

Laboratorio de Datos

Información general

Primer Cuatrimestre 2024
Turnos tarde y noche

Facultad de Ciencias Exactas y Naturales, UBA

Turno tarde

- Darío Elías (ayudante de primera)
- Nazareno Fallace (ayudante de primera)
- Santiago Laplagne (profesor)

Turno noche

- Gonzalo Barrera Borla (ayudante de primera)
- Nazareno Fallace (ayudante de primera)
- Santiago Laplagne (profesor)

Vacantes

- ¿A todos los presentes les llegó un mail confirmando la vacante?
- Si no pueden cursar, avisen a la brevedad porque hay inscriptos en lista de espera.

Estructura de la materia

- Teórico-práctica
- Laboratorio 1108 (turno tarde) y laboratorio 1111 (turno noche)
- Python
- Para aprobar la materia es necesario:
 - 1 Asistir al 80% de las clases.
 - 2 Trabajar en clase.
 - 3 Aprobar 2 TPs grupales (nota mayor o igual a 4).
 - 4 Aprobar el parcial individual (nota mayor o igual a 6).
- Nota final = $0.6 \times \text{nota parcial} + 0.2 \times \text{nota TP1} + 0.2 \times \text{nota TP2} + x$,
donde

$$x = \begin{cases} 1 & \text{si el promedio de los TPs es mayor o igual a 8,} \\ 0 & \text{si el promedio de los TPs es menor a 8.} \end{cases}$$

Material, información y comunicación

- Campus de la materia
- Github (<https://github.com/fcen-amateur/ldd>)
- Telegram

Fechas importantes

- Entrega del primer TP: martes 21/05
- Entrega del segundo TP: martes 18/06
- Parcial: viernes 14/06
- Recuperatorio del parcial: martes 25/06

¿Qué es la ciencia de datos?

- Se define por las preguntas que se hace: es la ciencia que usa datos para describir, explicar y predecir.
- No se define por las técnicas: no es la ciencia que usa deep learning.

Taxonomía de preguntas

- **Descriptiva:** Resumir características de un data set. Sin interpretación, como atributos de los datos.
- **Exploratorias:** Buscar patrones, tendencias, relaciones entre variables. Sirve para generar hipótesis
- **Inferenciales:** Evaluar hipótesis, respecto a un patrón encontrado en un análisis exploratorio.
- **Predictivas:** Adivinar qué va a suceder, sin importar la causa.
- **Causales:** ¿Cambiar una variable, cambia el valor de otra variable?
- **Mecanísticas:** ¿Cómo ocurre?

Objetivos generales de la materia

¿Cómo trabajar con datos?

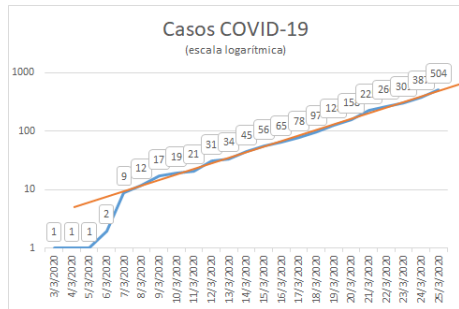
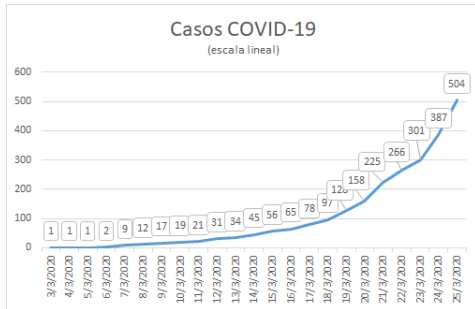
- Organización de datos.
- Visualización, descripción y análisis exploratorio de datos.
- Modelado de datos (modelos explicativos y predictivos).

Programa

Algunos de los temas que trataremos en esta materia, en orden cronológico, serán los siguientes:

- 1 Introudcción a Python y Numpy
- 2 Python: lectura de archivos con Pandas, dataframes
- 3 Introducción a GitHub
- 4 Python: elaboración de gráficos
- 5 Mínimos cuadrados y regresión lineal
- 6 Introducción al modelado: datos de entrenamiento, validación y prueba
- 7 Mínimos cuadrados regularizados
- 8 Limpieza de datos, variables categóricas, datos faltantes
- 9 Normalización de datos
- 10 Aprendizaje no supervisado: clustering
- 11 Valores singulares y componentes principales
- 12 Aplicaciones de componentes principales
- 13 Optimización no lineal - Método del descenso
- 14 Introducción a Redes neuronales

Ejemplo: ajuste de funciones



¡Largamos!

