculo numérico, el análisis de datos y el aprendizaje automático.

Tras superar la asignatura, el estudiante estará en condiciones de abordar con satisfacción y madurez técnica materias cuyo contenido teórico es más específico y complejo, pero cuyos problemas prácticos pueden resolverse mediante la combinación de técnicas vistas en esta asignatura.

## Contenidos

### Tema 1 | Introducción a la programación en Python

En este tema se comenzará presentando la segunda mitad de la asignatura, consistente en programación en Python, concretando la metodología de trabajo y los mecanismos de evaluación. Se explicará de forma superficial la sintaxis de Python, comparándola con la sintaxis de R, ya familiar. Se presentará el entorno de desarrollo en el que se trabajará durante las sesiones y se recordará el concepto de «variable» como el mecanismo más simple de almacenamiento de datos, presentando los distintos tipos básicos de datos soportados en Python.

- Presentación de los contenidos y la evaluación.
- Introducción a la sintaxis de Python.
- Entorno de desarrollo.
- Declaración de variables y tipos de datos.

### Tema 2 | Estructuras de datos y colecciones

En Python, existen mecanismos para almacenar colecciones de datos, tales como listas, diccionarios, tuplas y conjuntos. En este tema se explicarán los dos primeros con un nivel elevado de detalle, y se presentarán otras estructuras complementarias con algunos casos de uso.

- Colecciones en Python.
- Listas.
- Diccionarios.
- Otras estructuras de datos.

### Tema 3 | Control del flujo

Como en la mayoría de lenguajes de programación, Python cuenta con mecanismos para permitir el control del flujo de la ejecución de un programa. Los más notables son las sentencias condicionales, que permiten ejecutar bloques de código en función de si se cumplen o no determinadas condiciones; y los bucles, que permiten repetir la ejecución de un bloque de código un número determinado de veces. En este tema se verán en profundidad ambos mecanismos. Además, se introducirá la captura de excepciones como mecanismo para controlar el flujo de un programa en caso de errores.

- Condicionales.
- Bucles e iteradores.
- Excepciones y gestión de errores.

### Tema 4 | Funciones

Las funciones son un mecanismo esencial de todo lenguaje de programación para permitir la reutilización de código y la implementación organizada de programas complejos. En este tema se presenta el concepto de función y se explica cómo diseñar funciones para la transformación de datos.

- Funciones en Python.
- Transformación de datos mediante funciones.

### Tema 5 | Python para ciencia de datos

Python es un lenguaje ampliamente usado para resolver problemas de análisis de datos. Para ello, se suele hacer uso de la librería Pandas, que permite mediante una interfaz sencilla realizar numerosas transformaciones, filtrados y resúmenes de los datos, además de simplificar el proceso de carga y exportación. En este tema, se presentarán estos conceptos para hacer frente a algunos problemas sencillos de análisis de datos.

- Librerías en Python: Pandas.
- El concepto de Dataframe.
- Carga de datos desde ficheros.
- Filtrado de datos.
- Transformación de datos.
- Agrupación y agregación de datos.
- Exportación.

### Tema 6 | Cálculo numérico

En muchas ocasiones, es habitual trabajar con datos numéricos, de los que se quiere obtener determinado valor. Aunque Python proporciona algunos mecanismos sencillos para llevar a cabo esta operativa, lo más habitual es emplear librerías específicas para el cálculo numérico, siendo Numpy y Scipy las más habituales. En este tema, se presenta el concepto de array numérico y mecanismos para almacenar datos en arrays y para procesarlos empleando Numpy y Scipy. Además, con el fin de ilustrar todo lo visto anteriormente de forma práctica y alimentando el ingenio y el pensamiento matemático, repasaremos el concepto de derivada e integral y plantearemos su resolución numérica (no simbólica) empleando Python.

- El concepto de Array.
- Cálculo numérico empleando Numpy.
- Conceptos estadísticos empleando Scipy.
- Pensamiento matemático y resolución de problemas.

## Tema 7 | Visualización

En el trabajo con datos, es muy frecuente realizar visualizaciones de los mismos con el fin de entender mejor su estructura y disposición. En este tema, se presentarán algunas visualizaciones ampliamente utilizadas, discutiendo sus ventajas y desventajas en diferentes escenarios. Para ello, se hará uso de la librería Pandas conjuntamente con Seaborn.

- Introducción a la visualización de datos.
- Histogramas y gráficos de cajas en Pandas.
- Visualizaciones avanzadas con Seaborn.

## Tema 8 | Datos en APIs

Aunque hasta este momento habremos trabajado extensamente con datos disponibles en ficheros con formatos familiares para el estudiante, como Excel o CSV, lo cierto es que el trabajo con datos da un salto de gigante cuando estas fuentes de datos se complementan con otras accesibles a través de Internet, cuya información se puede recuperar a través de las denominadas APIs. En este tema, discutiremos el formato JSON como formato habitual de intercambio de datos y resaltaremos su analogía con un diccionario de Python. Nos conectaremos a diversas fuentes de datos e incluso trataremos de combinar estos datos para obtener un valor adicional.

- Introducción a las APIs.
- Introducción al formato JSON.
- Conexión con APIs para la obtención de datos abiertos.
- Combinación de datos. Operaciones join y merge.

## Tema 9 | Pensamiento computacional e introducción al aprendizaje automático

Este tema no presenta conceptos nuevos, pero aprovecha para dar una importante vuelta de tuerca a todo lo visto hasta el momento para ponerlo a funcionar de un modo que el estudiante probablemente no hubiera imaginado. El tema se introduce con un problema en apariencia ambicioso: resolver un problema de predicción numérico (regresión). El estudiante difícilmente sabrá cómo proceder, pero a lo largo de la sesión se irá desvelando la resolución al problema aplicando todo lo aprendido: la carga de datos, su estructuración de forma adecuada, el filtrado y transformación de valores, la extracción de muestras aleatorias, y el empleo de los datos para aprender los parámetros de un modelo de regresión lineal y posteriormente visualizarlo y validarlo. A excepción de algunas funciones y, tal vez, del funcionamiento de la regresión lineal, el estudiante no hará uso de ningún concepto nuevo. Sin embargo, al finalizar la sesión, habrá podido aplicar los conocimientos adquiridos para resolver un problema complejo, que nos anticipa al mundo del aprendizaje automático que el estudiante verá posteriormente en el Máster.

# Metodología

La asignatura tiene un enfoque fundamentalmente práctico. Si bien en cada tema se presentarán conceptos de programación y se revisará la sintaxis para poder implementar estos conceptos en Python, el planteamiento del curso está dirigido a resolver problemas concretos y, en particular, aquellos orientados a la ciencia de datos.

Así, buena parte de las sesiones estarán dedicadas a escribir código en Python. Para ello, se empleará un «cuaderno de Python», que es una herramienta contenida en un servidor y que permite al estudiante trabajar sin necesidad de instalar ningún software.

Asimismo, las pruebas de evaluación están planteadas de un modo práctico, consistiendo en ejercicios de programación que, según avance el curso, irán más destinadas al análisis de datos y la resolución de problemas.

# Cronograma

📅 L 02/10/2023 ⌚ 19:30 – 21:00 Sesión 1

**Tema 1:** Introducción a la programación en Python (I).

📅 J 05/10/2023 ⌚ 17:30 – 21:00 Sesiones 2 y 3

**Tema 1:** Introducción a la programación en Python (II).

**Tema 2:** Estructuras de datos y colecciones.

📅 L 09/10/2023 ⌚ 17:30 – 21:00 Sesiones 4 y 5

**Tema 3:** Control del flujo.

**Tema 4:** Funciones.

📅 M 10/10/2023 ⌚ 17:30 – 19:00 Sesión 6

**Tema 5:** Python para ciencia de datos (I).

📅 L 16/10/2023 ⌚ 17:30 – 21:00 Sesiones 7 y 8

**Tema 5:** Python para ciencia de datos (II).

**Repaso y actividades prácticas:** Se repasarán de forma práctica los contenidos vistos hasta el momento, presentando el primer ejercicio práctico y resolviendo ejercicios preparatorios conjuntamente en clase.

📅 M 17/10/2023 ⌚ 17:30 – 19:00 Sesión 9

**Tema 6:** Cálculo numérico (I).

📅 L 23/10/2023 ⌚ 17:30 – 19:00 Sesión 10

**Tema 6:** Cálculo numérico (II).

📅 M 24/10/2023 ⌚ 17:30 – 19:00 Sesión 11

**Tema 7:** Visualización.

📅 L 30/10/2023 ⌚ 17:30 – 19:00 Sesión 12

**Tema 8:** Datos en APIs.

📅 M 31/10/2023 ⌚ 17:30 – 19:00 Sesión 13

**Tema 9:** Pensamiento computacional e introducción al aprendizaje automático (I).

📅 M 07/11/2023 ⌚ 17:30 – 19:00 Sesión 14

**Tema 9:** Pensamiento computacional e introducción al aprendizaje automático (II).

📅 M 14/11/2023 ⌚ 17:30 – 19:00 Sesión 15

**Repaso y actividades prácticas:** Esta sesión se dedicará enteramente a repasar de forma práctica los contenidos vistos hasta el momento. Además, se presentará el segundo ejercicio práctico, trabajando en algunos ejercicios previos para poder acometerlo satisfactoriamente.

# Evaluación

La evaluación se llevará a cabo mediante proyectos de programación, siguiendo un esquema de evaluación objetivo. Según este esquema, se valorará que la funcionalidad solicitada funcione correctamente, para lo cual se probará de forma automatizada sobre una serie de casos de prueba, lo que determinará la puntuación obtenida por el estudiante.

## **Ejercicio Práctico 1**    **Puntuación: 20%**    **Aprox. 22/10/2023**

Realización de un proyecto de programación en Python que incluirá todos los conceptos vistos entre los temas 1 y 5, pensado en evaluar sobre todo la capacidad de desarrollar programas sencillos empleando la sintaxis de Python y la capacidad de llevar a cabo un sencillo análisis de datos.

## **Ejercicio Práctico 2**    **Puntuación: 30%**    **Aprox. 19/11/2023**

Realización de un proyecto de programación en Python que incluirá todos los conceptos vistos entre los temas 1 y 9, haciendo un mayor hincapié en el pensamiento computacional y la resolución de problemas.

## **Examen Final**    **Puntuación: 50%**    **TBD**

Examen final que se realizará con ordenador, consistente en resolver problemas de programación. Se valorará que el planteamiento de la resolución sea adecuado y que el código desarrollado funcione correctamente.

### **Importante**

Los trabajos de evaluación, así como el examen final, son individuales y cada estudiante debe trabajar autónomamente. Cualquier signo de trabajo fraudulento o plagio en la elaboración de los ejercicios de evaluación tendrá como consecuencia el suspenso en la asignatura.

# Bibliografía

No es necesario adquirir o tomar prestado ningún material para completar la asignatura. No obstante, si se desea, las siguientes referencias bibliográficas son útiles para ampliar y profundizar en los contenidos de la asignatura.

### **Think Python: How to Think Like a Computer Scientist, 2nd Ed.**

*Allan B. Downey. O'Reilly, 2012. ISBN: 978-1491939369.*

Se trata de una estupenda referencia para aprender Python con un enfoque dirigido a la resolución de problemas, fomentando el pensamiento computacional. El libro está disponible en formato PDF de forma gratuita, incluyendo una versión en castellano, que se proporcionará en el aula virtual.

### **Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython, 3rd Ed.**

*Wes McKinney. O'Reilly, 2022. ISBN: 978-1098104030.*

Este volumen (solo disponible en inglés) es una estupenda referencia para profundizar en el tratamiento de datos con Python, ampliando sustancialmente los conceptos vistos en la asignatura.