

# Actividad 3 de Física Computacional 1

José Antonio Sanabria Vázquez  
Departamento de Física  
Universidad de Sonora

January 28, 2021

## 1 Introducción

Esta actividad fue hecha y desarrollada para poder aprender a manejar una de las herramientas más utilizadas en el mundo de Python ( Pandas ). En la cual tomamos una serie de datos en formato .txt dadas por la página oficial gubernamental CONAGUA, donde se dan a conocer datos en formato RAW las temperaturas, precipitaciones y demás datos diarios desde la década de los 70 en la ciudad de Bahía de Kino ( Datos utilizados para hacer un resumen en la actividad 1 )

## 2 Biblioteca Pandas

En Computación y Ciencia de datos, pandas es una biblioteca de software escrita como extensión de NumPy para manipulación y análisis de datos para el lenguaje de programación Python. En particular, ofrece estructuras de datos y operaciones para manipular tablas numéricas y series temporales. Es un software libre distribuido bajo la licencia BSD versión tres cláusulas. El nombre deriva del término "datos de panel", término de econometría que designa datos que combinan una dimensión temporal con otra dimensión transversal.

Particularmente esta biblioteca es utilizada para poder manipular una gran extensión de datos crudos donde ya sea para poder sacar información de ellos, o tan solo para poder organizarlos en archivos CSV, JSON,ETC.

### Comandos mas utilizados en la practica 3:

- `df.head()` # Enlista los primeros 5 renglones por default. Si se desea imprimir n renglones: `df.head(n)`
  - `df.tail()` # Enlista los últimos 5 renglones por default
  - `df.shape` # Regresa el número de (renglones, columns)
  - `df.info()` # Brinda información del DataFrame: número de renglones, columnas, nombres, tipo de datos, número de entradas no nulas, y otros
  - `df.describe()` # Describe las estadísticas de las variables flotantes.
  - `df['variable categórica'].describe()` # Describe cuantos valores hay, cuantos distintos.
  - `df['variable categórica'].value_counts().head(10)` # Cuenta las ocurrencias de cada variable categórica y muestra las primeras 10.
  - `df.columns` # Imprime los nombres de las columnas
  - `df.columns = ['col1', 'col2', 'col3']` # Asignar mediante una **lista**, nombres a todas las columnas
  - `df.rename(columns={'Old1' : 'New1', 'Old2' : 'New2'}, inplace=True)` # Renombrar algunas columnas
  - `df.drop_duplicates(inplace=True)` # Elimina los renglones duplicados
  - `df.isnull().sum()` # Regresa la suma de número faltantes de cada variable.
  - `df.dropna()` # Elimina los renglones que tengan algún número nulo.
  - `df.dropna(axis=1)` # Elimina las columnas que tengan algún número nulo.
  - `df.mean()` # Calcula el promedio.
  - `df.fillna(x_mean, inplace=True)` # Reemplaza los valores nulos por el valor promedio.
  - `df.corr()` # Muestra la correlación entre variables
- 
- `df['B']` # Selecciona la columna 'B'.
  - `df[['A', 'B']]` # Selecciona las columnas 'A' y 'B'.
  - `df_new = df[['A', 'B']]` # Crea un nuevo DataFrame a partir de 2 columnas de **df**
  - `df['C'] = df['A'] + df['B']` # Crea una nueva columna a partir de sumar dos columnas 'A' y 'B'.
  - `df_copy = df.copy()` # Crea una copia del DataFrame **df**.
  - `df.drop('D', axis=1, inplace=True)` # Elimina la columna 'D' y reescribe a el DataFrame df.
  - `df.loc['2']` # Regresa el renglón de df con índice '2'.
  - `df.loc['2', 'C']` # Regresa un elemento específico: índice = '2', columna = 'C'.
  - `df.iloc[2]` # Regresa el renglón de índice 2
  - `df.loc['2':'4']` # Regresa los renglones '2','3', y '4'.
  - `df.iloc[2:4]` # Regresa los renglones de índices 2 y 3, no incluye el 4.
  - `df[df['B'] == 5.0]` # Filtra los renglones donde se cumple la condición: df['B'] es igual a 5.0.
  - `df[(df['B'] == 5.0) & (df['D'] <= 2.0)]` # Filtra los renglones donde se cumplan ambas condiciones simultáneas.

### 3 Impresiones de la practica 3

1. **¿Qué te pareció?**

Me pareció súper interesante el hecho de poder usar los DataFrame, ya que estos se van a poder manipular de cualquier forma, entonces, me parece una muy buena idea poder ver estos tipos de estructuras en código.

2. **¿Cómo estuvo la carga de trabajo?**

No estuvo tan pesado el trabajo, ya que es muy difícil aprender nuevos comandos, pero cuando uno le empieza a agarrar el rollo ya es mucho mucho mas fácil.

3. **¿Qué se te dificultó más?**

Se me dificultó mucho darle la estructura al DataFrame, ya que venia como serie de palabras y no sabía como separar, pero todo sale con esfuerzo.

4. **¿Qué recomendarías para mejorar la segunda Actividad?**

Nada, me pareció muy didáctica y la explicaciones fueron suficientes.

5. **¿Que grado de complejidad le asignarías a esta Actividad? (Bajo, Intermedio, Avanzado)**

Intermedio, ya que como siempre es difícil acostumbrarse al principio como todo nuevo comando o lenguaje.