

# **5101 - Análisis de Datos con Python**

## **Curso de Especialización en Desarrollo de Aplicaciones en Lenguaje Python**

# 1. ¿Dónde se enmarca este módulo?

Este módulo forma parte del:

- **Curso de Especialización de Formación Profesional de Grado Superior en Desarrollo de aplicaciones en lenguaje Python.**
- Al superar el curso, obtendréis el título de **Máster de Formación Profesional.**

Está compuesto por cuatro módulos profesionales:

- *5098. Entornos y sintaxis en Python.*
- *5099. Estructuras de control en Python.*
- *5100. Programación orientada a objetos.*
- **5101. Análisis de datos con Python.**

## 2. Visión General del Módulo 5101

Vamos a centrarnos en los detalles de **nuestro** módulo:

- **Denominación:** Análisis de datos con Python.
- **Código:** 5101.
- **Duración total:** 75 horas.
- **Créditos ECTS:** 9.

El objetivo es claro: aprender el ciclo completo del análisis de datos, desde la obtención hasta la creación de modelos.

### 3. Objetivo General del Módulo

Al finalizar este módulo, seréis capaces de:

Manejar, limpiar, analizar y desarrollar modelos predictivos a partir de diversos conjuntos de datos utilizando Python y sus librerías especializadas.

Esto se articula a través de **4 Resultados de Aprendizaje (RA)** que iremos cubriendo progresivamente.

## 4. Nuestra Hoja de Ruta: Unidades Didácticas

El módulo está estructurado en **10 Unidades Didácticas (UD)**, cada una con una duración aproximada de 7.5 horas.

1. **Introducción** al Análisis de Datos y Entornos Python.
2. Lectura y **Exportación de Datos** Estructurados.
3. Interacción con **Bases de Datos**.
4. **Verificación** y Preparación Inicial de Datos.
5. **Normalización** y Estandarización de Datos.
6. Manejo de **Variables Categóricas** y Cuantitativas.
7. Análisis Exploratorio I: **Resumen y Patrones**.
8. Análisis Exploratorio II: **Correlación**.
9. Modelos Predictivos I: **Relaciones Lineales y No Lineales**.
10. Modelos Predictivos II: **Evaluación y Validación**.

## RA 1: Adquisición y Manejo de Datos

Nos centraremos en obtener y manipular datos de diversas fuentes.

- **Unidades:** UD 1, UD 2, UD 3 y UD 4.
- **Contenidos clave:**
  - Importación de librerías esenciales (Pandas, NumPy, etc.).
  - Lectura y exportación de ficheros **CSV y XLSX**.
  - Conexión y operaciones **CRUD** con bases de datos.
  - Verificación inicial de la estructura y contenido de los datos.

## RA 2: Limpieza y Estandarización

Una vez tenemos los datos, ¡hay que prepararlos! La calidad del dato es fundamental.

- **Unidades:** UD 4, UD 5 y UD 6.
- **Contenidos clave:**
  - Análisis de los datos leídos para comprenderlos.
  - **Normalización y estandarización** de datos.
  - Procesos de **limpieza de datos** basados en el problema a resolver.
  - Tratamiento de **variables categóricas** y su conversión a numéricas.

## RA 3: Análisis Exploratorio de Datos (EDA)

Con los datos limpios, empezamos a extraer conocimiento. ¿Qué nos cuentan los datos?

- **Unidades:** UD 7 y UD 8.
- **Contenidos clave:**
  - Técnicas para **resumir grandes volúmenes** de datos.
  - Identificación de **patrones y tendencias**.
  - Formulación y respuesta a preguntas relevantes sobre los datos.
  - Cálculo e interpretación de la **correlación** entre variables.



## RA 4: Desarrollo de Modelos Predictivos

El paso final: utilizar los datos para predecir el futuro.

- **Unidades:** UD 9 y UD 10.
- **Contenidos clave:**
  - Modelado de **relaciones lineales y no lineales**.
  - Selección de **muestras de ensayo (train) y evaluación (test)**.
  - Realización de **predicciones**.
  - **Comprobación de la precisión y validez** del modelo con datos de muestra.

## 5. Evaluación

La evaluación será eminentemente **práctica y continua**, basada en los criterios oficiales del módulo. Se valorará principalmente vuestra capacidad para:

- **Escribir código Python** que resuelva problemas específicos de cada unidad (leer datos, limpiarlos, analizarlos, etc.).
- **Analizar y deducir** las operaciones necesarias para tratar un conjunto de datos.
- **Desarrollar y validar** modelos que den solución a los problemas planteados.

Realizaremos ejercicios prácticos por cada unidad y, previsiblemente, un proyecto final que integre todos los conocimientos.

## 6. Herramientas y Librerías Clave

Durante el curso, nos apoyaremos en el ecosistema de Python para la ciencia de datos. Las protagonistas serán:

- **Entorno de Trabajo:**
  - Visual Studio Code
  - JetBrains PyCharm
  - Jupyter Notebooks
- **Librerías Básicas:**
  - **Pandas:** Para la manipulación y análisis de datos tabulares.
  - **NumPy:** Para el cálculo numérico y trabajo con arrays.
- **Visualización:**
  - **Matplotlib:** Para la creación de gráficos estáticos y de calidad.
  - **Seaborn:** Basada en Matplotlib, para gráficos estadísticos más atractivos.
- **Modelado:**
  - **Scikit-learn:** La librería de referencia para Machine Learning en Python.

## 7. Preguntas y Dudas

Este es el viaje que vamos a emprender juntos. Es un campo apasionante y con un potencial enorme.

**¿Alguna pregunta inicial?**

...