Actividad 3
Introducción a Pandas

Elizabeth Torres Torrecillas
Departamento de Física
Universidad de Sonora

January 29, 2021

1 Introducción

La actividad 3 tiene como fin el utilizar e irnos familiarizando con la biblioteca Pandas del lenguaje Python. Para ello, realizamos una actividad en Google Colab donde analizamos datos que ya habíamos obtenido y analizado en la actividad 1. Dicha serie de datos se descargaron en formato .txt por medio de la página web gubernamental de la CONAGUA
Siendo esta biblioteca, la tercera utilizada a lo largo del curso, los datos climatológicos de la estación meteorológica seleccionada, El Fresnal, y almacenada en el repositorio de Física Computacional en Github.

2 Pandas

Pandas es una biblioteca de software escrita como extensión de NumPy para manipulación y análisis de datos para el lenguaje de programación Python.
Esta ofrece estructuras de datos y operaciones para manipular tablas numéricas y series temporales. Los principales tipos de datos que pueden representarse son los datos tabulares con columnas de tipo heterogéneo con etiquetas en columnas y filas, como también las series temporales.
Pandas proporciona herramientas que permiten:

- Leer y escribir datos en diferentes formatos: CSV, Microsoft Excel, bases SQL y formato HDF5
- Seleccionar y filtrar tablas de datos
- Fusionar y unir datos
- Transformar datos aplicando funciones tanto en global como por ventanas
- Manipulación de series temporales

- Graficar

En pandas existen tres tipos básicos de objetos:

1. Series (listas, 1D)
2. DataFrame (tablas, 2D)
3. Panels (tablas 3D)

3 Actividades realizadas

Para poder realizar el presente trabajo, primero se subió el archivo con los datos climatológicos de la estación meteorológica de El Fresnal por medio de un documento .csv a Github, después se seleccionó la opción "raw" y se copió el link que hace referencia al documento.
Después se abrió un nuevo cuaderno de trabajo Jupyter en Google Colab, llamado Actividad3.Después se declararon las bibliotecas a utilizar, en mi caso fueron NumPy, Pandas y Matplotlib.
Después de ello, se definió un DataFrame con ayuda de una función de pandas, siendo esta quien nos ayudará a leer el archivo .csv que tenemos en Github, aquí es donde utilizamos el link de referencia y lo introducimos en dicha función.

Una vez que se leyó el archivo de datos, estudiamos la estructura del archivo original para determinar que renglones tienen información sobre los nombres de las columnas y en qué número de renglón comienzan los datos. Creo importante mencionar que este se introdujo sin modificar datos, en su estado original.
Después, ya que tenemos un DataFrame con los nombres de las columnas en el primer renglón, entonces el resto serán los datos diarios de la estación.
Ya con ello, realizamos varias funciones para ir conociendo el DataFrame, por ejemplo imprimimos el encabezado, el final, el contenido, las dimensiones del data frame, lo cual se nos brinda como (renglones, columnas).
Notamos que al imprimir los datos originales, en ellos se incluye una cadena de caracteres presentados como 'Nulo', lo cual indica que no hubo registro de el dato que lo presenta. Por lo que decidimos, reemplazarlo por un espacio en blanco, así haciendo una limpia de datos. Como también contabilizamos dicho número de datos faltantes.

A continuación, convertimos a número flotante a las variables que se tienen para las columnas de tipo de datos presentados y realizamos una cuadro de estadística con datos básicos de las variables numéricas presentes en el DataFrame.
Proseguimos a realizar una interpretación de los resultados, por ejemplo podemos notar que la temperatura mínima registrada es de 5°C y la máxima de 45°C, como también que la temperatura esperada en El Fresnal es de 28°C, además que tenemos una cantidad de alrededor de doce mil datos de evaporación sin registro alguno, por lo que no contamos con el conocimiento de cómo se comporta este en dicha estación. Otra interpretación interesante es que cuando llueve, la

cantidad de agua es poca, por lo que se espera alrededor de 5 mm.
El útlimo ejercicio se basó en el análisis de las variable 'Fecha', y 'Tiempo'. Datos con los que ya contabamos, sin embargo procedimos a convertirlos de objeto a una variable que Python pudiera comprender.

4 Comentarios personales

La actividad desarrollada me pareció interesante, ya que analizamos datos e hicimos que Python los asimilaran como el tipo de datos que puede reconocer. En principio me pareció que no tenía tanta dificultad sobre todo porque ahora sentí una explicación más sencilla y profunda por medio de profesor, lo cual fue de bastante ayuda para la manipulación de los datos. Sin embargo, en el último paso, siendo este el número cuatro, me pareció que me perdí un poco y fue donde más batalle para realizarlo.
Haciendo uso de la función para la variable fecha. Me pareció entretenida, ya que conocí una biblioteca más con bastante utilidad y más porque sigo aprendiendo un poco más de Python, como utilizarlo para aprovechar al máximo sus herramientas.
La carga de trabajo esta semana siento que fue un poco menor, supongo que también porque desde el primer día me fijé objetivos de avanzar lo más que pudiera y porque como mencioné antes, sentí mucho apoyo del profesor. Por ello, le asignaría un grado de dificultad intermedio, me agradó la actividad.