

Las Herramientas del científico de datos

Prólogo: La ciencia de datos

Juan Manuel Moreno — jmmoreno@profesorescol.imf.com



Juan Manuel Moreno Lamparero

- Grado en Sistemas de Información (UAH)
- Máster en Big Data & Business Analytics (UAH)
- Everis – Desarrollador Java para el sector público.
- Investigador, Cátedra en Big Data y Analítica Predictiva Bancaria.
- Cognicor Technologies — Consultant ML Architect.
- IMF – Tutor máster Big Data & Business Analytics.
- IMF – Tutor, autor, profesor máster Data Science.



TU TURNO

Calendario de sesiones

FECHA	MÓDULO	Unidad
17/04/2023	I	1 — Fundamentos de Python
18/04/2023	I	1 — Fundamentos de Python
19/04/2023	I	1 — Fundamentos de Python
20/04/2023	I	1 — Fundamentos de Python
24/04/2023	I	2 — Análisis de datos con Numpy
25/04/2023	I	2 — Análisis de datos con Pandas
26/04/2023	I	3 — Visualización de datos con Python (Matplotlib, Seaborn, Plotly)
27/04/2023	I	4 — Fundamentos de R
03/05/2023	I	4 — Fundamentos de R
04/05/2023	I	4 — Fundamentos de R
08/05/2023	I	5 — Paquetes de R (Tidyverse)
09/05/2023	I	6 — Visualización de datos con R (Graphics, Ggplot2, Plotly)

Objetivos del módulo

- Descubrir en qué consiste la ciencia de datos
- Conocer los fundamentos de programación en Python.
- Saber implementar arrays multidimensionales en Python.
- Diseñar y trabajar con dataframes en Python.
- Aprender a utilizar diferentes librerías de visualización en Python.
- Conocer los fundamentos de programación en R.
- Diseñar y trabajar con dataframes en R.
- Aprender a implementar diferentes gráficos en R.

Caso práctico

- Realización de un análisis exploratorio (EDA) en Python.
- Realización de un análisis exploratorio (EDA) en R.

Competencias del módulo

- Conocer los principales elementos fundamentales de programación, tanto para Python como para R.
- Desarrollar y aplicar funciones básicas para transformar datos
- Aprender a utilizar estructuras de datos esenciales como tuplas, listas, diccionarios de datos, arrays y dataframes.
- Saber controlar el flujo de un programa.
- Escoger un tipo de gráfico apropiado para una visualización de datos.
- Conocer los procedimientos para instalar un paquete tanto de Python como de R.
- Volcar información desde archivos a dataframes.
- Aplicar diferentes transformaciones y funciones a dataframes.

01

Prólogo: La ciencia de datos

1.1.– Introducción

Introducción

- Cada vez nos son más frecuentes los términos de Big Data, Internet de las Cosas (IoT), Inteligencia Artificial (IA) o la Industria 4.0.
- Seguimos generando ingentes cantidades de datos día a día.
- ¿Cómo podemos procesar, entender y sacar valor de estos datos ? – **Científico de datos.**

Competencias de un Científico de datos



1.– Prólogo la ciencia de datos

1.2.– Algunas herramientas

Algunas herramientas

- *Magic Quadrant Gartner*

<https://powerbi.microsoft.com/en-us/blog/microsoft-named-a-leader-in-the-2023-gartner-magic-quadrant-for-analytics-and-bi-platforms/>

Figure 1: Magic Quadrant for Analytics and Business Intelligence Platforms



Source: Gartner

Conocer la base a través de Python y R

- La gran mayoría de estas herramientas no son de código libre y, por lo tanto, suponen costes de adaptación e implantación, además, es importante para un científico de datos ser conocedor de las principales transformaciones que se pueden realizar sobre los datos desde una perspectiva de desarrollador software.
- En este módulo se trabajará íntegramente con los de los lenguajes de programación estrella dentro del mundo de la ciencia de datos: Python y R

IMF

Smart Education