

# Actividad 8: Uniendo DataFrames en Pandas

Ángela Morales Zamudio

Lun-Mié 10-11. Jue 4-6

Marzo, 2019

En esta actividad se trabajará con dos conjuntos de datos: con los datos meteorológicos de la estación de nogal de la actividad pasada y con otros datos del suelo (variables medidas por sensores bajo el suelo), ambos coincidiendo en el periodo 2009 que será estudiado.

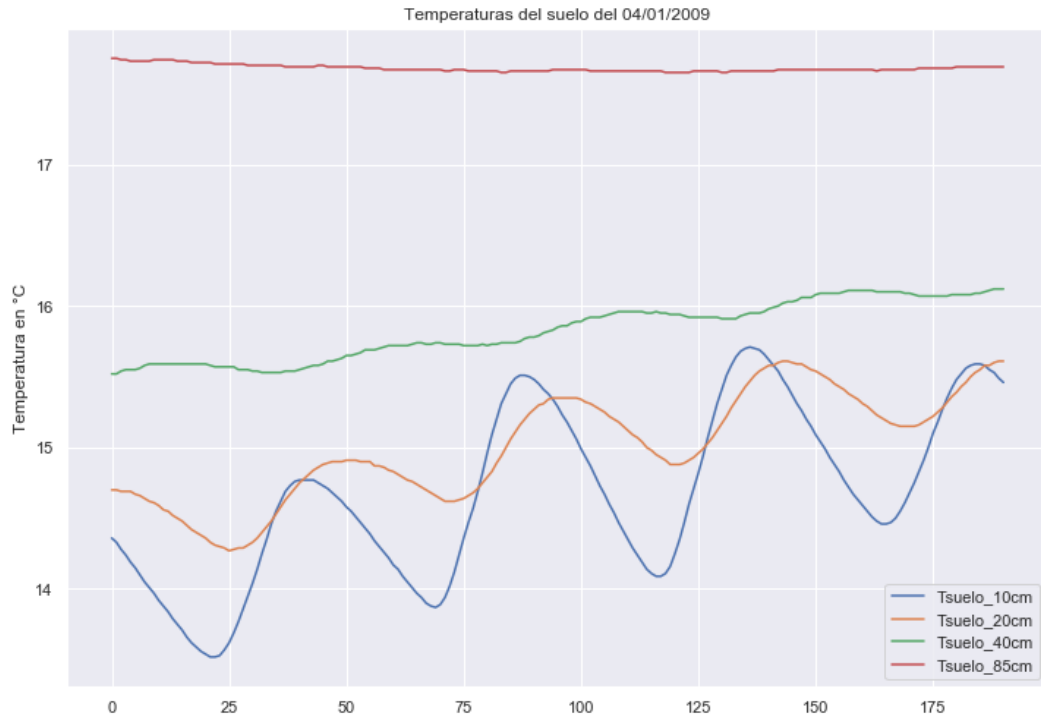
En el segundo conjunto se creará una nueva columna para la variable temporal como en el primer archivo de datos para poder incorporar ambos en un solo DataFrame. Se estudiarán los datos relacionados con las lecturas de las temperaturas del suelo en 8 profundidades distintas, incorporando la temperatura del aire.

Se seleccionará luego un día de Enero (por ejemplo, el 4 de Enero) y se graficará la temperatura del aire y las temperaturas del subsuelo (10, 20, 40 y 85 cm). Se realizará una gráfica con las temperaturas máxima, mínima y promedio diaras para el 2009. También se calculará el promedio cada media hora al día para el mes de Enero para la temperatura del aire y las 8 temperaturas promedio del subsuelo y se graficará esta variación en 24 horas. Además, se graficará el promedio móvil (rolling mean) para suavizar la evolución temporal de una serie de tiempo.

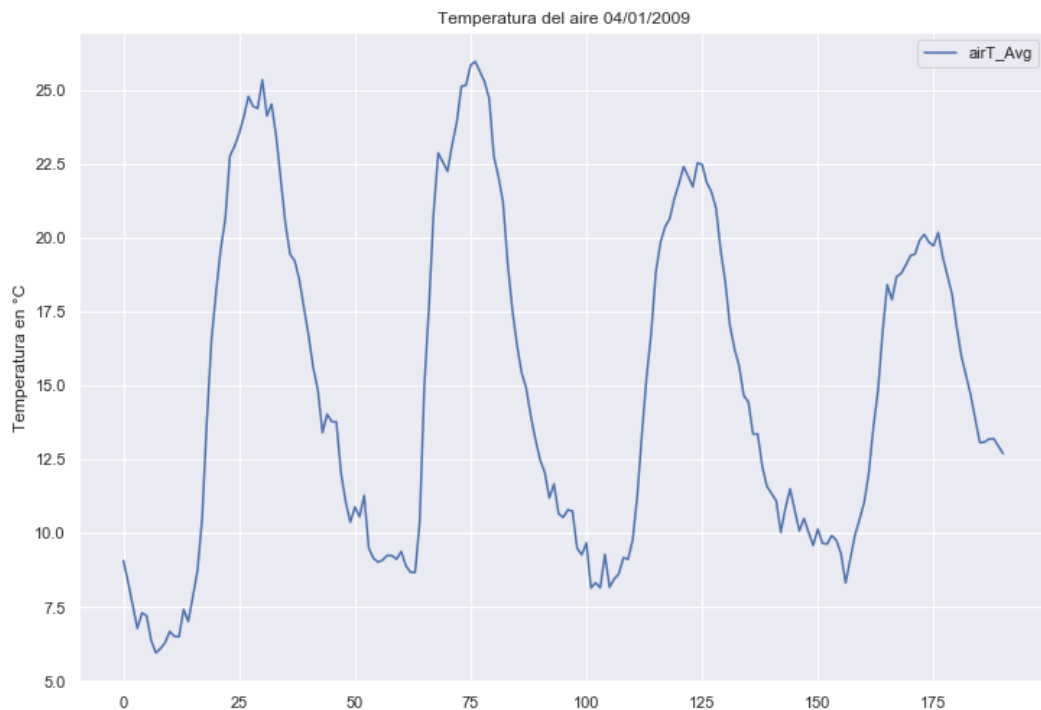
Nos interesa estudiar la variación de la temperatura en el subsuelo como función de la hora del día y la profundidad del suelo.

Para unir ambos DataFrames fue necesario una rigurosa serie de pasos, como evaluar y modificar los tipos de variables, eliminar datos innecesarios, crear variables temporales para la obtención de la variable *Fecha*, realizar loops para el llenado de los arreglos temporales, y el uso del comando *merge* para unir los DataFrames una vez ajustados.

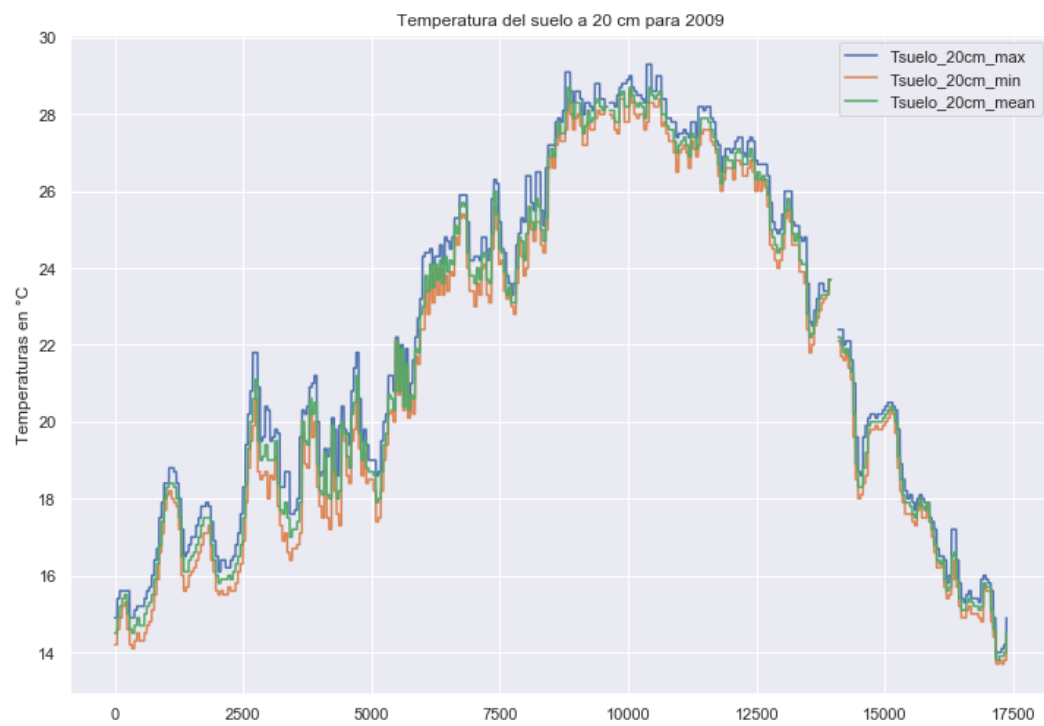
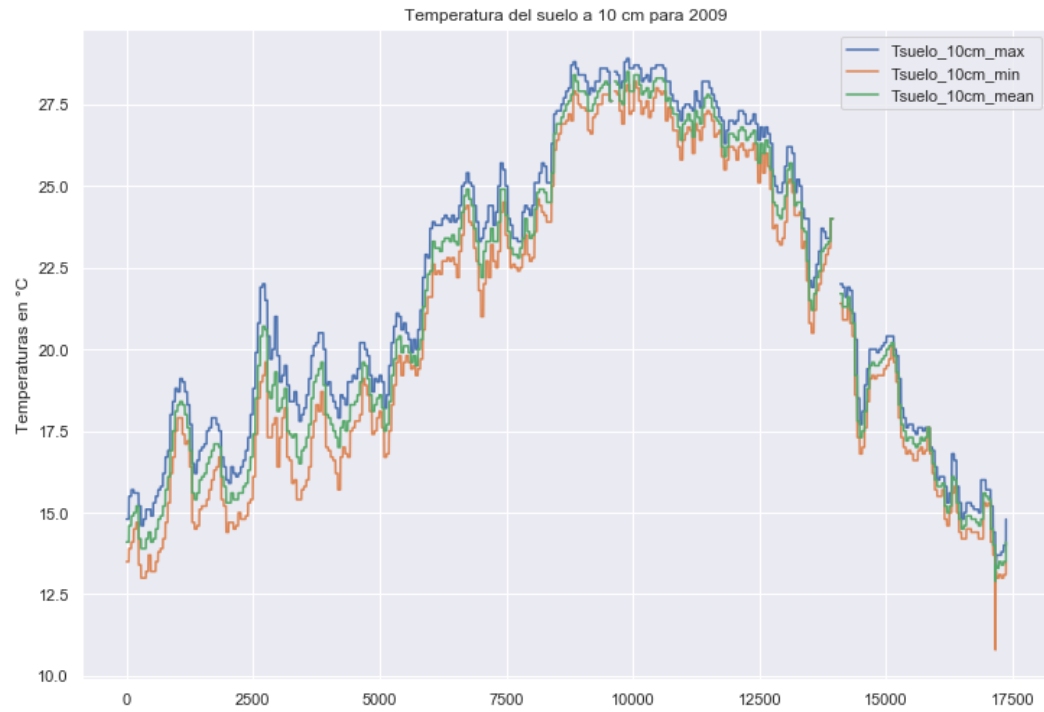
Teniendo el nuevo DataFrame, se prosiguió a la obtención de las temperaturas del subsuelo a 10, 20, 40 y 85 metros del día 4 de Enero, obteniendo la siguiente gráfica:

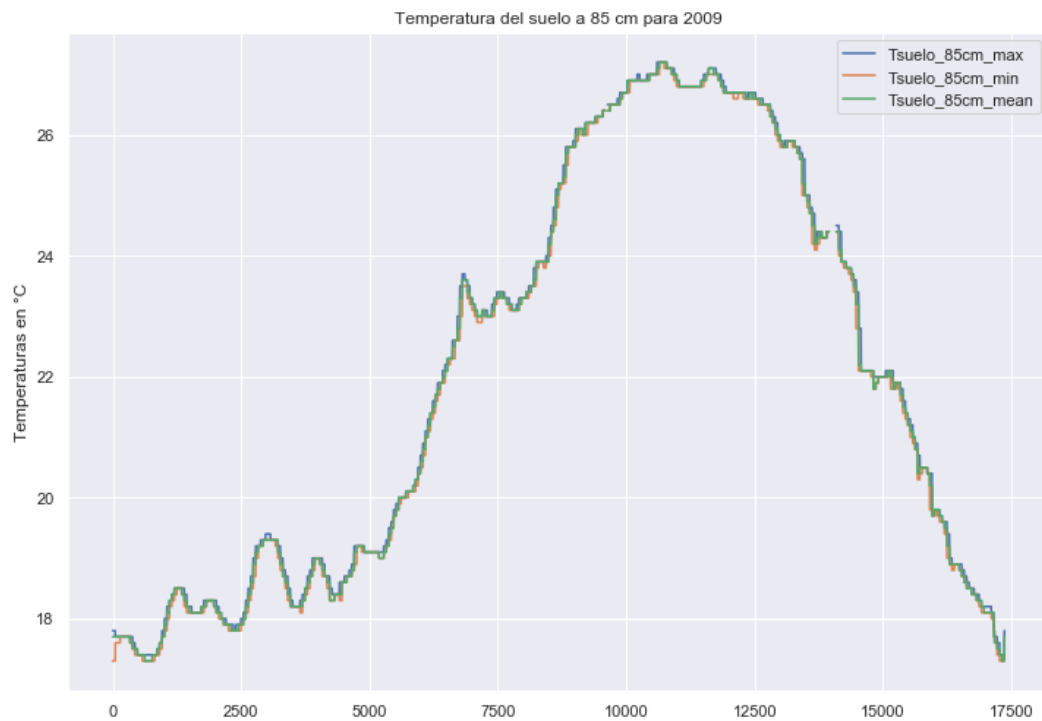
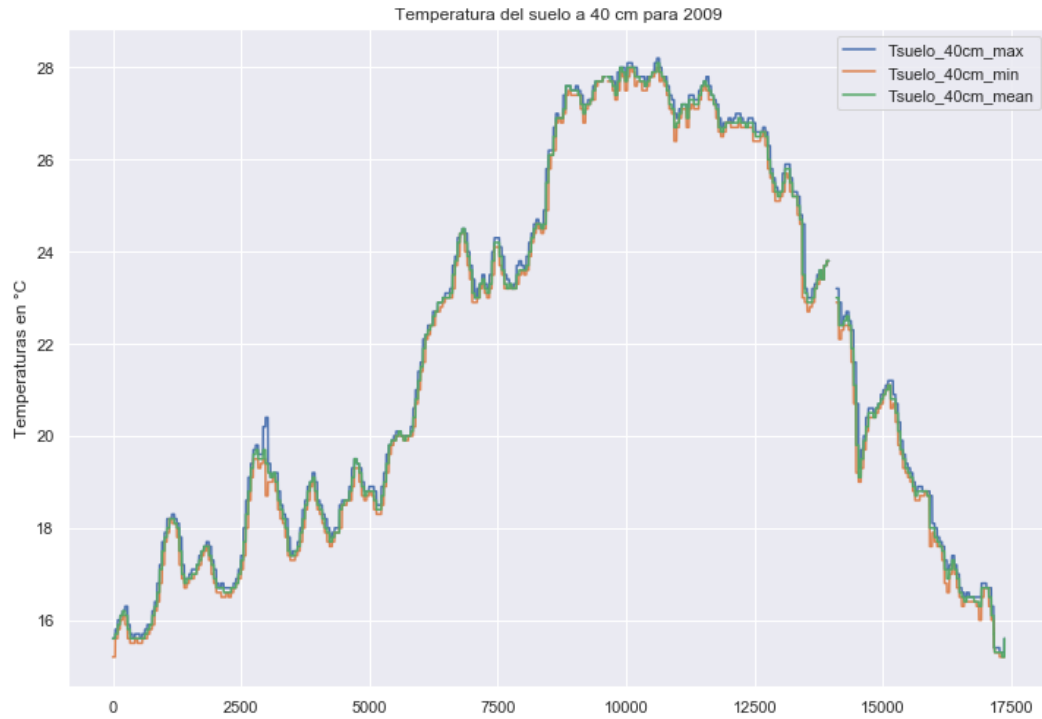


Luego se obtuvo la gráfica del 4 de Enero de la temperatura del Aire:

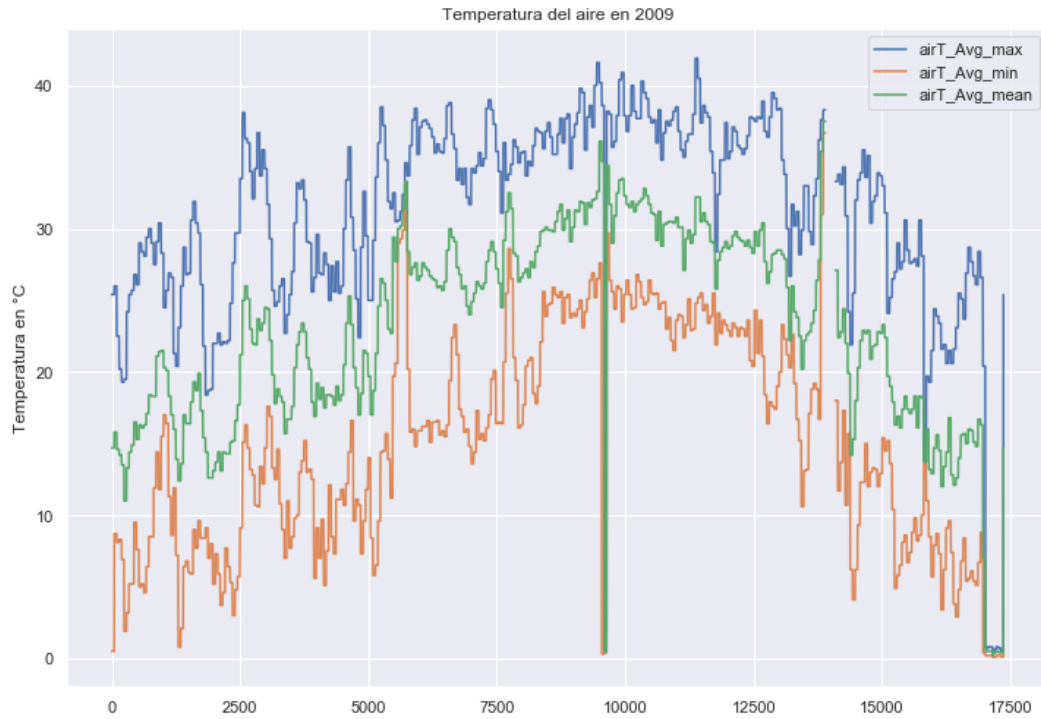


Luego se realizó una gráfica de las temperaturas máxima, mínima y promedio diarias para el año completo de datos del 2009, utilizando la función *groupby* para la temperatura con su respectivo argumento. Se obtuvieron las gráficas de las 4 temperaturas del subsuelo:

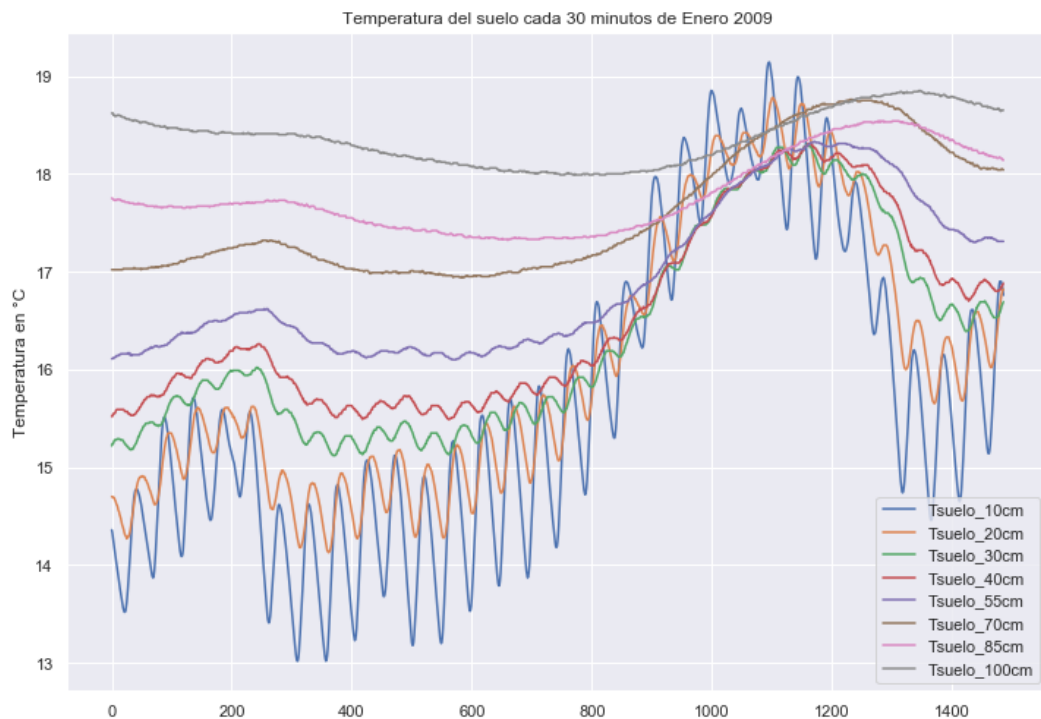




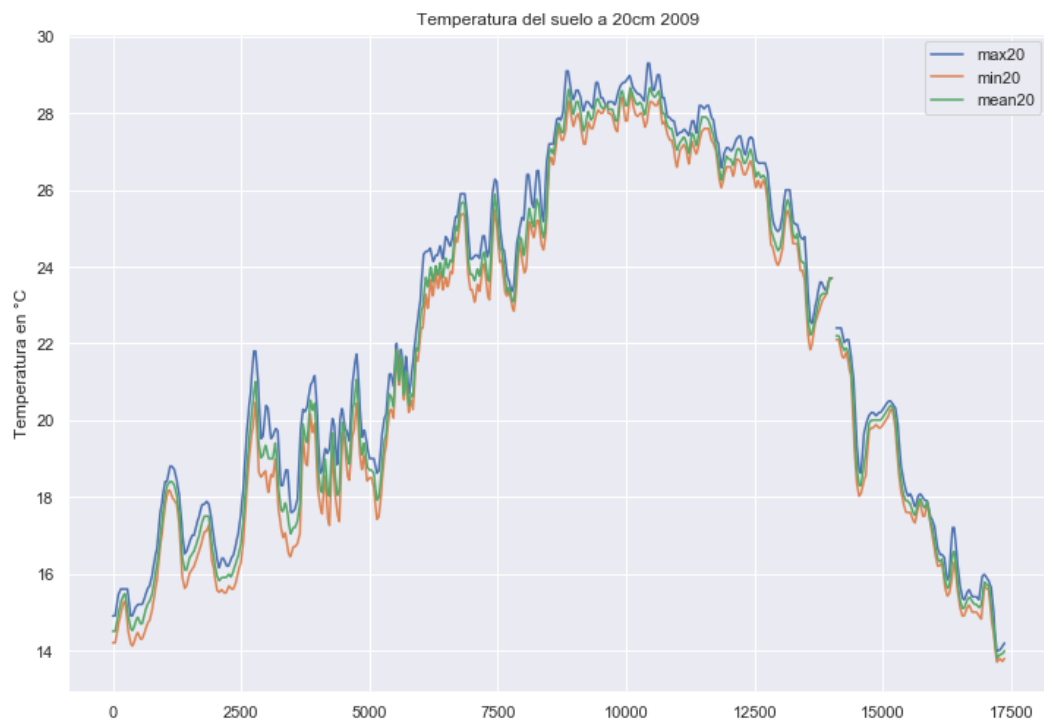
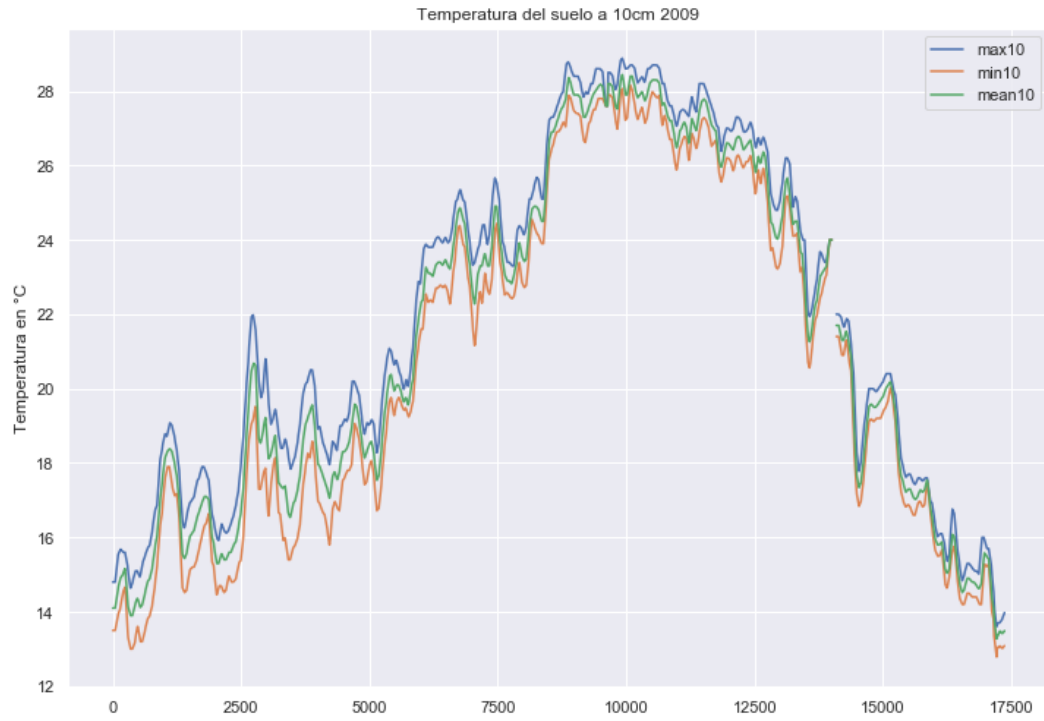
Para la gráfica de la temperatura del Aire, seguimos los mismos pasos anteriores para las temperaturas máxima, mínima y promedio en el 2009, y se graficaron, obteniendo la siguiente gráfica con estos tres criterios:

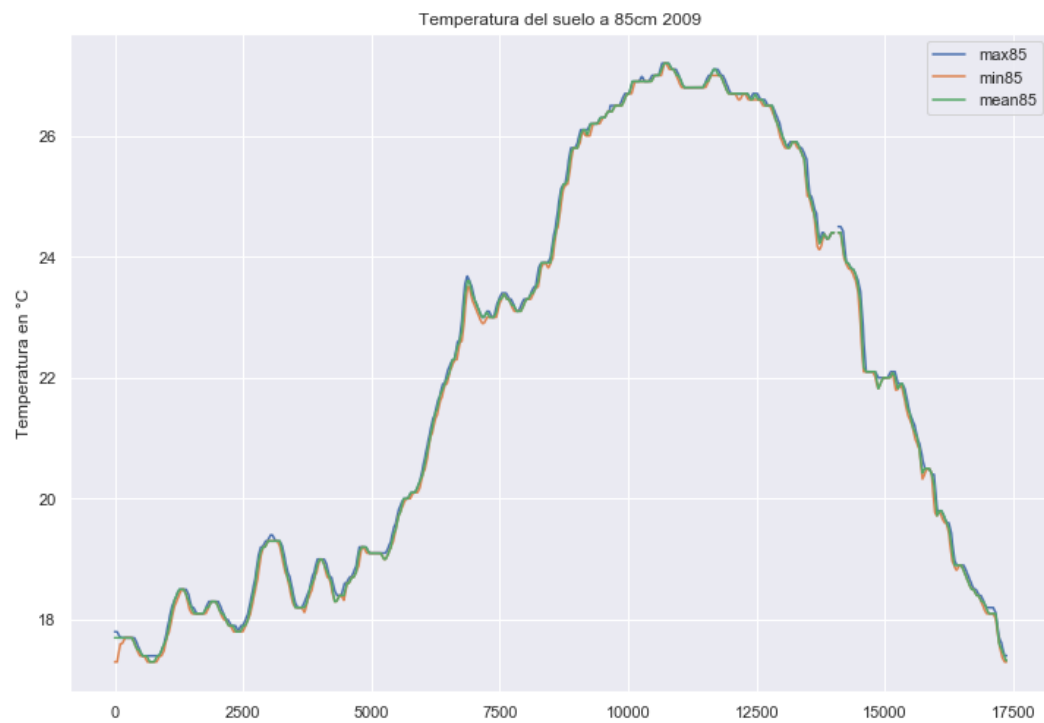
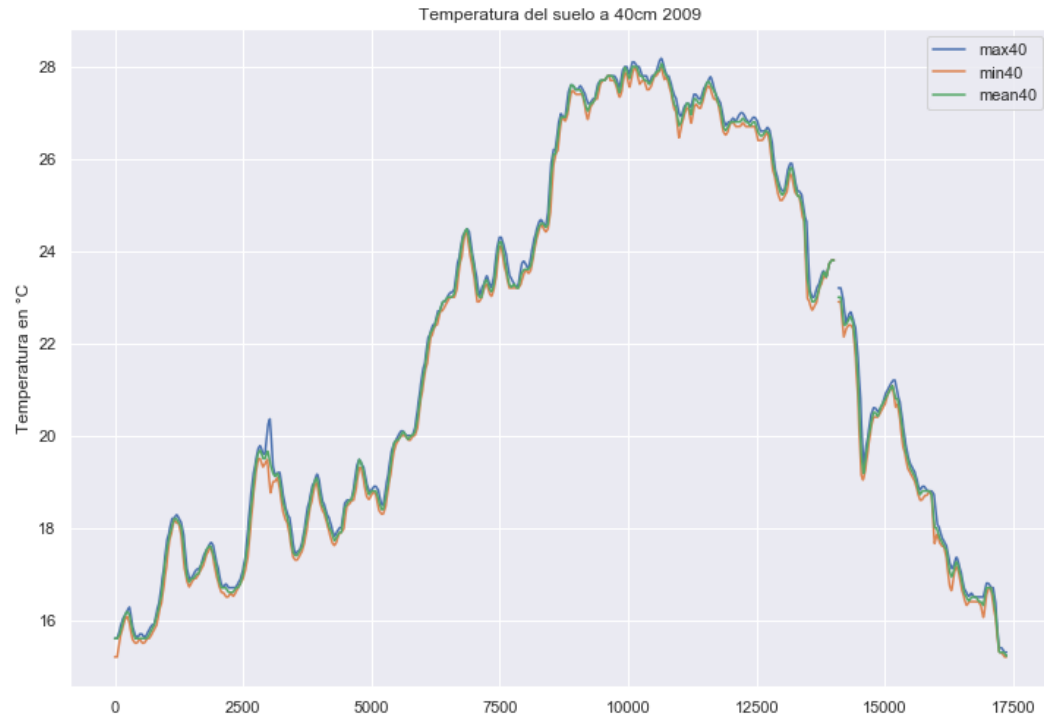


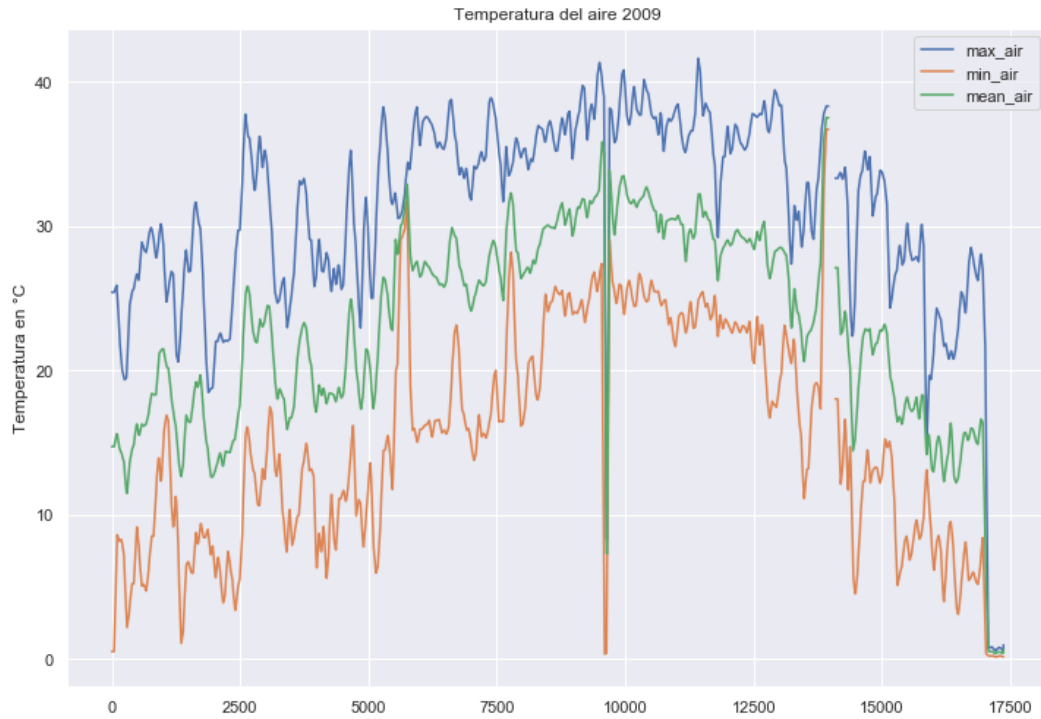
Para calcular el promedio cada 30 minutos durante el día para el mes de Enero de la temperatura del aire y las 8 temperaturas promedio del subsuelo, para posteriormente graficar la variación en 24 horas de las temperaturas de interés, se redujo el DataFrame a las puras temperaturas del subsuelo cada media hora y se graficaron :



Para las gráficas del promedio móvil de las temperaturas del suelo y del aire se utilizó la función *rolling*. Se obtuvieron las siguientes gráficas:







Todas las herramientas usadas fueron esenciales para lograr la unificación de los DataFrames, el cual dicho proceso fue lo más difícil de realizar, pues fue necesario estar atento a pequeños factores como la longitud de cada uno, la indentación, incluso los tipos de variables. Una vez obtenido el DataFrame final, se utilizaron herramientas con las que ya estamos familiarizados para poder graficar.