
Uniendo dataframes en Pandas

Rolando Salazar

¹ Universidad de Sonora. Hermosillo, Sonora.

5 de mayo de 2019

Resumen de la actividad 8 donde se unen dataframes con la biblioteca Pandas y se crean gráficas de rolling mean.

1. Introducción

En esta actividad se trabaja con dos conjuntos de datos los cuales pertenecen a una estación de un Nogal y que coinciden en el año 2009. El primer conjunto es el que se trabajó en la actividad 7, donde se encuentran datos meteorológicos, mientras que en el segundo conjunto tenemos variables medidas por sensores en el subsuelo del lugar en cuestión.

La parte fundamental de la actividad es la unión de ambos conjuntos de datos en un solo DataFrame mediante una variable temporal similar, para poder juntar las variables de interés. En este caso, nos interesa realizar gráficas de las variables Tsuelo_10cm, Tsuelo_20cm, Tsuelo_30cm, Tsuelo_40cm, Tsuelo_55cm, Tsuelo_70cm, Tsuelo_80cm, Tsuelo_100cm y airT_Avg.

1.1. Objetivos específicos

- Seleccionar un día de Enero y graficar la temperatura del aire y las temperaturas del subsuelo a 10, 20, 40 y 85 centímetros.
- Realizar una gráfica de temperaturas T_max, T_min y T_promedio diarias para el año completo de 2009.
- Calcular el promedio cada 30 minutos durante los días del mes de Enero de la temperatura del aire y de las 8 temperaturas del subsuelo, para posteriormente graficar las variaciones cada 24 horas.
- Introducir el concepto de rolling mean (promedio móvil) como método para suavizar la evolución temporal de una serie de tiempo y reproducir las gráficas suavizadas.

2. Desarrollo de la actividad 8

Para llevar a cabo todas las gráficas que se piden, fue necesario unir los dos conjuntos de datos. Para ello, se utiliza el siguiente comando

```
df = pd.merge(Nogal, Suelo, on = ['DATETIME'])
```

donde Nogal es el DataFrame de los datos meteorológicos del nogal, Suelo es el DataFrame de los datos del subsuelo del nogal y 'DATETIME' es una variable previamente creada y colocada en ambos DataFrames para que Pandas pueda usarla como referencia para unir ambos conjuntos. Todo esto después de filtrar las variables de interés en cada uno de los DataFrames.

2.1. Primer objetivo específico

Para la realización de estas gráficas, se filtran los datos del día 26 de Enero (día seleccionado por ser el día de mi cumpleaños). Los bloques de comandos no se muestran aquí por ser un poco extensos (al igual que en los demás objetivos, solo se mostrarán los resultados gráficos). Los resultados fueron los siguientes.

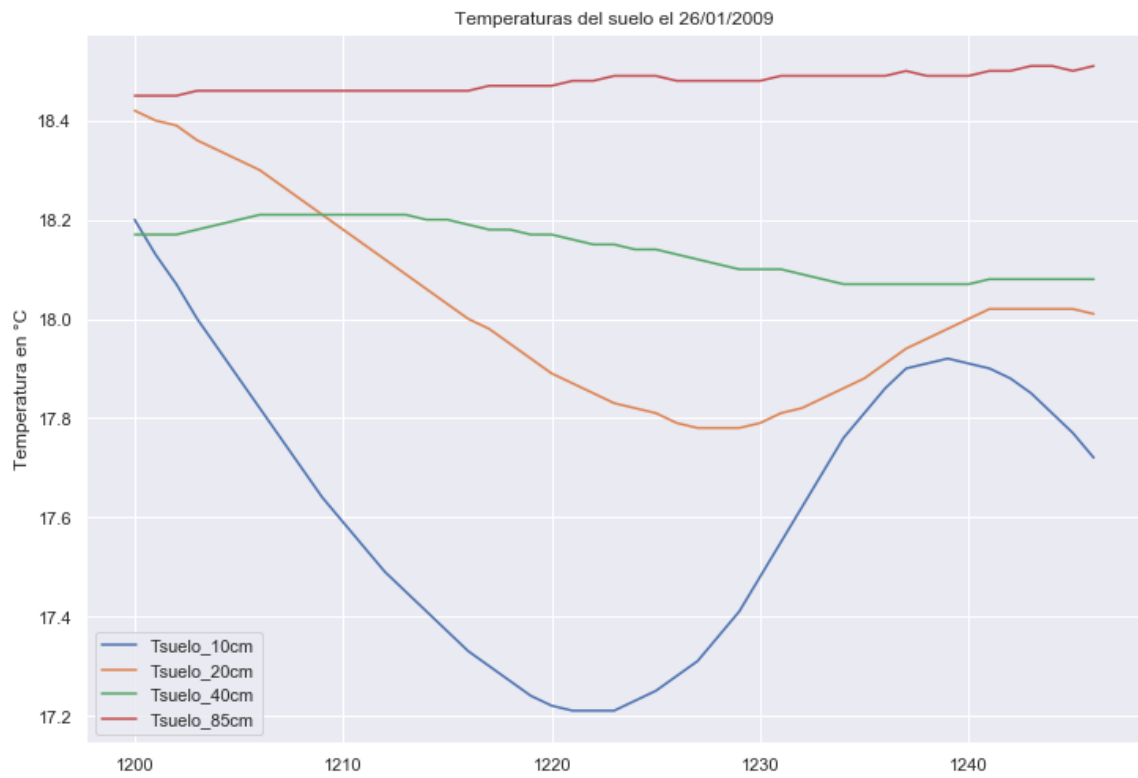


Figura 1: Temperaturas del subsuelo el 26 de Enero de 2009.

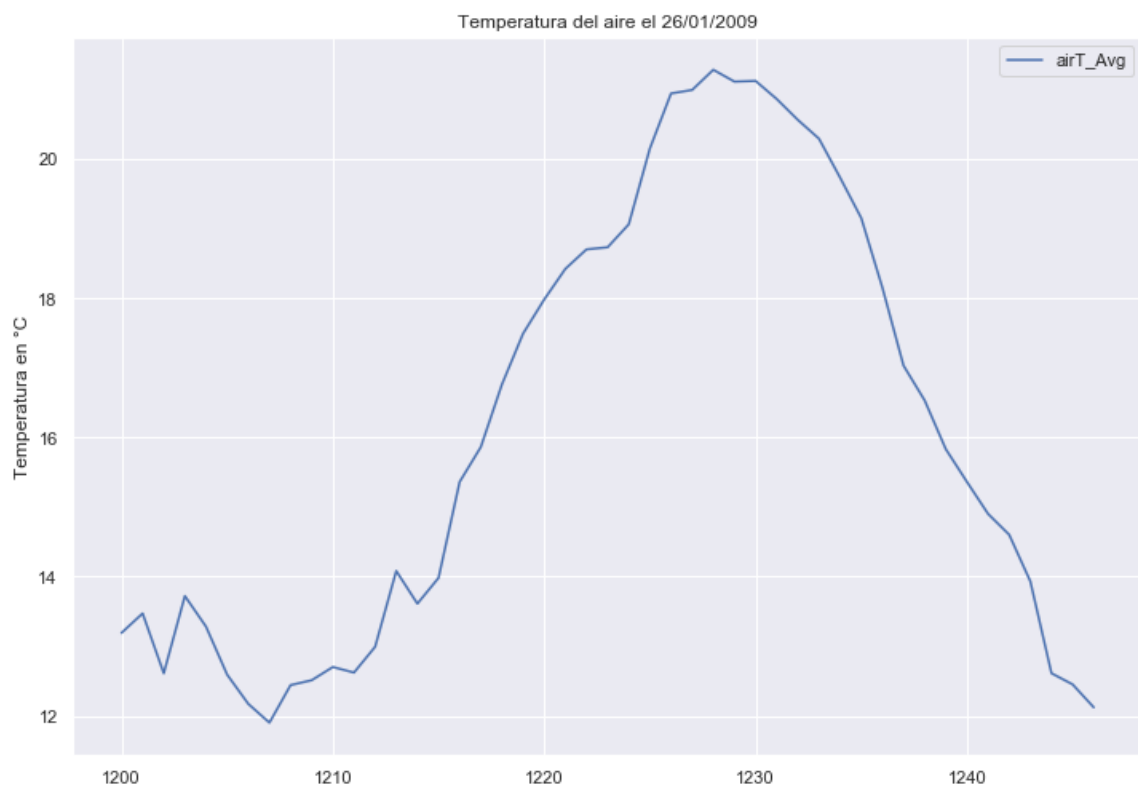


Figura 2: Temperaturas del aire el 26 de Enero de 2009.

2.2. Segundo objetivo específico

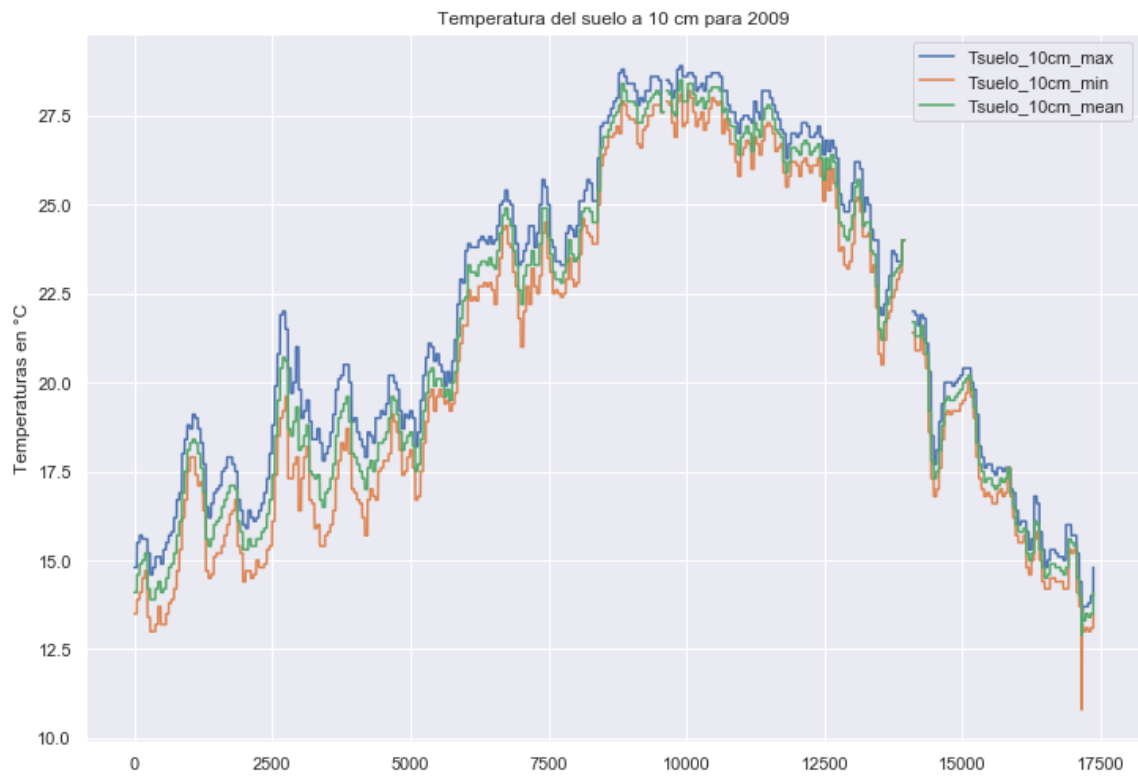


Figura 3: Temperaturas máxima, mínima y promedio diarias a los 10 centímetros bajo suelo del 2009.

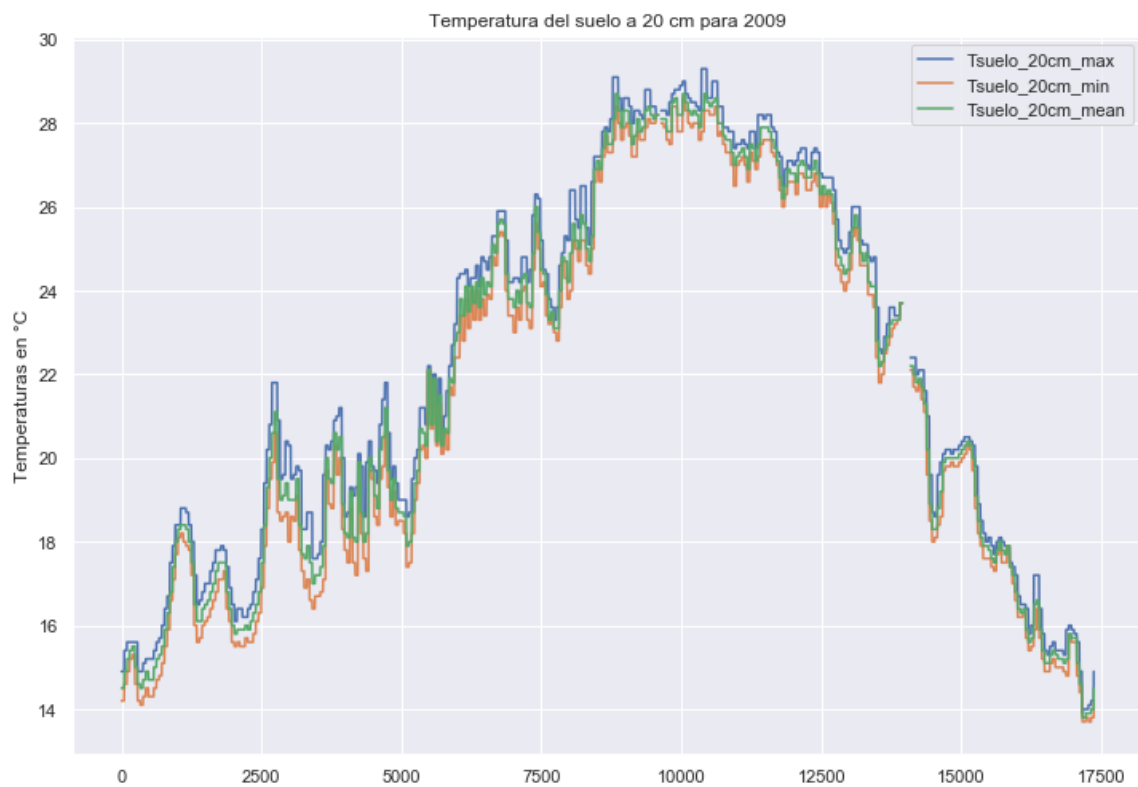


Figura 4: Temperaturas máxima, mínima y promedio diarias a los 20 centímetros bajo suelo del 2009.

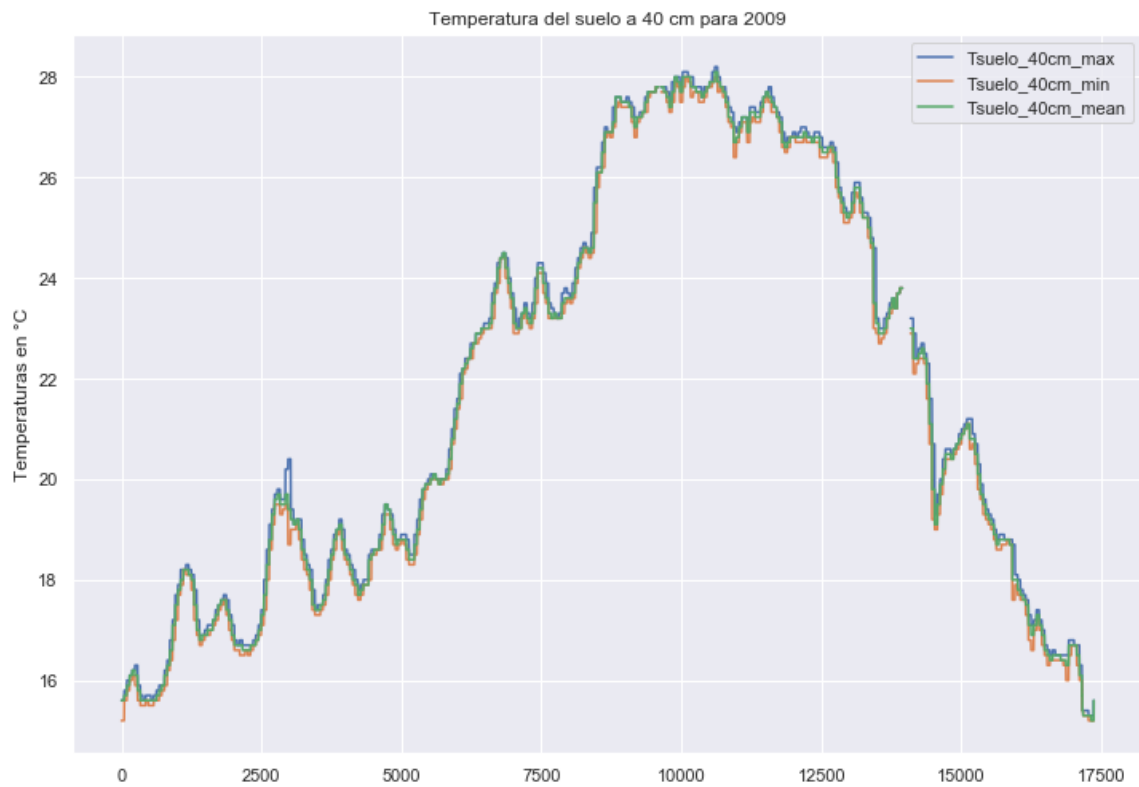


Figura 5: Temperaturas máxima, mínima y promedio diarias a los 40 centímetros bajo suelo del 2009.

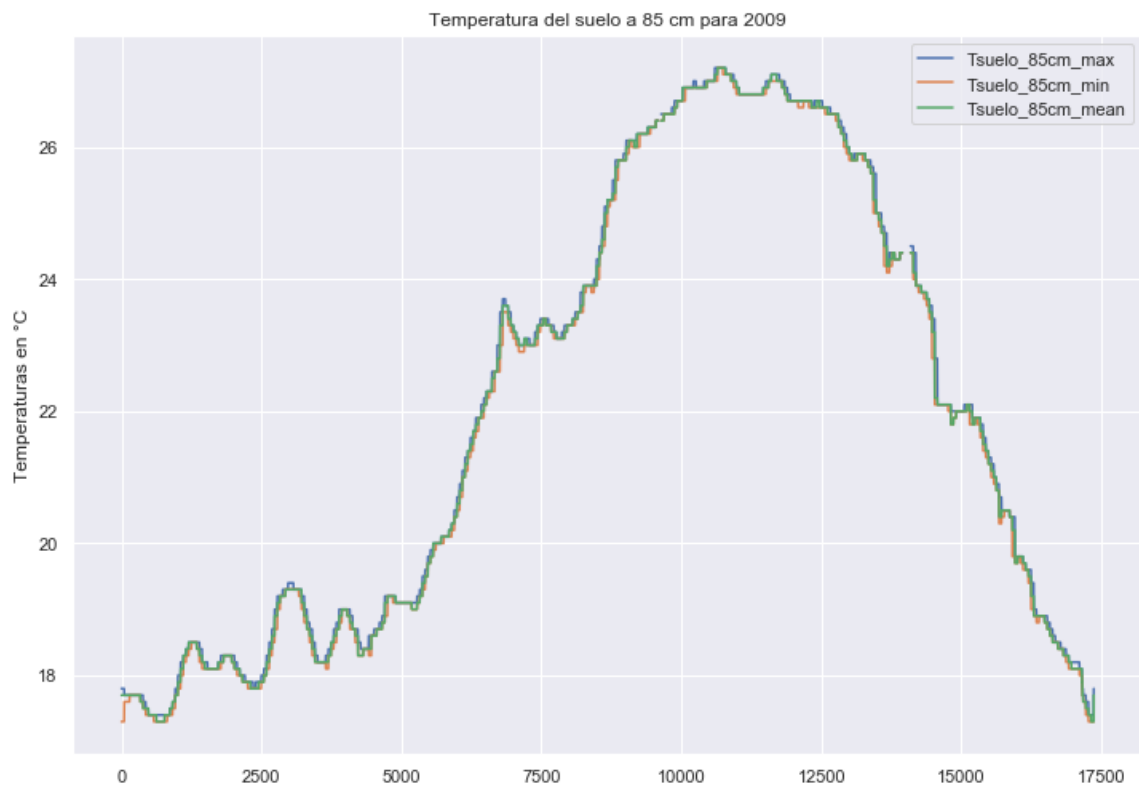


Figura 6: Temperaturas máxima, mínima y promedio diarias a los 85 centímetros bajo suelo del 2009.

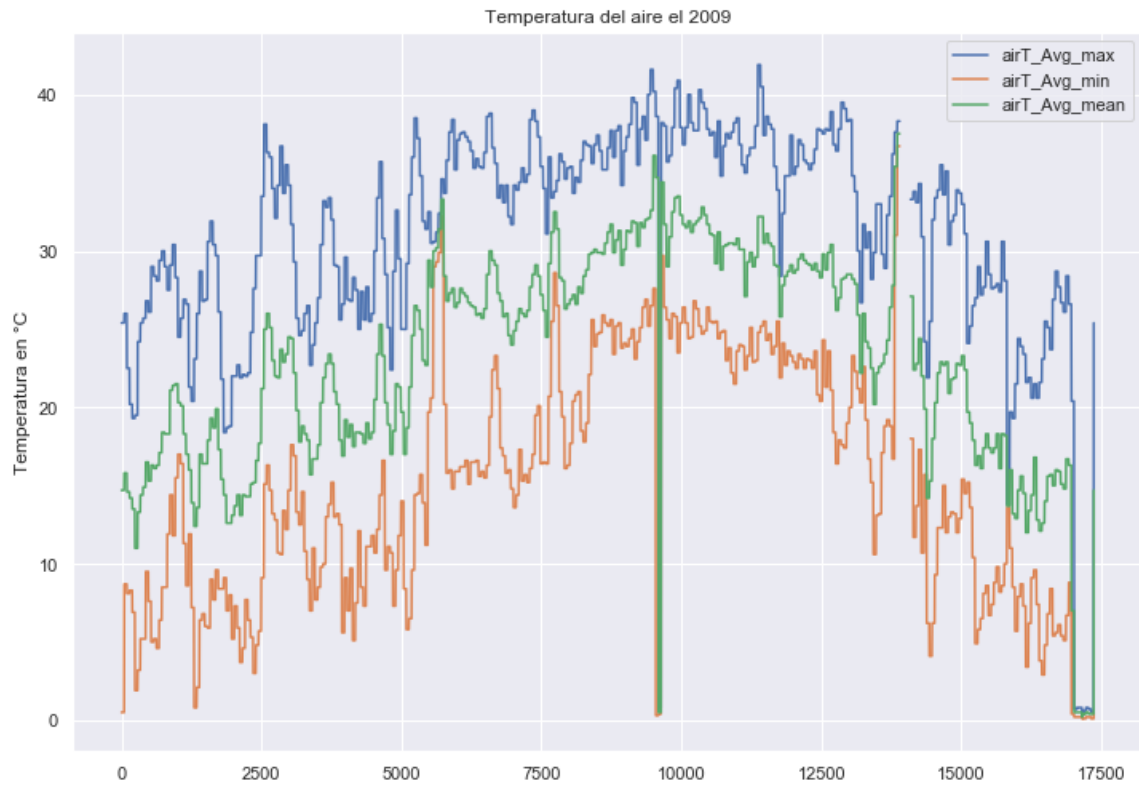


Figura 7: Temperaturas máxima, mínima y promedio diarias del aire del 2009.

2.3. Tercer objetivo específico

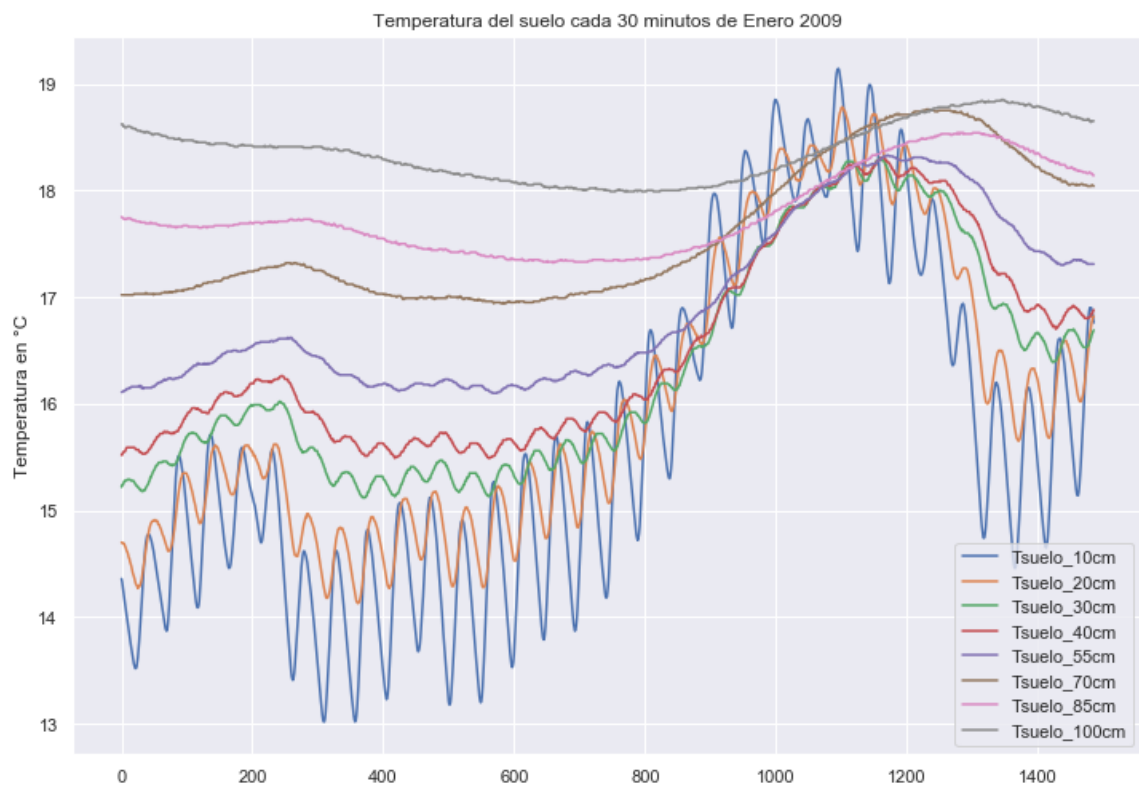


Figura 8: Temperaturas promedio cada 30 minutos en todos los niveles de subsuelo del mes de Enero de 2009.

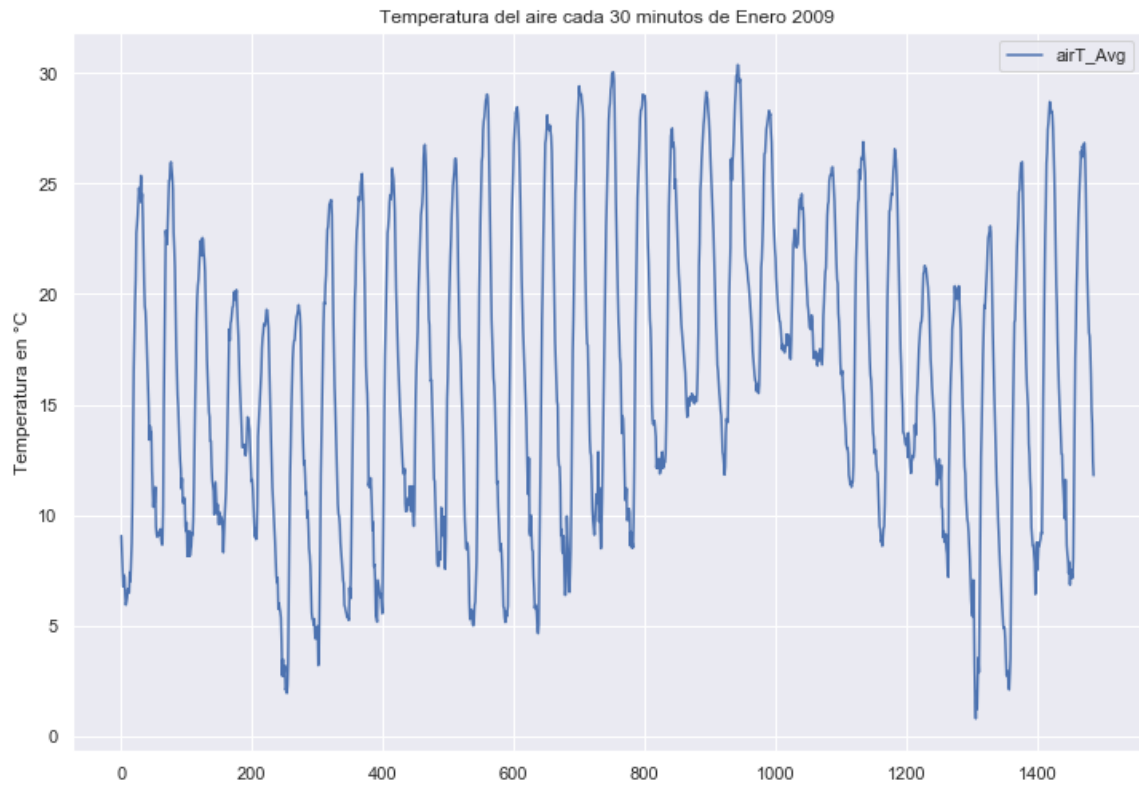


Figura 9: Temperatura promedio cada 30 minutos del aire del mes de Enero de 2009.

2.4. Cuarto objetivo específico

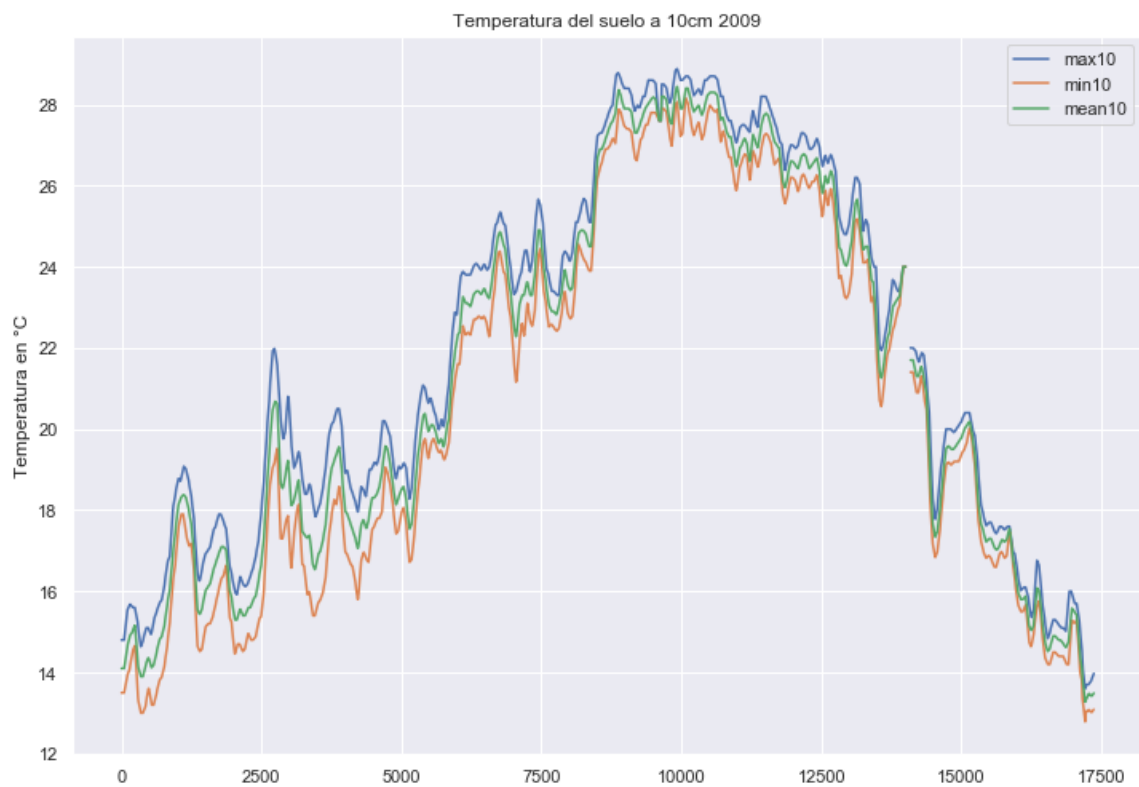


Figura 10: Temperaturas máxima, mínima y promedio diarias a los 10 centímetros bajo suelo usando rolling mean del 2009.

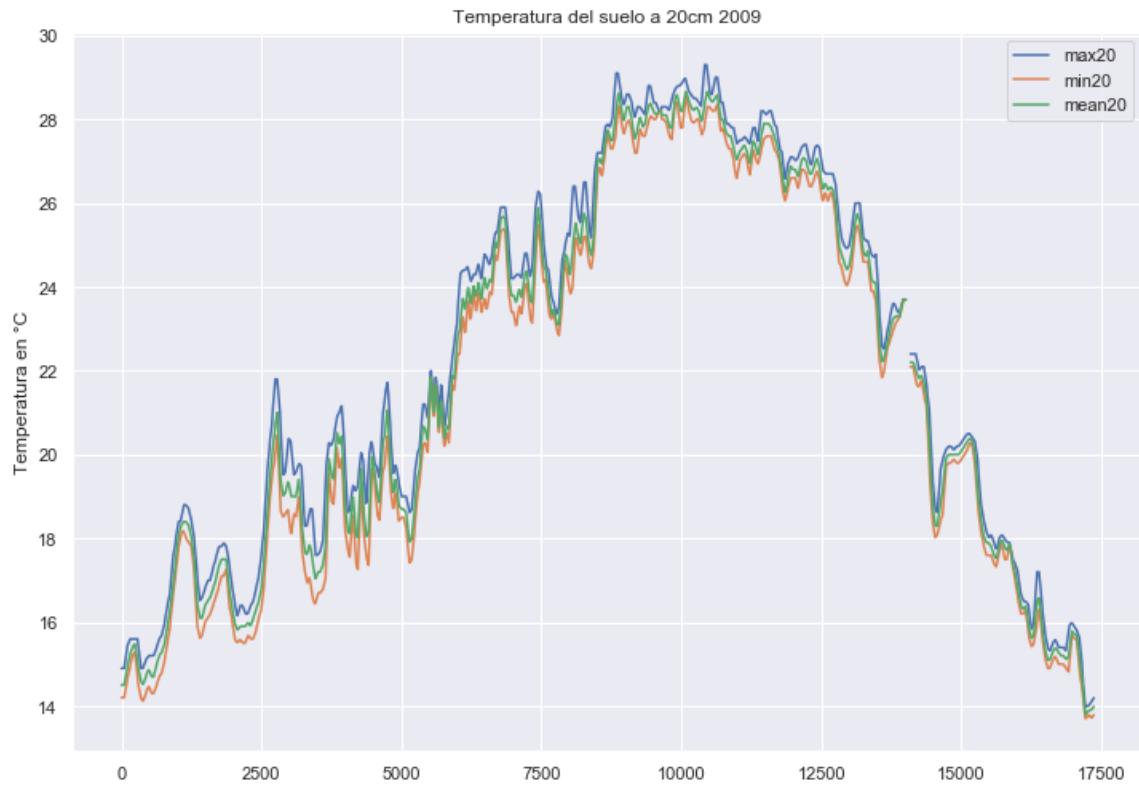


Figura 11: Temperaturas máxima, mínima y promedio diarias a los 20 centímetros bajo suelo usando rolling mean del 2009.

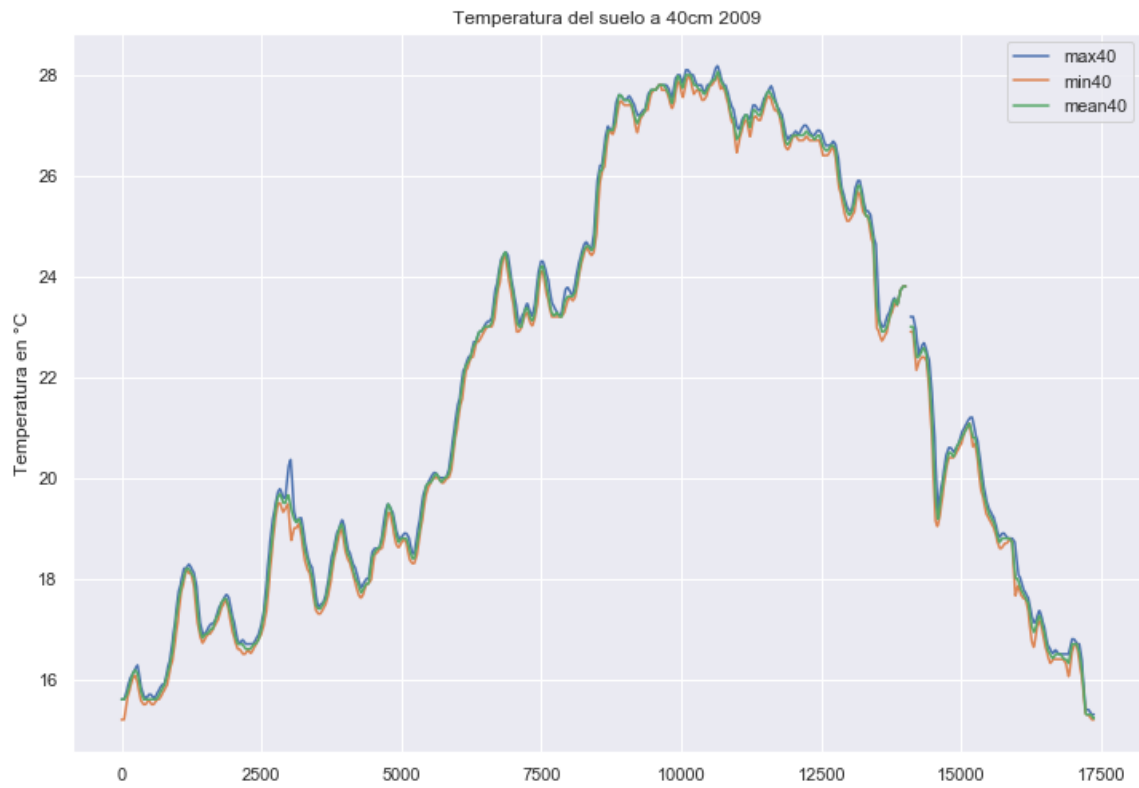


Figura 12: Temperaturas máxima, mínima y promedio diarias a los 40 centímetros bajo suelo usando rolling mean del 2009.

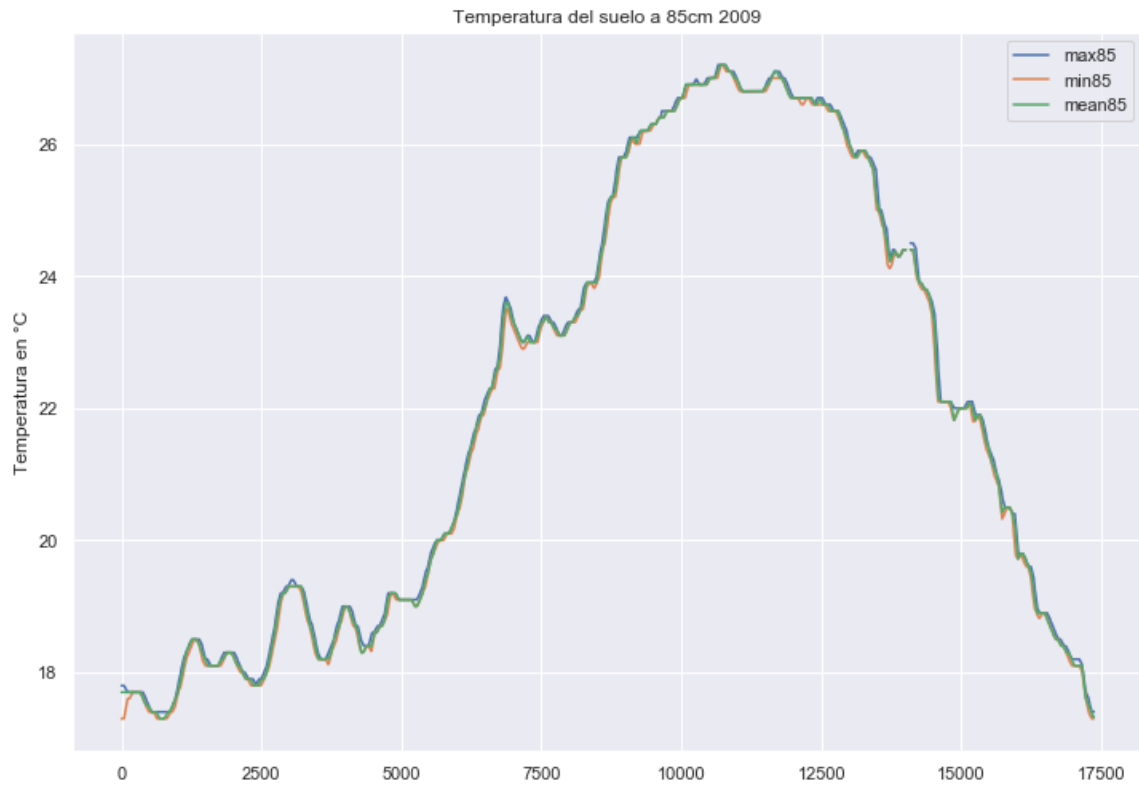


Figura 13: Temperaturas máxima, mínima y promedio diarias a los 85 centímetros bajo suelo usando rolling mean del 2009.

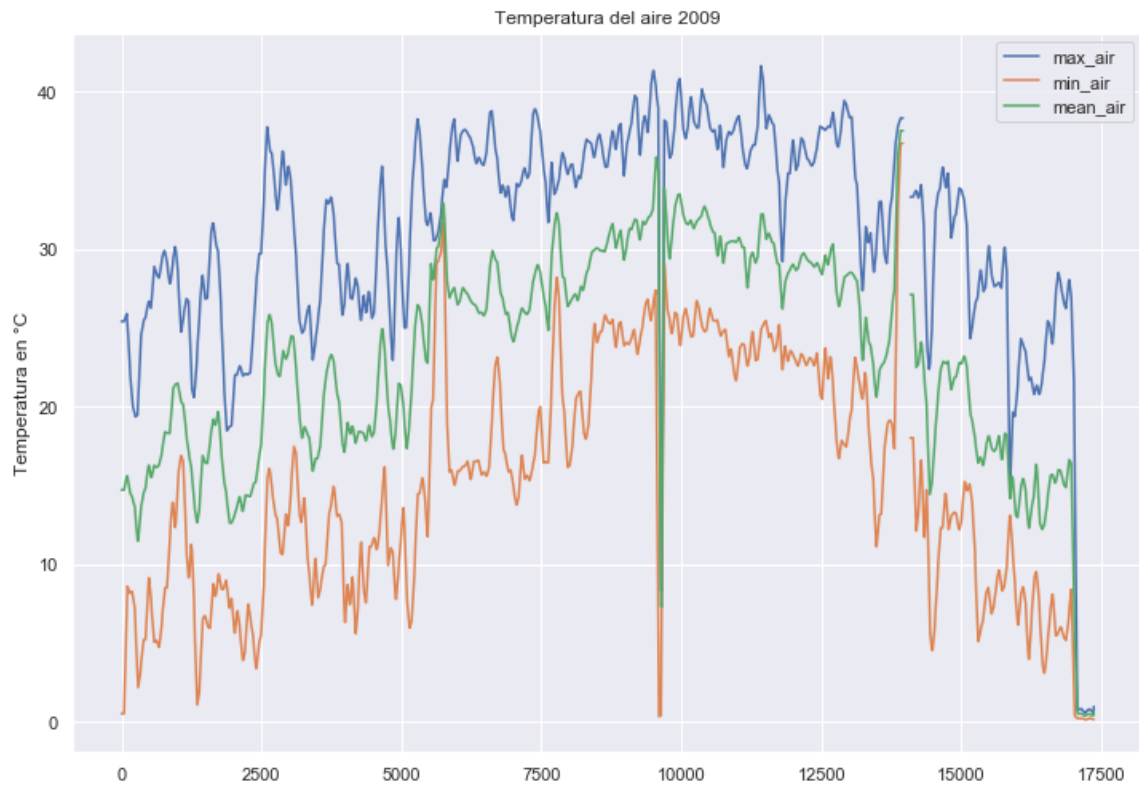


Figura 14: Temperaturas máxima, mínima y promedio diarias del aire usando rolling mean del 2009.

3. Conclusiones

En general, se cumple que a mayor profundidad de suelo, mayor temperatura se obtiene en las mediciones y, por el contrario, a menor profundidad de suelo, menor temperatura. Además, la temperatura del aire se mantiene oscilante durante todo el año, pero acercándose a mayores temperaturas durante el verano y a menores durante el invierno.

4. Bibliografía

- Seaborn: statistical data visualization. (2018). Consultado en Mayo del 2019, de Seaborn. Sitio web: <https://seaborn.pydata.org/>
- Merge, join, and concatenate. Consultado en Mayo del 2019, de PandasPydata.org. Sitio web: https://pandas.pydata.org/pandas-docs/stable/user_guide/merging.html