



Curso: Programación Nivel II – Introducción a Pandas con Python

Taller 1

Preparación

El ejercicio propone el manejo de la librería *Pandas* en *Python* para la carga y el análisis de un *dataset* que corresponde a las mediciones de ciertos parámetros climáticos de una ciudad al norte de Europa durante un año.

El *dataset* contiene las columnas:

- Fecha
- Temp (C)
- Temp (C) Punto de Rocío
- Hum Rel (%)
- Vel. Viento (km/h)
- Visibilidad(km)
- Presión (kPa)
- Pronóstico

Actividades

1. El material se encuentra en la carpeta de *github*
https://github.com/dfloaiza/Python_TTIC2022/tree/main/Semana2/Tarea1
 - a. El archivo plano **weather2012_d.csv** desde donde debe cargarse el dataframe
 - b. Código de ejemplo trabajado en clase **practica_pandas1.py**
 - c. **Documentos de referencia**
2. Realizar la carga de los datos en un *dataframe* desde el archivo plano
3. Realizar un análisis previo del dataframe: nombres de las columnas, celdas nulas, número total de filas
4. Realizar la limpieza respectiva del dataset de acuerdo a lo visto en clase
5. Realizar el siguiente análisis:
 - a. Determinar la columna Fecha como la columna índice
 - b. Obtener una copia del dataframe que incluya solo las columnas Fecha, Temp (C), Hum Rel(%) y Pronóstico
 - c. Realizar agrupaciones por pronóstico con las siguientes agregaciones:
 - i. Promedio de temperatura



- ii. Promedio de humedad relativa
- d. BONUS: Obtener una tabla de correlación¹ para las columnas humedad, temperatura, y pronóstico². ¿Cuál es su interpretación?

Evaluación

- Se evalúa que el código esté completo y realice las tareas correctas con el dataset dado

Condiciones de Entrega

- Debe compartirse el código completo en Python que realice todas estas tareas de forma **correcta**
- La entrega debe ser individual
- Puede enviarse el código por correo a diego.loaiza02@usc.edu.co o compartir el enlace al repositorio donde se encuentra
- Plazo de Entrega: **Lunes 11 de Julio**

Referencias

(Disponibles en el repositorio)

- *Python Programming Fundamentals*
- *A Begginer's Guide to Python 3 Programming*
- *Python pandas tutorial*

¹ <https://es.wikipedia.org/wiki/Correlaci%C3%B3n>

² El método `corr()` aplicado al dataframe devuelve una tabla de correlación entre las distintas columnas de datos. Así, si su dataset se llegara a llamar `df_clima`, `print(df_clima.corr())` imprimiría la tabla de correlación.