

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 — Programas en: Big Data | Data Science | IA
- Freelance



TU TURNO

Calendario de sesiones

FECHA	MÓDULO	Unidad
20/11/2023	I	1 — Fundamentos de Python 1
21/11/2023	I	1 — Fundamentos de Python 2
22/11/2023	I	1 — Fundamentos de Python 3
23/11/2023	I	1 — Fundamentos de Python 4
27/11/2023	I	2 — Análisis de datos con Numpy
28/11/2023	I	2 — Análisis de datos con Pandas
29/11/2023	I	3 — Visualización de datos con Python (Matplotlib, Seaborn, Plotly) 1
30/11/2023	I	4 — Visualización de datos con Python (Matplotlib, Seaborn, Plotly) 2
Puente de la constitución		
11/12/2023	I	4 — Fundamentos de R 1
12/12/2023	I	4 — Fundamentos de R 2
13/12/2023	I	5 — Paquetes de R (Tidyverse)
14/12/2023	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.
- Emplear diferentes estructuras de datos en R.
- Trabajar con dataframes en R.
- Aprender a implementar diferentes gráficos en R.



Evaluación final

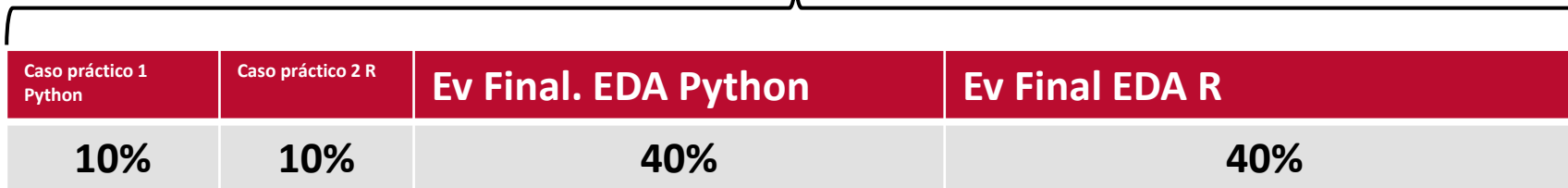
- Ejercicios de seguimiento:
 - Caso práctico 1 Python
 - Caso práctico 2 R
- Análisis Exploratorio de Datos (EDA)
 - Realización de un análisis exploratorio (EDA) en Python.
 - Realización de un análisis exploratorio (EDA) en R.

Caso práctico 1 Python	Caso práctico 2 R	Ev Final. EDA Python	Ev Final EDA R
10%	10%	40%	40%

Evaluación final

- Ejercicios de seguimiento:
 - Caso práctico 1 Python
 - Caso práctico 2 R
- Análisis Exploratorio de Datos (EDA)
 - Realización de un análisis exploratorio (EDA) en Python.
 - Realización de un análisis exploratorio (EDA) en R.

Los casos prácticos se entregan desde el campus virtual



Caso práctico 1 Python	Caso práctico 2 R	Ev Final. EDA Python	Ev Final EDA R
10%	10%	40%	40%

Evaluación final

- Ejercicios de seguimiento:
 - Caso práctico 1 Python
 - Caso práctico 2 R
- Análisis Exploratorio de Datos (EDA)
 - Realización de un análisis exploratorio (EDA) en Python.
 - Realización de un análisis exploratorio (EDA) en R.

Los casos prácticos se entregan desde el campus virtual

Caso práctico 1 Python	Caso práctico 2 R	Ev Final. EDA Python	Ev Final EDA R
10%	10%	40%	40%

Cuando se finaliza el módulo, plazo largo generalmente 2 semanas

Se manda cuando se termina el bloque de R, plazo corto 2-3 días

Se manda cuando se termina el bloque de Python, plazo corto 2-3 días

Competencias del módulo

- Conocer los principales elementos fundamentales de programación, tanto para Python como para R.
- Desarrollar y aplicar funciones básicas para transformar y limpiar 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.

01

Prólogo: La ciencia de datos

Introducción

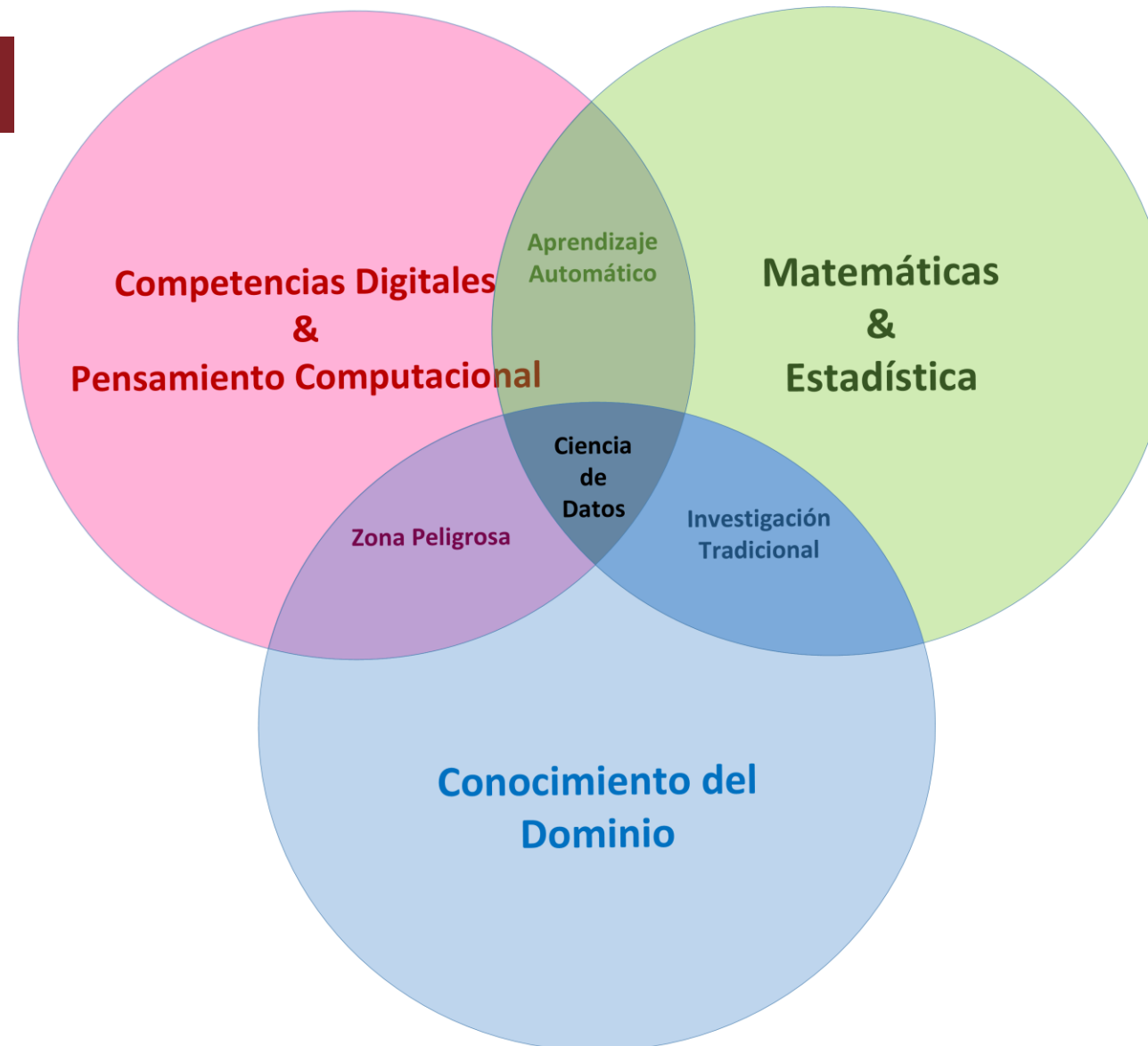
- Cada vez nos son más frecuentes los términos de Big Data, Internet de las Cosas (IoT), Inteligencia Artificial (IA) o la Industria 4.0.
- Seguimos generando ingentes cantidades de datos día a día.
- ¿Cómo podemos procesar, entender y sacar valor de estos datos ? — **Científico de datos.**

Competencias de un Científico de datos



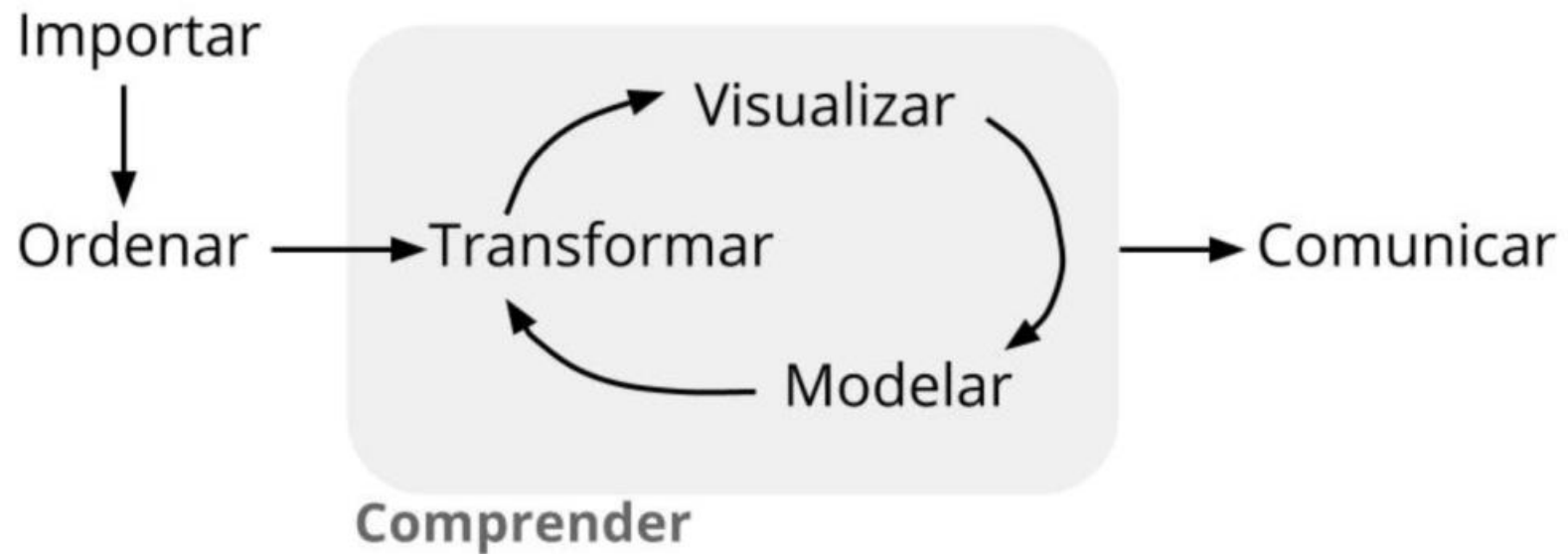
1.– Prólogo la ciencia de datos

1.1.– Introducción



1.- Prólogo la ciencia de datos

1.1.- Introducción



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

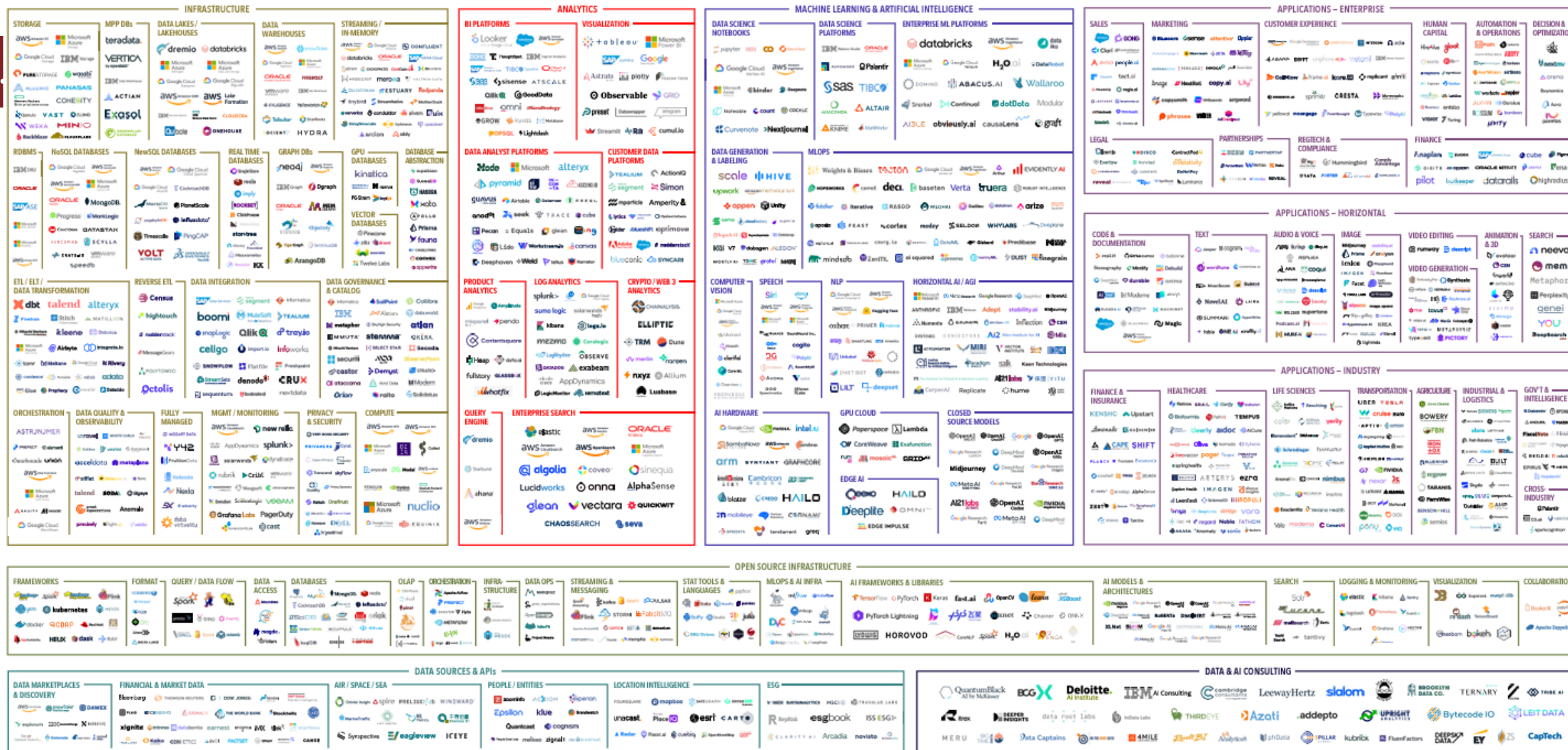
Algunas herramientas

- *The 2023 MAD (Machine Learning, IA, Data) Landscape*

<https://mathtturck.com/landscape/mad2023.pdf>

1.- Prólogo la ciencia de datos

THE 2023 MAD (MACHINE LEARNING, ARTIFICIAL INTELLIGENCE & DATA) LANDSCAPE



1.2.– Algunas herramientas

- Pero ¿Porqué tantas herramientas? ¿Tantas tengo que saber? ¿En qué punto nos encontramos ahora?
- En la industria actual se sigue una tendencia de tratar de automatizar al máximo las tareas que involucran desde que llegan los datos, hasta que se toman decisiones sobre ellos (plataformas E2E)
- No, no tenéis que saber utilizar tantas herramientas, la gran mayoría de ellas trabajan siempre a través de:
 - Organización de orígenes de datos
 - Limpieza de datos
 - Transformación, pre-procesamiento y selección de variables
 - Aplicación de modelos de aprendizaje automático | IA
 - Toma de decisiones (servicios web, cuadros de mando...)
- De hecho, muchas de estas herramientas llevan “por detrás” código Python, es decir, en su origen están basadas en OpenSource, así que, vamos a comenzar desde el OpenSource con Python.

1.3.– Competencias a través de Python y R

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