

Reporte de la Actividad 8

Uniando DataFrames en Pandas

Luis Fernando Duarte Gonzalí
Universidad de Sonora
Física Computacional

Abril del 2019

1. Introducción

En esta actividad trabajaremos con 2 conjuntos de datos: Los datos meteorológicos de la estación de Nogal que ya se trabajó en la actividad 7. El segundo conjunto de datos se refieren a datos del suelo. Ambos conjuntos de datos coinciden en el periodo 2009, que será el periodo que nos interesa estudiar. El segundo conjunto de datos son sobre variables medidas por sensores bajo el suelo, gestionados y almacenados por un segundo data logger.

Nos interesa crear en los datos del suelo una nueva columna con una variable temporal similar a la que se tiene en los datos meteorológicos, para poder incorporar un conjunto de variables nuevas. Las variables que nos interesan estudiar son los relacionados con las lecturas de temperatura del suelo Tsuelo_10cm, Suelo_20cm, ..., Tsuelo_100cm en 8 profundidades distintas. Adicionalmente se desea incorporar la temperatura del aire del dataframe de datos meteorológicos.

2. Análisis de Datos

Primero tenemos que cargar a nuestro código las librerías que vamos a necesitar durante la actividad.

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import matplotlib.dates as mdates
import datetime
```

Leemos el archivo de la manera que hemos estado usando:

```
df1 = pd.DataFrame( pd.read_csv("MeteoN.csv", engine="python" ) )
df1.head()
```

Se siguió el mismo procedimiento que en la actividad 7, pero como ahora uniremos ambos dataframes en uno, se creó un formato de fecha.

Leemos el segundo archivo que vamos a utilizar:

```
#Leemos el segundo archivo sobre los datos del suelo
df2 = pd.DataFrame( pd.read_csv("SoilN.csv", engine="python" ) )
df2.head()
```

Filtramos los datos que nos interesan del segundo archivo

```
df2 = df2.filter(['2 Year_RTM L', '3 Day_RTM L', '4 Hour_Minute_RTM L', 'Tsuelo_10cm', 'Tsuelo_20cm', 'Tsuelo_30cm',  
                 'Tsuelo_40cm', 'Tsuelo_55cm', 'Tsuelo_70cm', 'Tsuelo_85cm', 'Tsuelo_100cm'], axis=1)  
df2.head()
```

Creamos dos arreglos para separar las horas y minutos. Convertimos a tipo string la columna de minutos y horas.

```
df2['4 Hour_Minute_RTM L'] = df2['4 Hour_Minute_RTM L'].astype(str)
```

Extraemos las horas y minutos de esa columna por separado:

```
hora=[]  
minuto=[]  
  
for i in range(0, len(df2)):  
    #Primero revisaremos si contiene 4 caracteres.  
    if (len(str(df2['4 Hour_Minute_RTM L'][i]))==4):  
        #Si resulta ser de 4 caracteres, revisaremos el caso en que la hora sea 2400.  
        if (str(df2['4 Hour_Minute_RTM L'][i])[0:2]=='24'):  
            hora.append('00')  
            minuto.append('00')  
        else:  
            hora.append(str(df2['4 Hour_Minute_RTM L'][i])[0:2])  
            minuto.append(str(df2['4 Hour_Minute_RTM L'][i])[2:4])  
    #Ahora revisamos el caso de que contenga 3 caracteres.  
    elif (len(str(df2['4 Hour_Minute_RTM L'][i]))==3):  
        hora.append(str(df2['4 Hour_Minute_RTM L'][i])[0:1])  
        minuto.append(str(df2['4 Hour_Minute_RTM L'][i])[1:3])  
    #Por último revisaremos el caso de 2 caracteres.  
    elif (len(str(df2['4 Hour_Minute_RTM L'][i]))==2):  
        hora.append('00')  
        minuto.append(str(df2['4 Hour_Minute_RTM L'][i])[0:2])  
  
    #Arreglo para los días  
    dias=[df2['3 Day_RTM L'][i] for i in range(0,len(df2))]
```

Creamos un DataFrame nuevo con los días, horas y minutos.

```
d = {'dias': dias, 'hora': hora, 'minuto':minuto}  
df_fechas = pd.DataFrame(data=d)  
df_fechas.head()
```

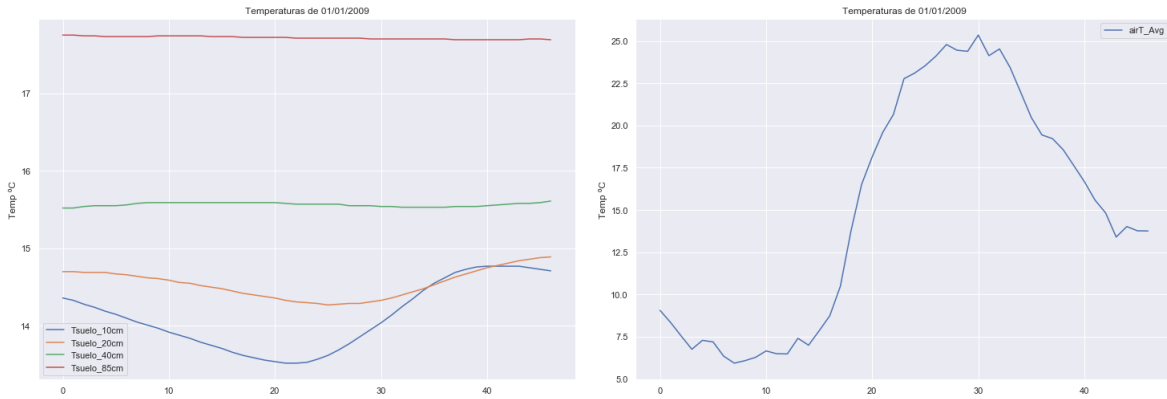
Filtrados los datos, creando el formato de fecha, agregando un día cuando se marcaran las 0 horas, después se unieron ambos DataFrames en uno solo:

```
df3 = pd.merge(df1, df2, on=['FECHA'])
```

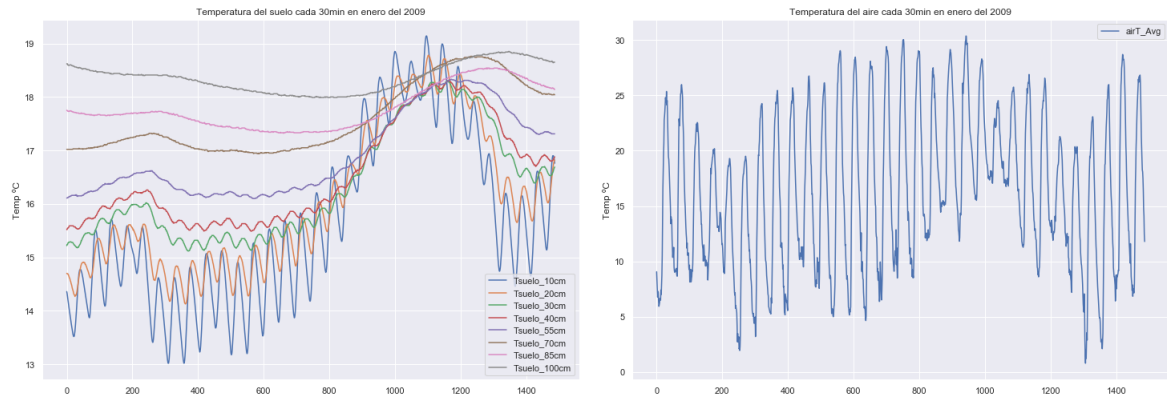
Después de todo eso, se empezaron a generar las gráficas que buscamos desde el principio para mostrar el análisis de datos completo.

2.1. Resultados

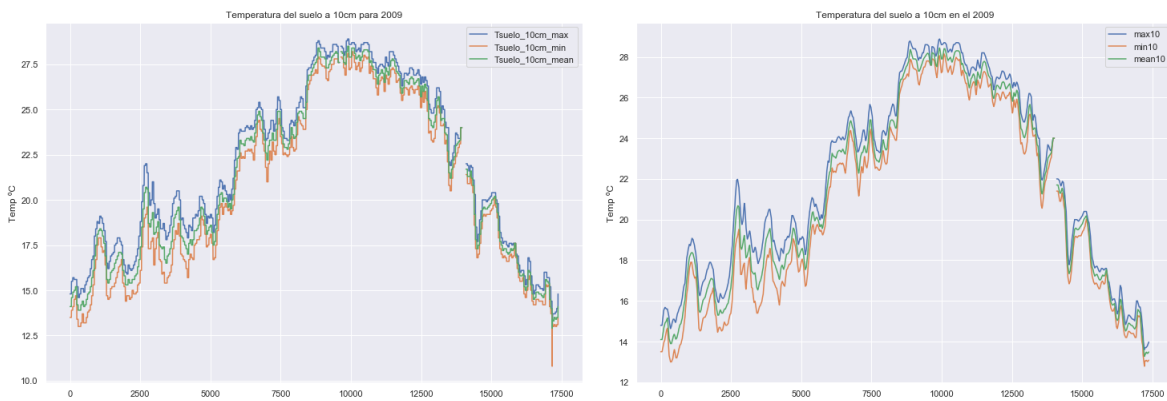
En las siguientes gráficas se muestran las temperaturas de un día (1 de Enero de 2009) del suelo y del aire.

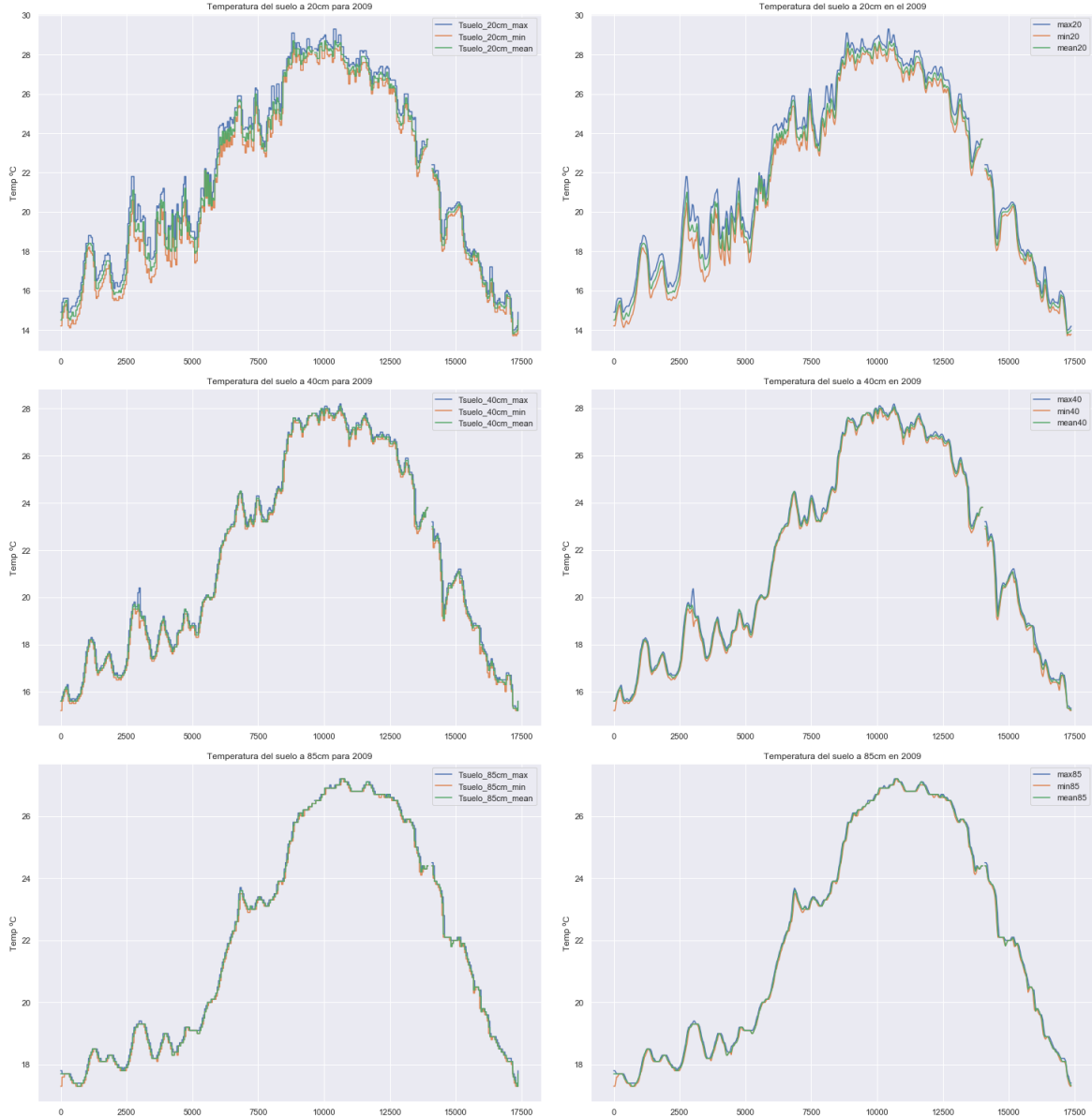


Temperaturas del suelo y del Aire.



En las gráficas que se muestran a continuación se muestra la temperatura del aire y del suelo en diferentes profundidades del mismo. A lado de cada gráfica se muestra las mismas gráficas usando el método de rolling mean, suavizando la evolución temporal de cada una.





3. Conclusión

Al momento de observar las gráficas podemos notar el cambio de temperatura mientras más profundo en el suelo estamos, va de menor a mayor igual que la profundidad, además es más inestable en la superficie, mientras que a cierta distancia hay menos variaciones. Este efecto también se puede observar en las gráficas de cada 30 minutos, donde vemos que la temperatura es más caliente entre más profundo en el suelo estamos tomando las mediciones.

Al observar las gráficas de promedio, máximo y mínimo de temperaturas se hace evidente lo que se menciona anteriormente.

En sí la práctica tuvo su punto de mayor dificultad a la hora de combinar dos dataframes en uno para poder trabajar con sus datos en conjunto.

4. Bibliografía

- Seaborn: statistical data visualization. (2018). Consultado en Abril del 2019, de Seaborn. Sitio web:
<https://seaborn.pydata.org/>
- Python Strings, Functions and Examples (2019). Consultado en Abril del 2019, de Tech Beamers
Sitio web:
<https://www.techbeamers.com/python-strings-functions-and-examples/>
- Merge, join, and concatenate. Consultado en Abril del 2019, de PandasPydata.org. Sitio web:
https://pandas.pydata.org/pandas-docs/stable/user_guide/merging.html