

# Actividad 3

David Fernando Rodríguez Arnold  
Universidad de Sonora

17 de febrero de 2021

## Resumen

En esta actividad retomaremos los datos utilizados de la estación climatología que utilizamos anteriormente. Aquí utilizaremos una serie de datos que nos serán útiles para generar una tabla de datos utilizando la biblioteca de **Pandas**. Esto nos permitirá tener un mejor manejo de información, asignando nombre a las columnas, analizando la cantidad de datos totales que estemos manejando, etc.

Para poder hacer algún análisis de los datos, primeramente creamos una copia del trabajo para tener así un **DataFrame** vinculado al original, para así no alterar nada en los datos originales.

## 1. Librería Pandas

Pandas es una librería de Python especializada en el manejo y análisis de estructura de datos, utilizada en la ciencia y en Machine Learning, Pandas es un paquete muy utilizado en Python ya que ofrece una estructura de datos expresiva y flexible que facilita la manipulación y el análisis. Pandas es una librería de código abierto que proporciona una serie de herramientas de alto rendimiento utilizando su estructura de datos. El nombre de Pandas está derivado del término "*Panel Data*".

Un **Data Frame** es la estructura fundamental de Pandas, la cual es una estructura de datos etiquetados bidimensionalmente con columnas. Este DataFrame consta de tres componentes principales: *los datos, el índice y las columnas*.

El índice indica la diferencia en las filas, mientras que los nombres de las columnas indican la diferencia en las columnas. Estos componentes son muy útiles cuando se requiera manipular los datos.

Algunas de las principales características de la librería de Pandas son:

- Define nuevas estructuras de datos basadas en los arrays de la librería NumPy pero con nuevas funcionalidades.
- Permite leer y escribir fácilmente archivos en formato CSV, Excel y bases de datos SQL.
- Permite acceder a los datos mediante índices o nombres para filas y columnas.
- Ofrece métodos para reordenar, dividir y combinar conjuntos de datos.
- Permite trabajar con series temporales.
- Realiza todas estas operaciones de manera muy eficiente.

Pandas dispone de tres estructuras de datos diferentes:

- Series: Estructura de una dimensión.
- DataFrame: Estructura de dos dimensiones (tablas).

- Panel: Estructura de tres dimensiones (cubos).

Estas estructuras se construyen a partir de arrays de la librería NumPy, añadiendo nuevas funcionalidades.