
Actividad 3 (Análisis de datos) - Semana 1



Gustavo Alberto Medina Ferrer¹

A219223438¹

Numero(s):(81) 8309 6131¹

gustavomedinaferrer@gmail.com¹

Universidad de Sonora

Licenciatura en Física

Física Computacional

28/01/2021

Resumen

En este documento se hara una recapitulizacion sobre la actividad 3 (primera semana) de la materia de Fisica Copmutacional, impartida en la Universidad de Sonora por el profesor C. Lizarraga.

1. INSTRUCCIONES

En la actividad se nos pidio hacer varias cosas, sin embargo, tenia como centro la biblioteca *Pandas*, la cual es usada para analizar datos, que es lo que se comenzo a hacer esta semana. El tipo de instrucciones que se nos dieron fueron:

- Crear dataframes y dentro de ellos leer el archivo de texto de los datos de la estacion meteorologica que elegimos.
- Crear mas dataframes a partir del inicial y comenzar a manipular los datos (limpiarlos, acomodarlos, ponerles el tipo de dato correcto, etc.)

Y eso es lo que engloba generalmente la primera semana de este trabajo.

2. ACTIVIDAD

Primero, dando una descripcion sobre lo que es la biblioteca *Pandas*:

Pandas es una biblioteca orientada para el analisis de datos. Es una extension de *NumPy* y usualmente se utiliza para analizar tablas de datos y series de tiempo.

Ahora, como sabemos, la actividad en la que estamos trabajando es acerca de *series de tiempo*, por lo que no es ninguna sorpresa que estemos usando *Pandas* para analizar los datos que tenemos. Los datos se tomaron de una estacion meteorologica a la eleccion de cada estudiante (en este caso, elegi la estacion meteorologica 26068, ubicada en la presa Alvaro Obregon, en el municipio de Cajeme), de la cual obtendriamos los datos que ha recaudado de la pagina oficial de CONAGUA.

Los datos son:

- Fecha
- Precipitacion
- Evaporacion

- Temperatura maxima
- Temperatura minima

Despues, comenzariamos a usar Python para analizar los datos. Sin embargo, esta primera semana solo "*limpiamos*" los datos. Cabe mencionar que los datos son diarios, por lo que son un gran numero de ellos, ya que la mayoría de las estaciones lleva mas de 50 años corriendo.

Lo que se realizo en esta semana es un analisis de los datos, leyendo el archivo y guardandolo en un dataframe, el cual copiariamos a otro con el cual trabajariamos, para asi dejar intacto el original. Despues, comenzamos a analizar los datos que teniamos, viendo las dimensiones del dataframe para saber cuantos datos teniamos, analizando cuantos datos estaban vacios y sustituyendo estos datos vacios por un espacio realmente vacio, en vez de solo 'Nulo'.

2.1. Analisis de datos:

Tras esto, realizamos un analisis estadistico inicial, el cual (personalmente) dio el siguiente resultado:

	PRECIP	EVAP	TMax	TMin
count	24364.000000	24150.00000	24359.000000	24359.000000
mean	1.120149	7.05853	33.732982	18.278041
std	5.962872	3.14731	6.025166	5.906540
min	0.000000	0.10000	10.000000	-2.000000
25%	0.000000	4.60000	29.500000	13.500000
50%	0.000000	6.70000	34.500000	18.000000
75%	0.000000	9.40000	38.500000	24.000000
max	137.500000	18.00000	48.000000	32.000000

Estos resultados fueron comentados en el código, sin embargo, haciendo un analisis breve, podemos observar que los datos no son tan sorprendentes, ya que las temperaturas son las que se dan en esta region (hablando desde un conocimiento empirico), a excepcion de la temperatura de $-2^{\circ}C$, ya que en la ciudad de Obregon nunca he experimentado una temperatura por debajo de $0^{\circ}C$, sin embargo, no es raro que se lleguen a tener temperaturas de $6^{\circ}C$ o alrededor de esta, por lo que tampoco es algo increíblemente inusual.

Podemos ver que la precipitacion maxima es de 137.5mm, la cual es mucha, pero puede atribuirse a alguna tormenta tropical o huracan (quisiera buscar el dato en el dataframe y luego investigar si habia alguno de los dos en esa fecha, para corroborar). Tambien, podemos ver que la evaporacion no es mucha, la maxima siendo de 18mm evaporados.

2.2. Cambio de datos:

Como ultimo cambio a los datos, cambiamos tanto los datos numericos a *realmente numericos*, es decir, valores flotantes en Python y valores de fecha, usando funciones de Pandas.

3. CONCLUSIONES

Como conclusion, esta actividad fue una muy grata, ya que pienso que aprendi bastante sobre la biblioteca *Pandas*, sin embargo no me siento satisfecho, ya que por distracciones personales, no puse la atencion que deberia esta semana en clases, lo que hizo que al momento de hacer la actividad tuviera que investigar casi todo acerca de la actividad, y aun asi me quedaron algunas dudas que tendre que saciar buscando e investigando mas acerca de algunos comandos de *Pandas*. En general, me gusta hacia donde va la actividad.

