

Herramientas avanzadas de programación en *Python* para procesamiento de datos

Glosario

- **Columna:** Sinónimo de serie. Estructura básica de organización de datos en *Pandas*. Un Data Frame tiene varias columnas, cada una con un tipo de datos.
- **Fila:** Instancia única de valores para cada set de datos. Cada fila tiene valores distintos por columnas. Un Data Frame tiene múltiples filas.
- **Index:** Identificador por fila. Puede ser un número o texto. No necesariamente es único, aunque algunas funciones de *Pandas* requieren que lo sea.
- **Índice:** Posición de una fila dentro de un Data Frame. Se empiezan a contar desde 0 hasta la cantidad de filas-1. Puede ser distinto al index (por ejemplo, que el index sea el RUT y el índice una posición determinada).
- **NaN:** Abreviación de “*Not a Number*”. En general, *Pandas* rellena valores faltantes con NaN.
- **Tipos de datos en *pandas*:** La librería *Pandas* tiene sus propios tipos de datos. Son equivalentes a los de *Python*, pero tienen nombres distintos y existen nuevos que permiten manejar datos de manera más eficiente. Los más importantes son:
 - **object:** Equivalente al string en *Python*.
 - **int64:** Equivalente a int en *Python*.
 - **float64:** Equivalente a float en *Python*.
 - **bool:** Igual que *Python*.
 - **datetime:** Dato de tipo fecha (no tiene equivalente en *Python*).