



---

IIC2115 – Programación como Herramienta para la Ingeniería (I/2024)

## Ejercicio Formativo 2 Capítulo 2

### Aspectos generales

- **Objetivos:** aplicar los contenidos de análisis de datos tabulares con Pandas.
- **Entrega:** lunes 01 de abril a las 17:30 hrs. en repositorio privado y ticket de salida.
- **Formato de entrega:** archivo E2.ipynb con los solicitado, ubicado en la carpeta **C2** del repositorio.

### Descripción del problema

Considere el conjunto de datos almacenado en el archivo `data_E2.csv`, que contiene datos obtenidos a lo largo de los años sobre los niveles de Ozono ( $O_3$ ) y material particulado de 2.5 micrómetros ( $PM_{2.5}$ ). Además de esta información, cada registro está categorizado en tres niveles, en base al riesgo ambiental que presentan las mediciones de  $O_3$  y  $PM_{2.5}$  para la fecha: bajo, medio y alto. En base a toda esta información, complete las misiones indicadas a continuación.

#### Misión 0: aspectos básicos

Para cumplir las misiones de este taller, es fundamental explorar inicialmente el contenido del archivo y familiarizarse con el formato en que está almacenada la información. Para eso, utilice los comandos `describe` y `head` de pandas.

#### Misión 1: limpieza y depuración

Tanto para  $O_3$  como para  $PM_{2.5}$ , el conjunto contiene datos extremos y datos incompletos para algunos días, que fueron generados por motivos desconocidos. Con el fin de facilitar el análisis futuro, deberá **primero**

corregir los datos extremos y luego ajustar los datos faltantes de 2 formas distintas. Para esto último, cree 2 nuevos **DataFrame**, en el primero de ellos complete los datos faltantes con la media, y en el segundo elimínelos.

## **Misión 2: descripción y comparación**

A continuación, para ambos **DataFrame** generados en el ítem anterior y de manera independiente, imprima en una tabla ordenada los siguientes indicadores para  $O_3$  y  $PM_{2.5}$ : media, desviación estándar, máximo, mínimo, Kurtosis. Además, agregue a esta tabla la correlación entre  $O_3$  y  $PM_{2.5}$ .

## **Misión 3: visualización**

Para ambos **DataFrame** obtenidos en el primer ítem y de manera independiente, genere las siguientes visualizaciones:

- Histograma de  $PM_{2.5}$
- Boxplot de  $O_3$  por mes
- Evolución promedio de  $O_3$  y  $PM_{2.5}$  por año.

## **Misión 4: categorización**

En base a todos los análisis realizados anteriormente, proponga e implemente en Python un esquema para asignar un nivel de riesgo medioambiental para cada registro que no tiene esta información. Complete esto para ambos **DataFrame** de manera independiente. Comente y analice los resultados.

**IMPORTANTE:** todas las celdas utilizadas deben estar ejecutadas al momento de entregar el ejercicio, de modo que las salidas generadas sean visibles. En caso de no cumplir esto, su entrega no será considerada.