

HERRAMIENTAS DE ANÁLISIS DE DATOS

Curso breve de técnicas modernas de análisis

De una forma o de otra, es común enfrentarse con datos. Este curso pretende ser una introducción de las diversas herramientas disponibles para su análisis. Se cubrirá programación básica-intermedia en R y en Python, así como buenas prácticas para controlar versiones de código, datos, y modelos.



TEMARIO

2 horas	Introducción Flujos de trabajo y pipelines Lenguajes de programación Editores de texto e IDEs Control de versiones Ambientes productivos Dataframes vs. databases
12 horas	R Editores e IDEs <i>RStudio</i> , <i>VS Code</i> , <i>Jupyter</i> APIs de manejo de datos <i>dplyr</i> , <i>data.table</i> , <i>base</i> Visualización de datos <i>ggplot2</i> , <i>ggforce</i> , <i>ggraph</i> Modelos <i>tidymodels</i> Comunicación de resultados <i>RMarkdown</i> , <i>Shiny</i>
12 horas	Python Editores e IDEs <i>Jupyter</i> , <i>VS Code</i> APIs de manejo de datos <i>pandas</i> Visualización de datos <i>seaborn</i> , <i>matplotlib</i> Modelos <i>scikit-learn</i> Comunicación de resultados <i>Jupyter</i>
5 horas	Control de versiones Código <i>git</i> , <i>GitHub</i> , <i>GitLab</i> Datos <i>DVC</i> Modelos <i>MLflow</i>
A disp. de tiempo	Temas adicionales Makefiles <i>GNU Makefiles</i> Modelos en producción <i>Docker</i> , <i>Flask</i> , <i>Kubernetes</i> , <i>Dask</i> , <i>unittest</i> , <i>testthis</i> , <i>plumber</i> Análisis de redes <i>Neo4j</i>



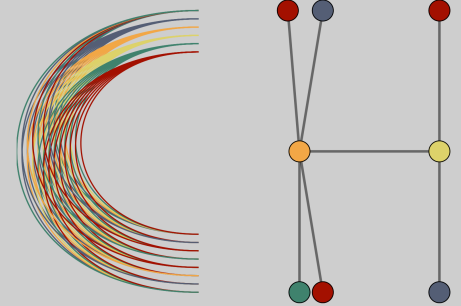
INFORMACIÓN ADICIONAL

Audiencia

Cualquier persona con interés en el trabajo con datos. Cada sesión comenzará desde cero e irá construyendo hacia ejemplos de mayor complejidad.

Requisitos de software

Se usará R, Python, SQL y Bash para demostrar el uso de RStudio, Jupyter Lab/Notebook, MLflow, git, MLflow, DVB, PostgreSQL, GNU Makefiles, Docker, Flask, Kubernetes, Dask. Para el material adicional, se necesitará la instalación de todas las herramientas. Para las sesiones se harán ejemplos rápidos de la mayor parte de ellas, así que es posible solo instalar R, Python, git, RStudio, Jupyter Lab/Notebook y VS Code.



CONTACTO

✉ haro_ca@outlook.com

🐦 [@haro_ca](https://twitter.com/haro_ca)

🐙 github.com/haro-ca

📖 medium.com/haro_ca

OBJETIVO

La intención es proporcionar un lenguaje mínimo para comenzar a practicar, indagar, y profundizar en el trabajo con datos.

Se cubrirán los temas que, a mi consideración, son los más importantes (e interesantes) para llevar a cabo un análisis de datos reproducible y en línea con buenas prácticas del ámbito.

No se pretende, de ninguna forma, que el curso sea suficiente para dar un conocimiento profundo.

En ese sentido, este temario, las sesiones, y el material adicional, funge solo como un compendio y una galería de técnicas.

Este temario fue realizado en R con [page-down](#).

Última actualización 2020-02-16.

Conocimientos previos

No se necesita ningún conocimiento previo de las herramientas de software, pero sí será necesario conocimiento básico de inferencia estadística para las secciones de modelos.

Material

Toda será publicado en github.com/haro-ca/teaching/herramientas_iniciales, tanto código cubierto, como ejercicios y ejemplos adicionales.



REFERENCIAS

Libros

H. Wickham, G. Grolemund. "R for Data Science". <https://r4ds.had.co.nz/>

Hadley Wickam. "Advanced R". <https://adv-r.hadley.nz/>

Yihui Xie, J.J. Allaire, Garrett Grolemund. "R Markdown". <https://bookdown.org/>

Wes McKinney. "Python for data analysis".

Richard McElreath. "Statistical Rethinking".

Yifan Wu. "Is a dataframe just a table?". <http://yifanwu.net/>

Blogs

Vincent Warmerdam. <https://koaning.io> y <https://calmcode.io>

Tom Augspurger. <https://tomaugspurger.github.io/>

Alex Ionnides. <https://alexioannides.com/>

Julia Silge. <https://juliasilge.com/>

Thomas P. Lindersen. <https://www.data-imaginist.com/>

Arturo Gonzales Bencomo.

<https://www.medium.com/arturo102964>