

Actividad 3

Física Computacional

Introducción a Pandas.

María Fernanda Vences Mendoza

29 de enero del 2021

1 Introducción

En esta actividad con la base de datos de la estación climatológica que se realizó en la actividad , analizamos como asignar los nombres de las columnas, el tamaño y que información es la que tiene. Para poder hacer algún análisis estadístico se realizó una copia de trabajo para así tener un **DataFrame** asegurado al original y no tener alguna alteración a los datos originales.

Al momento de realizar una copia de trabajo se observa que tenemos unos textos del tipo *Nulo* en la columna de Evaporación, para cambiar estos resultados reemplazamos donde dice *Nulo* por un espacio en blanco.

Al ver detalladamente, tenemos variables ambientales y tenemos que modificar a números flotantes, esto para que sea más sencillo y que *Python* pueda leer los datos. Al hacer esto se puede realizar un análisis estadístico de las variables.

Al analizar los datos, nos falta usar la variable *Fecha*, ya que puede tener mucha información condensada, para esto, se hace otra copia del **DataFrame** así tener un acceso más fácil sin modificar el **DataFrame** original y tenerlo separado. Con ayuda de *datetime64* nos muestra como clasificar la variable *Fecha* y así sabe que es una variable de tiempo. Para poder separar el mes, fecha y año, podemos clasificar los datos por año y mes para al final obtener una forma más específica de tener los datos acerca de la estación climatológica.

Creo que es una manera súper sencilla de ordenar los datos y tenerlos en un formato más accesible para lo que vamos hacer después. No sabía que los datos se podían ordenar de una manera más facil y vaya la palabra ordenada. Haciendo esto nos ayudará a realizar una clasificacion de datos ya sea saber que meses son los que cae más lluvia y cosas así. Poder realizar las gráficas que venían en el documento y haciendo esto es más facil saber cómo le hacen para obtener los graficos.

Es relativamente sencillo con el uso de los manuales proporcionados y las notas del profesor fue un poco más sencillo y no tan batalloso como pensé que iba a ser.

2 Biblioteca Pandas

Pandas es una biblioteca que forma parte de *Numpy* para poder manipular y hacer análisis de datos para el lenguaje de programación *Python*. Ofrece estructuras de datos y operaciones para manipular tablas numéricas, series temporales y hasta poder leer datos de un archivo externo.

Características de la biblioteca *Pandas*:

- El tipo de datos que “entran” o se almacenan se llaman **DataFrame**.
- Tiene herramientas para leer y escribir datos entre estructuras de datos en memoria y formato de archivos variados.
- Hace una alineación de datos y manejo de datos que puedan faltar, reacomoda y divide conjuntos de datos, divide en base a etiquetas, ordena y divide horizontal grandes conjuntos de datos, agrega y elimina columnas de los datos. Proporciona una herramienta que permite nombrar columnas y guardar copias de los datos a trabajar.
- Se realiza varias cadenas de operaciones, dividir, aplicar y combinar sobre los datos ya sea la mezcla y juntar datos.
- Ordena en secuencia de ejes para trabajar con datos de varias y menores dimensiones, funciona para las series de tiempo, genera rangos de fecha y proporciona datos estadísticos.

La biblioteca *Pandas* es una herramienta eficaz que permite un uso múltiple, lo cual es perfecta para el uso de datos. Aparte es sencilla y apta para personas poco expertas en *Python*.

3 Retroalimentación

Estuvo interesante esta actividad ya que nunca imaginé que se podía hacer tantas cosas con los datos, como facilitar la forma que un lenguaje como *Python* pueda leer tantos datos y tener la función de organizar los **DataFrame** de una manera que sea fácil de recordar.

Esta es una actividad de nivel intermedio ya que se tiene lo visto previamente pero agregar la lectura de datos, puede ser un poco batalloso al momento de querer hacer algo. El desarrollo estuvo bien no me aburrí nada, solo al momento de escribir el reporte, pero creo que esto me ayuda a mejorar y saber que estoy haciendo en mi programa. Escribir el reporte también me apoya como una retroalimentación a lo que vemos en la semana y me ayuda a no olvidar lo que estamos viendo al igual que hace que analice más mi código y cada función que usamos para no estar tan perdida en un futuro.