

# Actividad 3

Johana Guadalupe Miranda González

Enero 2021

## Introducción

En esta actividad el objetivo fue dar una introducción a la biblioteca **Pandas** de Python que fue desarrollada por Wes McKinney para manipulación y análisis de datos. En nuestro caso, desarrollaremos actividades en **Google Colab** y allí se encuentra también la biblioteca Pandas como tantas otras.

## Descripción de la actividad

### Paso 1

Lo primero que hicimos fue abrir un nuevo cuaderno de trabajo Jupyter en Google Colab llamado Actividad3, se cargaron las bibliotecas **Numpy** y **Pandas**.

### Paso 2

Con ayuda de la biblioteca pandas leí mi archivo y con las funciones **df.head(n)** y **df.tail()** estudié la estructura de él. Lo que buscamos es quedarnos con un DataFrame con los nombres de las columnas en el primer renglón y el resto de renglones serán sólo datos diarios de la estación meteorológica.

### Paso 3

Enseguida se pide explorar el dataframe recién cargado a la memoria. Para esto, se pide hacer una bitácora anotando o agregando un comentario de las acciones desarrolladas.

- ¿Qué dimensiones tiene tu dataframe? (11794, 5). Es decir, 11794 renglones y 5 columnas.
- ¿Cómo es el contenido de tu dataframe?: Índice de rango es de 11794 enumerado de 0 a 11793 con 5 columnas.
- Los datos originales incluyen la cadena de caracteres 'Nulo', indicando que no hubo datos para esa variable, ese día. Tenemos que reemplazar la palabra nulo con la función: **df.replace()**
- Después habrá que convertir a número flotante o numérico los datos de Precipitación, Evaporación, Temperatura Máxima y Temperatura mínima utilizando la función: **df.to\_numeric()**

- Se puede contabilizar el número de datos faltantes en esas variables mediante la función: **df.isnull().sum()**
- Realiza una estadística básica de las variables numéricas de tu dataframe usando la función: **df.describe()** y haz una interpretación de los resultados, para ver si tienen sentido físico (por ej. valores negativos de precipitación, valores extremos fuera de lo normal, etc)

#### Paso 4

Llevamos a cabo el análisis de la variable Fecha. Pandas maneja las variables tipo Fecha y Tiempo, además ofrece una serie de herramientas para su manipulación.

Lo que sigue es convertir el objeto que se ha leído a una variable que Python comprenda.

Lo primero que hicimos fue crear una copia del DataFrame y utilizamos la función de Pandas **pd.to\_datetime()** para convertir el objeto Fecha a formato de fecha que comprende Python.

Después utilizamos la función **df.dtypes** para verificar que todas las variables son del tipo deseado. Enseguida con la ayuda de las funciones de Pandas **df['Fecha'].dt.year** y **df['Fecha'].dt.month**, crea dos columnas nuevas adicionales **df['Año']** y **df['Mes']**.

## Conclusión

Considero que fue una actividad bastante clara y fácil de realizar. Lo que se me dificultó un poco fueron las funciones específicas con respecto al comportamiento de los datos, de ahí en fuera me gusto. Creo que tuvo un grado de dificultad intermedio gracias a que la ayuda que proporciona el maestro es grande.