

RAUL BENÍTEZ —

ANÁLISIS EXPLORATORIO DE DATOS

raul.benitez@upc.edu

S1 → INTRODUCCIÓN IA & ECOSISTEMA PYTHON + GOOGLE COLAB

S2 → MANIPULACIÓN DATOS CON PANDAS

S3 → AGRUPAMIENTO DE DATOS (CLUSTERING)

S4 → REDUCCIÓN DE LA DIMENSIONALIDAD

S5 → REGRESIÓN

S6 → SELECCIÓN DE ATRIBUTOS

S7 → PROYECTO HANDS-ON

ECOSISTEMA

PYTHON - language programación

GOOGLE COLAB - | plataforma programación
python notebook | cloud

win - win python } jupyter
mac - anaconda } notebook
linux - pycharm

GIT + WTB - control
versiones + compacity / colaborer
(repositorio ficheros)

desarrollo /
aprender

• ipynb

celdas código

+

celdas texto

~

~

telas fijas...

producción

• PY

|||||

LIBRERÍAS PYTHON

• MATEMÁTICAS, ESTADÍSTICA

[SCIPY — funciones matemáticas
NUMPY — operaciones vectores, matrices

• ESTRUCTURAS de DATOS

PANDAS — datos multivariados
en formato dataframe
& manipulación

• REPRESENTACIÓN GRÁFICA DATOS

└ MATPLOTLIB

└ SEABORN

└ PLOTLY — Gráficos interactivos

• DATA ANALYSIS + MACHINE LEARNING

— SKLEARN (ML TRADICIONAL)

| zoom
| punter...
| inspección

↑ R — software estadístico

IMAGES — scikit-image (skimage)

DEEP LEARNING — | KERAS (recurrent neural networks)
| TENSORFLOW