JORGE LUIS GONZALEZ CASTRO

CC: 1032395475

https://github.com/jlgingcivil/R.LTWB.CS2120

r.ltwb – SECTION 03

Descarga, procesamiento y análisis de datos hidroclimatológicos

Exploración y análisis de series - EDA - Representación gráfica

**TABLA DE CONTENIDO**

[1. Introducción 2](#_Toc145604542)

[2. Objetivo General 2](#_Toc145604543)

[3. Actividad 1: Procesamiento en software 2](#_Toc145604544)

[4. ACTIVIDAD 2: análisis de longitud de series 12](#_Toc145604545)

[5. ACTIVIDAD 3: análisis de OTROS PARÁMETROS 16](#_Toc145604546)

[6. ACTIVIDAD 4: análisis de resultados 17](#_Toc145604547)

[7. Conclusiones 23](#_Toc145604548)

[8. Referencias Bibliográficas 24](#_Toc145604549)

**ÍNDICE DE ILUSTRACIONES**

[Ilustración 3‑1. Descarga EDA-py 2](#_Toc145604511)

[Ilustración 3‑2. Script EDA.py 3](#_Toc145604512)

[Ilustración 3‑3. Ejecución inicial EDA.py 3](#_Toc145604513)

[Ilustración 3‑4. Eliminación estaciones problemáticas 4](#_Toc145604514)

[Ilustración 3‑5. Ejecución EDA.py 5](#_Toc145604515)

[Ilustración 3‑6. Resultados script EDA.py 7](#_Toc145604516)

[Ilustración 3‑7. Visualización Resultados EDA.md 9](#_Toc145604517)

[Ilustración 4‑1. Comparación Estaciones Caudal 12](#_Toc145604518)

[Ilustración 4‑2. Comparación Estaciones Evaporación 13](#_Toc145604519)

[Ilustración 4‑3. Comparación Estaciones Precipitación 14](#_Toc145604520)

[Ilustración 4‑4. Comparación Estaciones Temperatura Máxima 15](#_Toc145604521)

[Ilustración 4‑5. Comparación Estaciones Temperatura Mínima 16](#_Toc145604522)

[Ilustración 5‑1. Acceso plataforma IDEAM 17](#_Toc145604523)

[Ilustración 6‑1. Registros Evaporación 17](#_Toc145604524)

[Ilustración 6‑2. Correlación Evaporación 18](#_Toc145604525)

[Ilustración 6‑3. Registros Precipitación 19](#_Toc145604526)

[Ilustración 6‑4. Correlación Precipitación 19](#_Toc145604527)

[Ilustración 6‑5. Registros Caudal 20](#_Toc145604528)

[Ilustración 6‑6. Correlación Caudal 20](#_Toc145604529)

[Ilustración 6‑7. Registros Temperatura Máxima 21](#_Toc145604530)

[Ilustración 6‑8. Correlación Temperatura Máxima 22](#_Toc145604531)

[Ilustración 6‑9. Registros Temperatura Mínima 22](#_Toc145604532)

[Ilustración 6‑10. Correlación Temperatura Mínima 23](#_Toc145604533)

# Introducción

Se continua con curso Balance hidrológico de largo plazo para estimación de caudales medios usando SIG – LWTB con el desarrollo de la sección 3 Descarga, procesamiento y análisis de datos hidroclimatológicos. A continuación, se presenta en cada numeral las actividades realizadas de acuerdo con cada capítulo de la sección de estudio, incluyendo el resumen de actividades, logros alcanzados y capturas de pantalla de los ejercicios realizados. Se ha creado el repositorio <https://github.com/jlgingcivil/R.LTWB.CS2021> para la inclusión de los archivos y documentos de las actividades desarrolladas.

# Objetivo General

El objetivo general en esta sección es analizar a través de diferentes técnicas la calidad y confianza de las series de datos descargadas desde el portal del IDEAM, también realizar estadísticas principales de estos.

# Actividad 1: Procesamiento en software

En primera medida se realiza la descarga del script EDA.py y la creación de la carpeta para almacenamiento de archivos.

Ilustración ‑. Descarga EDA-py

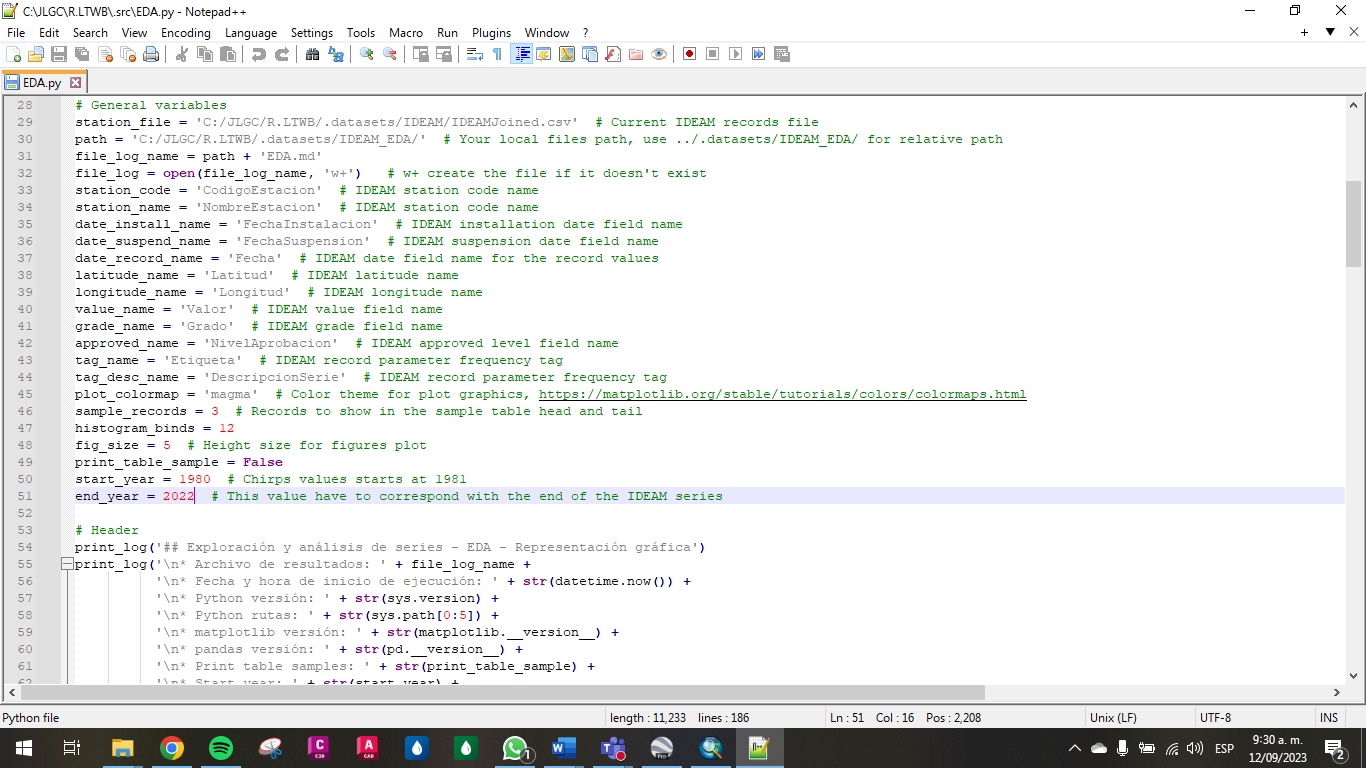
Captura de pantalla de computadora

Descripción generada automáticamente

Fuente: Elaboración Propia, 2023.

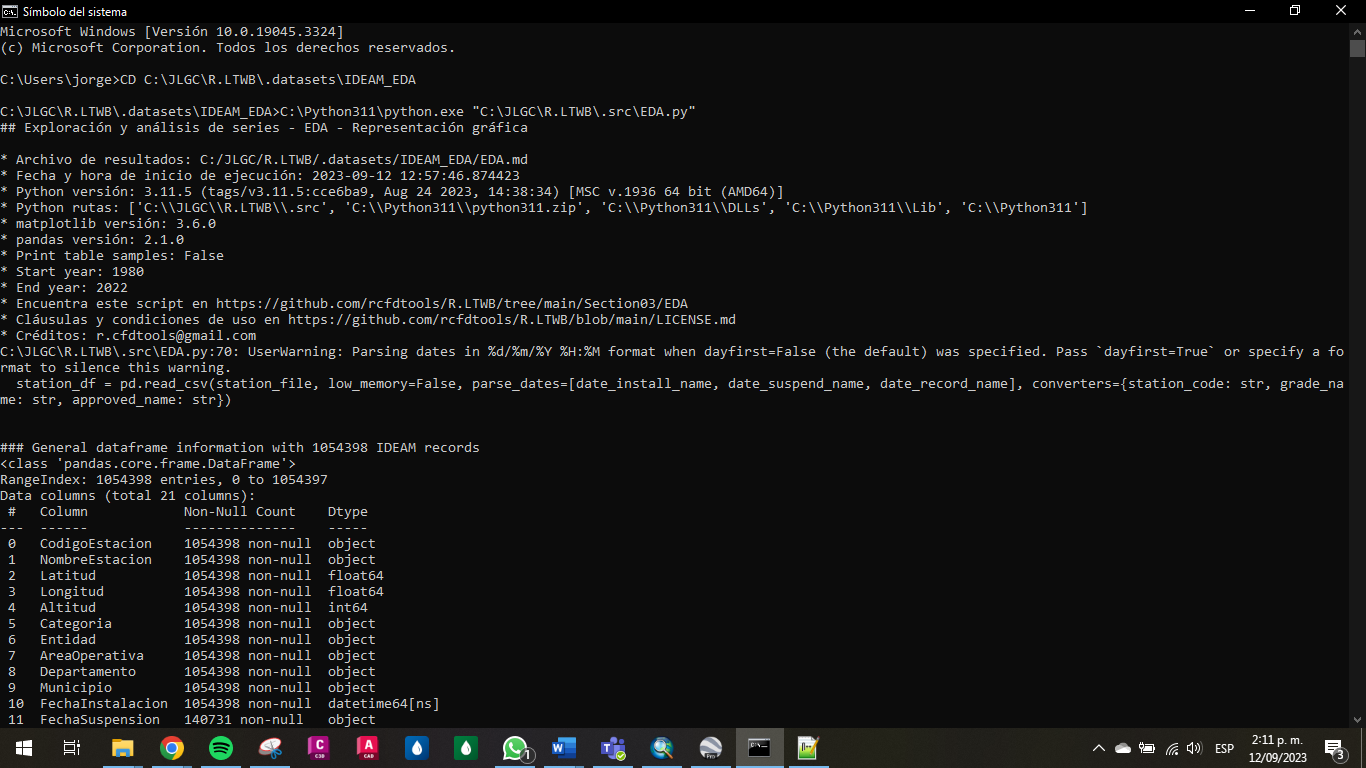
Luego se ajusta el script para que lea la ruta de almacenamiento de archivos y se inicia con la ejecución de la herramienta.

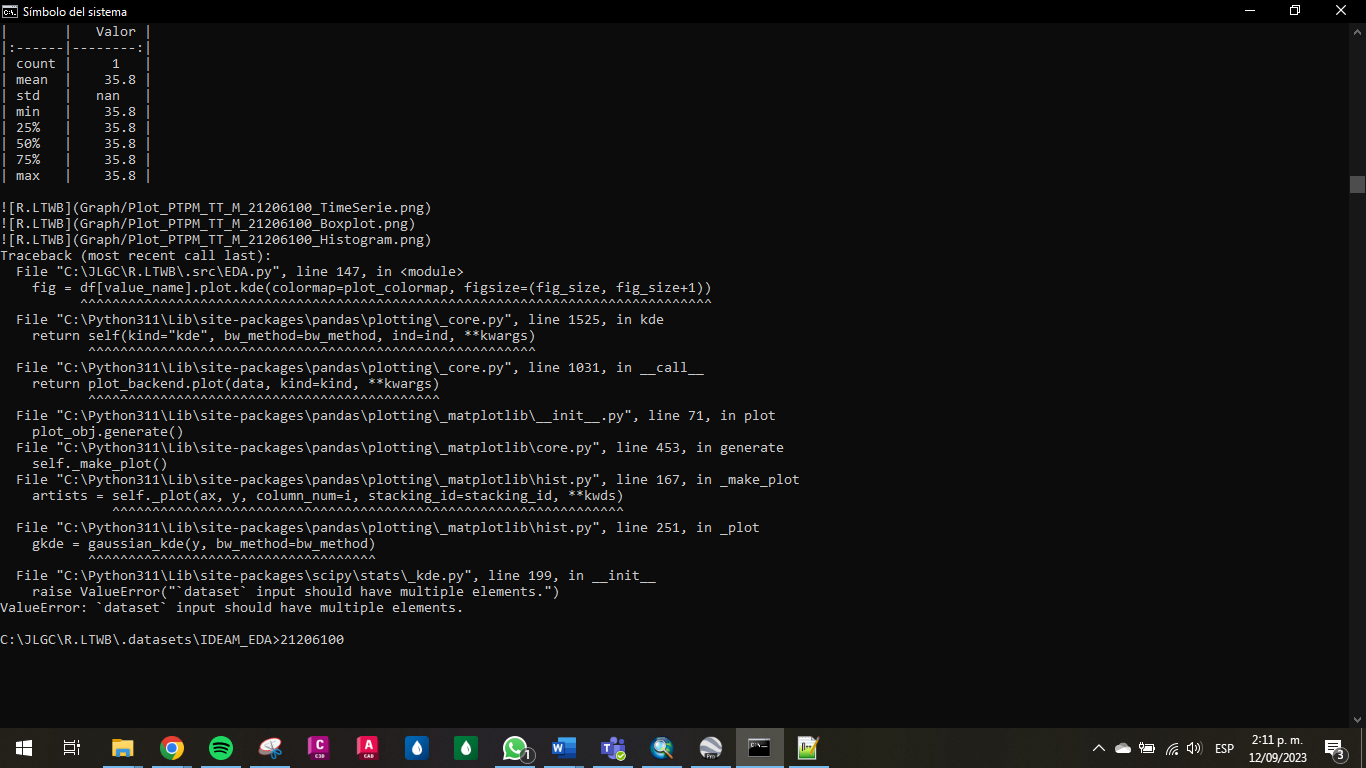
Ilustración ‑. Script EDA.py



Fuente: Elaboración Propia, 2023.

Ilustración ‑. Ejecución inicial EDA.py

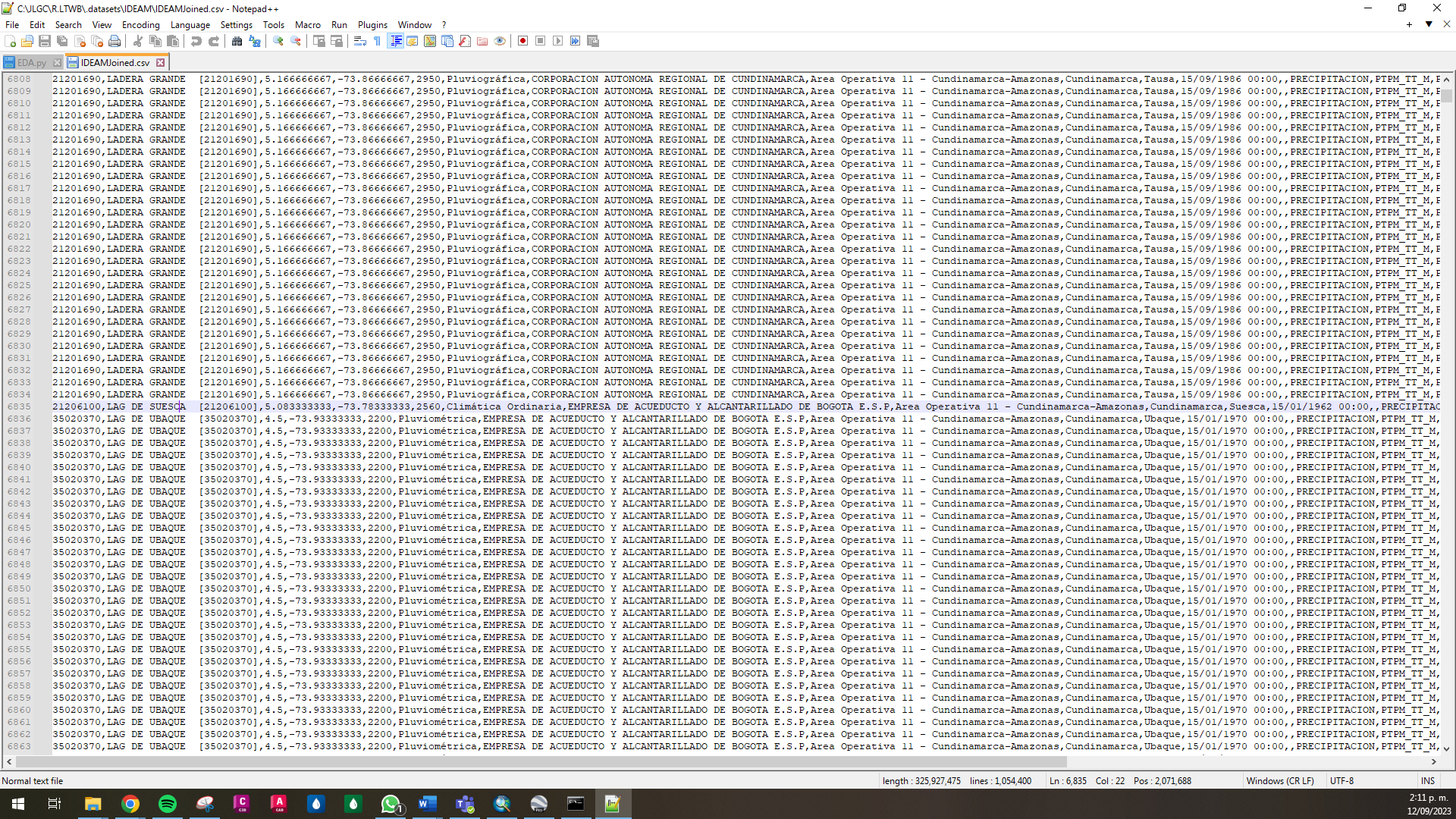


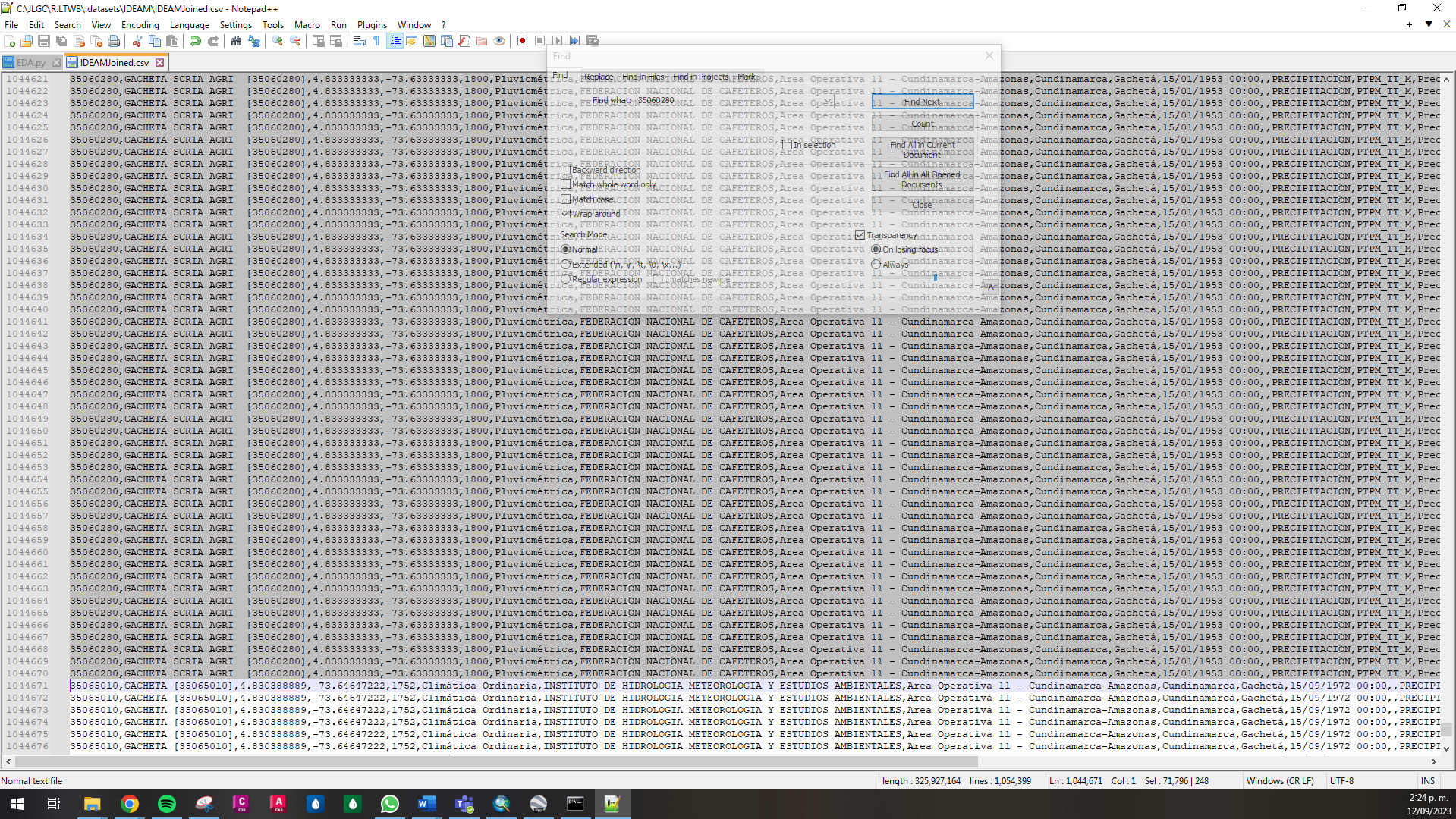


Fuente: Elaboración Propia, 2023.

Como se observa en la imagen anterior, el script produce un volcamiento dado un error en la cantidad de registros de algunas estaciones, por lo cual fue necesario realizar varias corridas del script para eliminar las problemáticas, las cuales fueron las codificadas con el número 21206100 y 35060280.

Ilustración ‑. Eliminación estaciones problemáticas

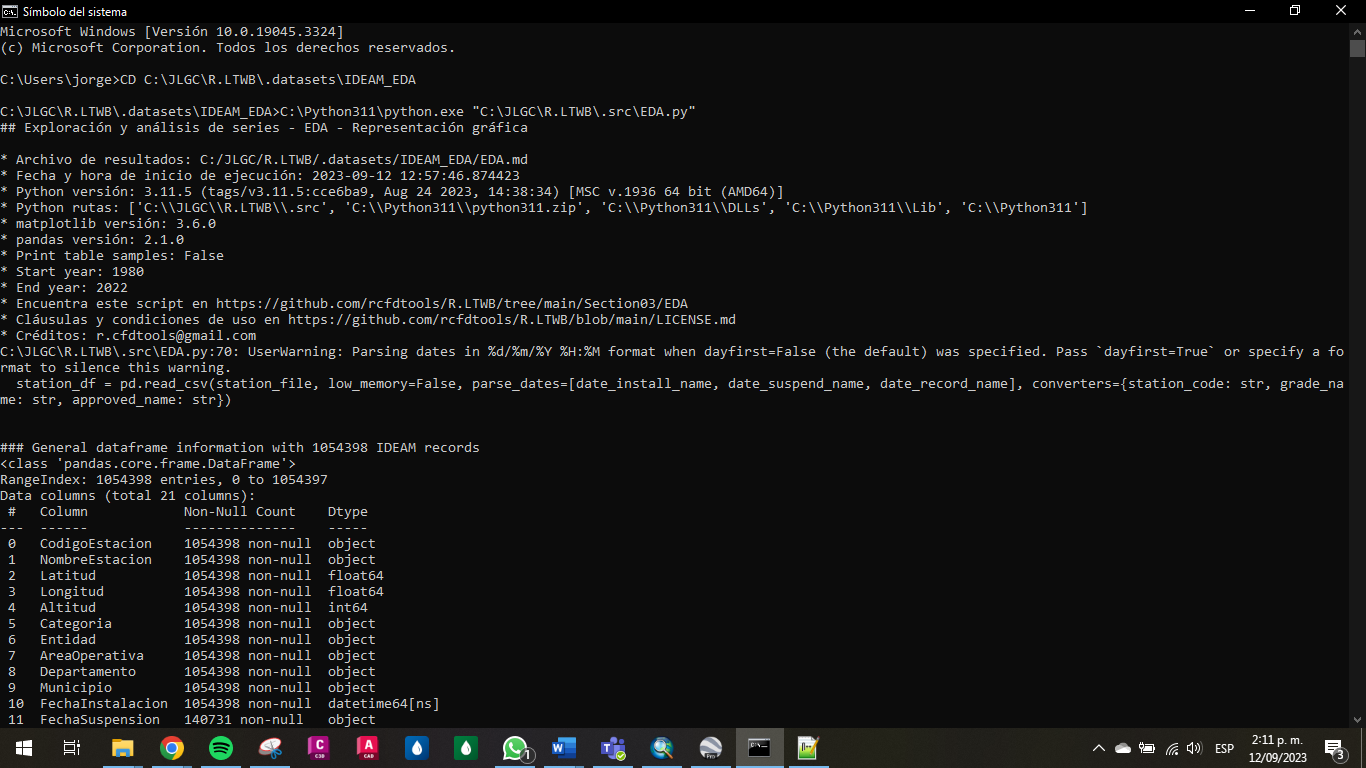


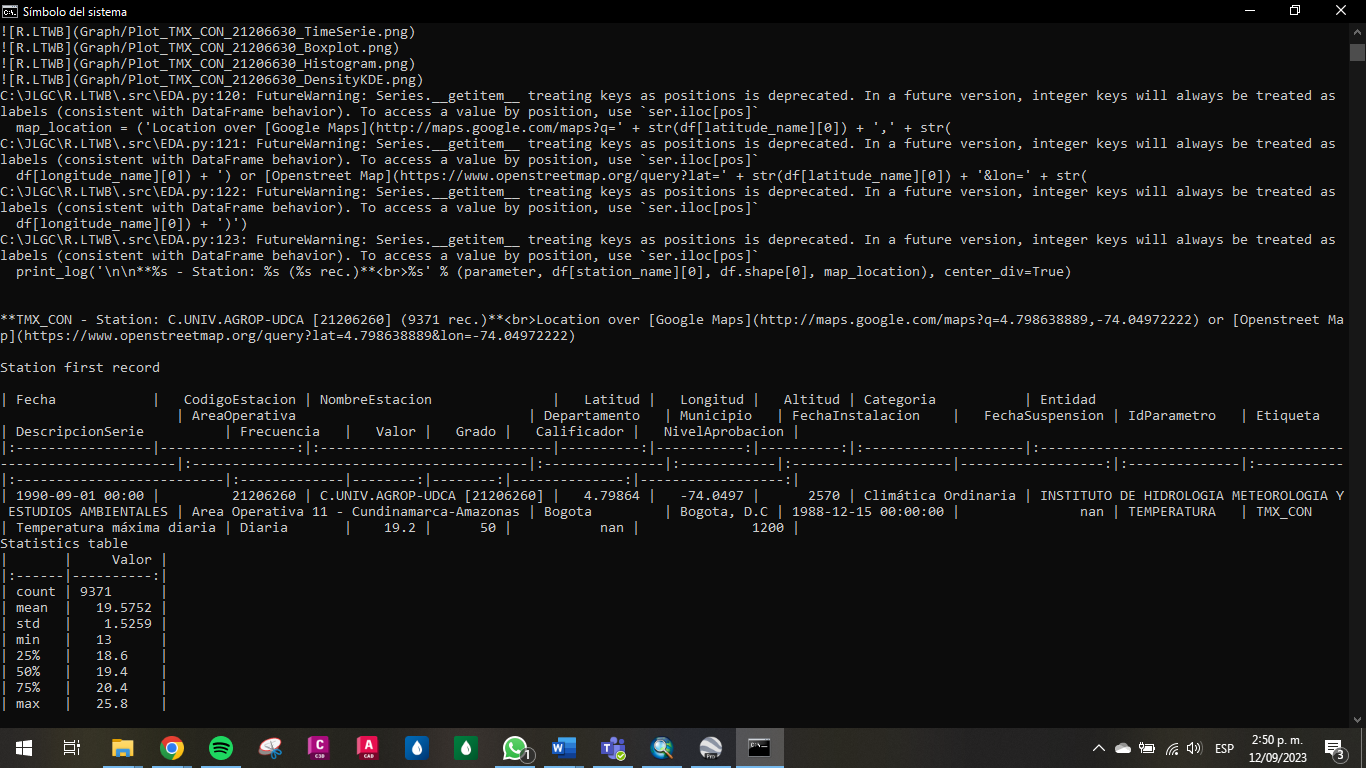


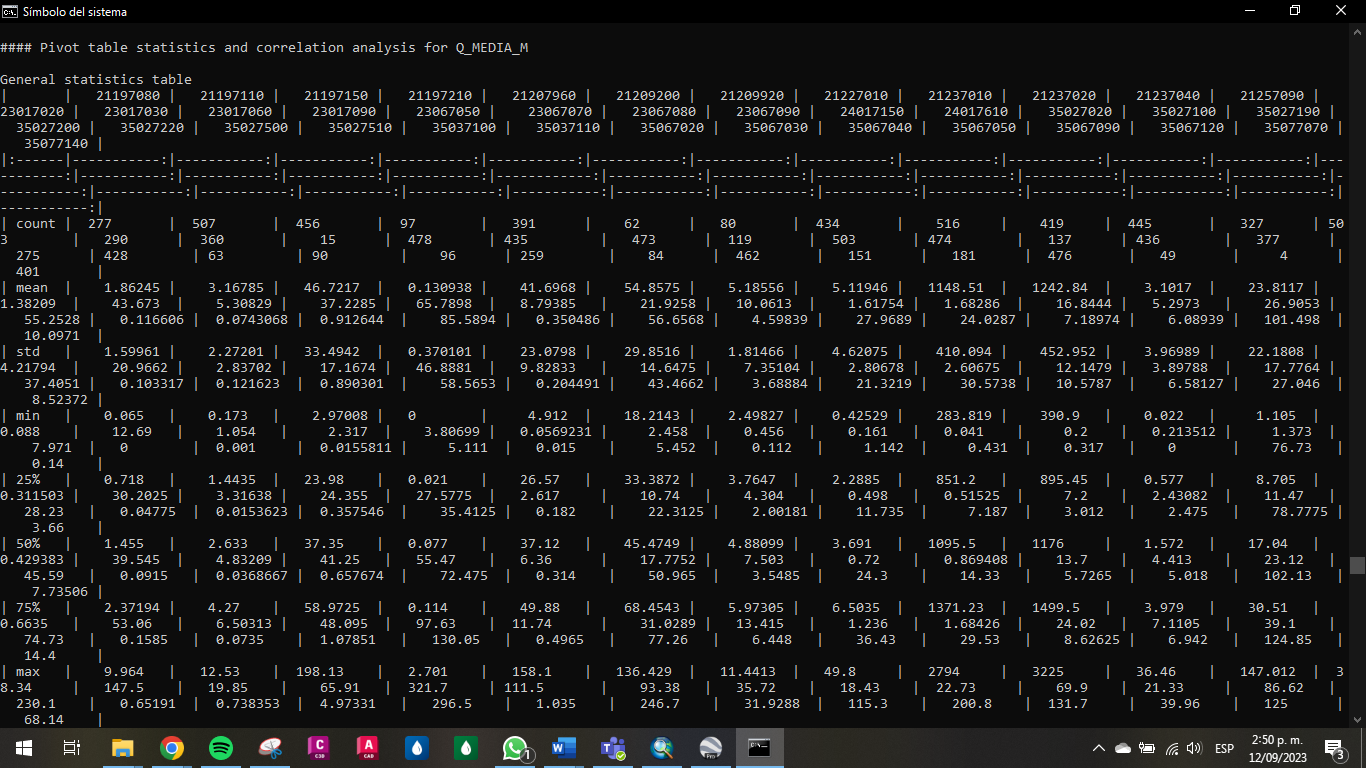
Se observa que en el primer caso la estación solo tenía un registro por lo que no se podía realizar una estadística, mientras que en el segundo aunque tenia varios registros no eran suficientes para realizar los cálculos de densidad KDE.

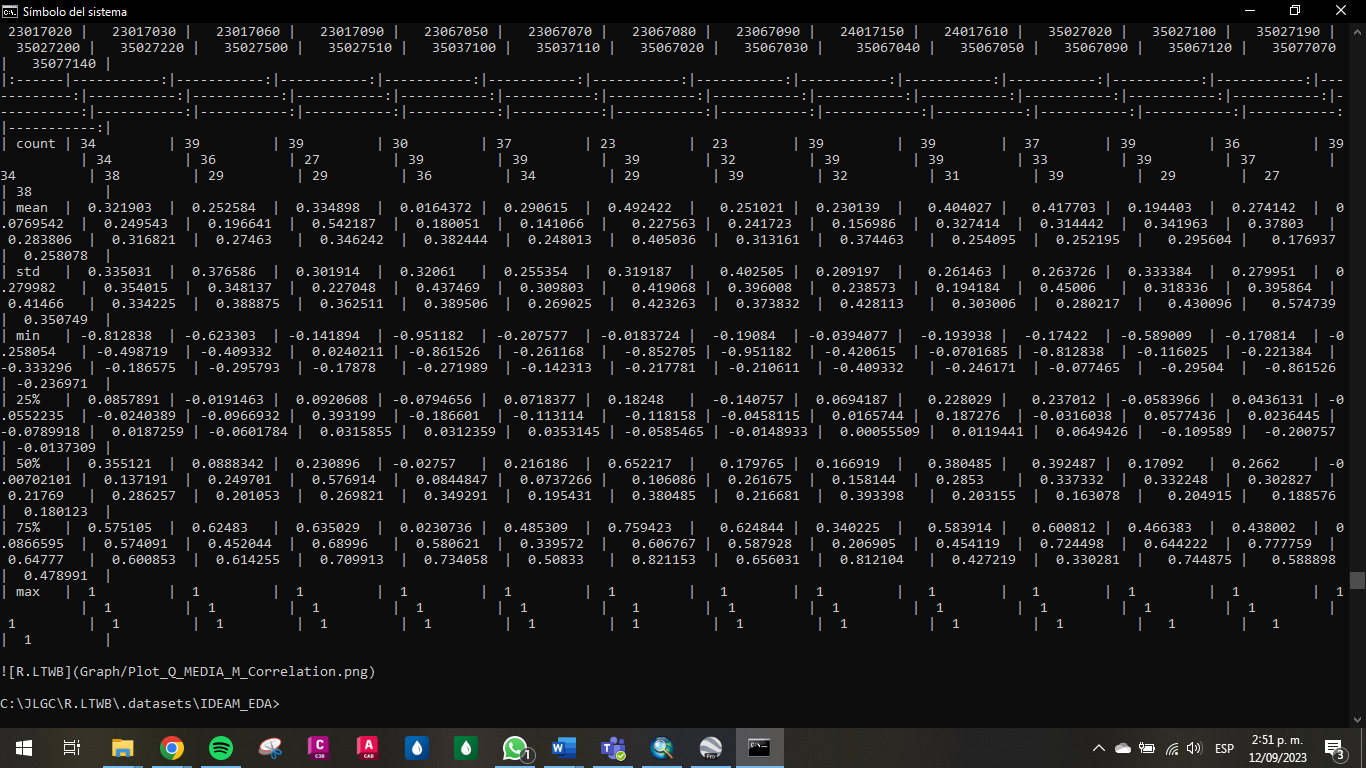
Una vez se eliminaron esta información la rutina se ejecutó correctamente. A continuación, se presenta los resultados obtenidos con el script.

Ilustración ‑. Ejecución EDA.py





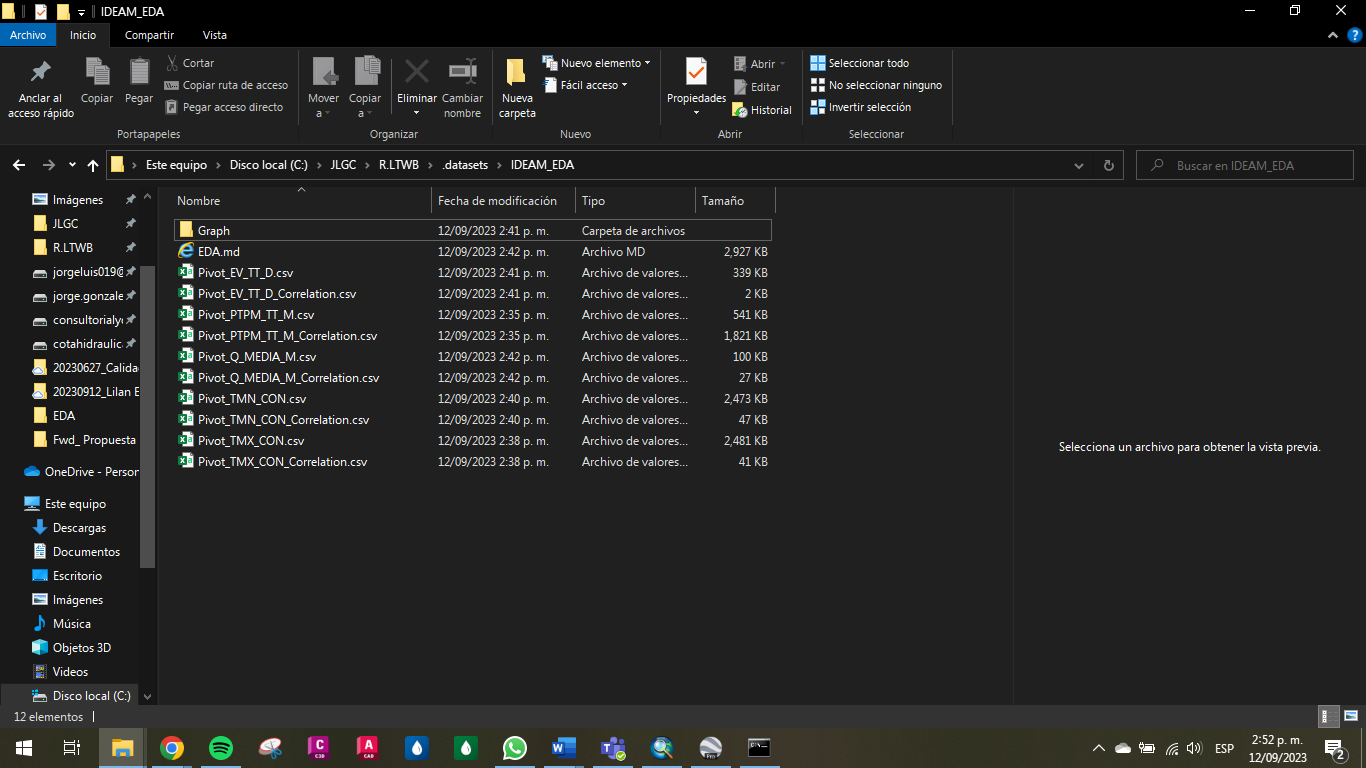


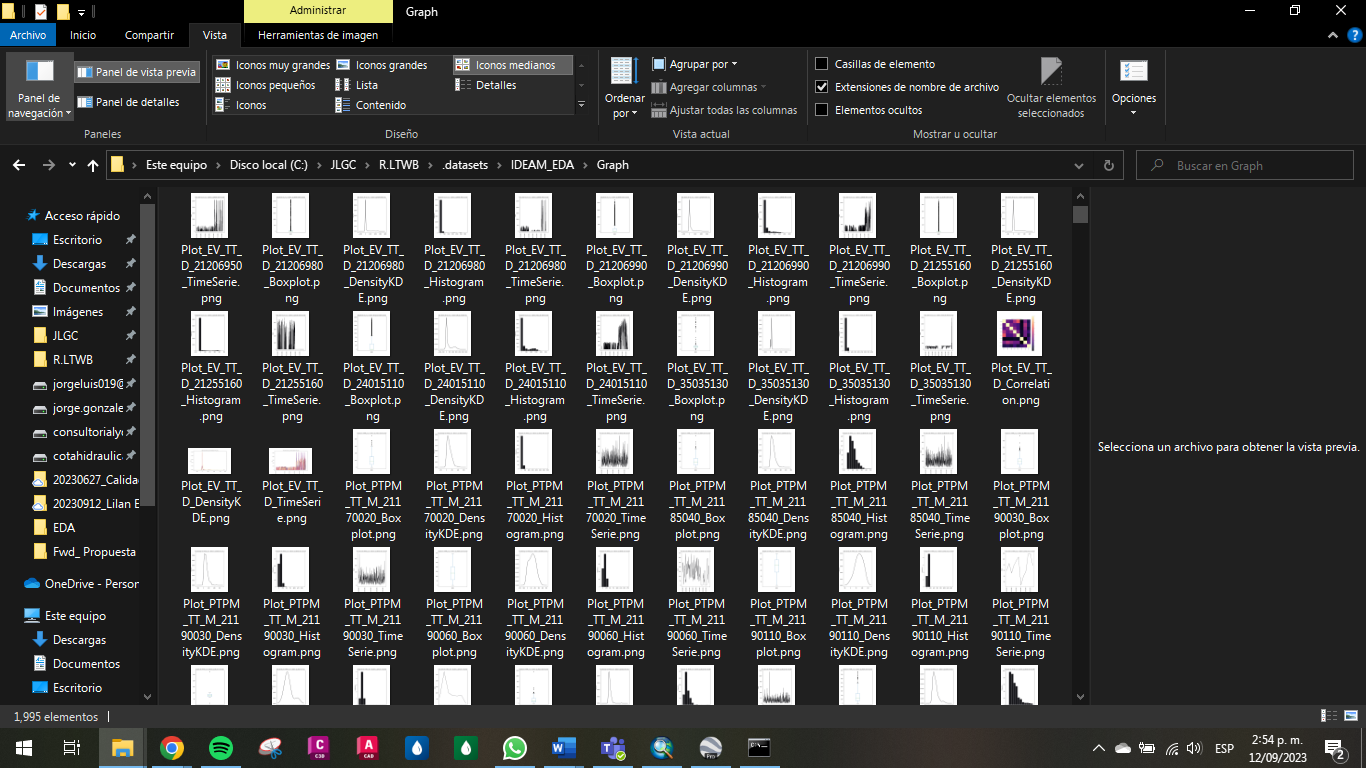
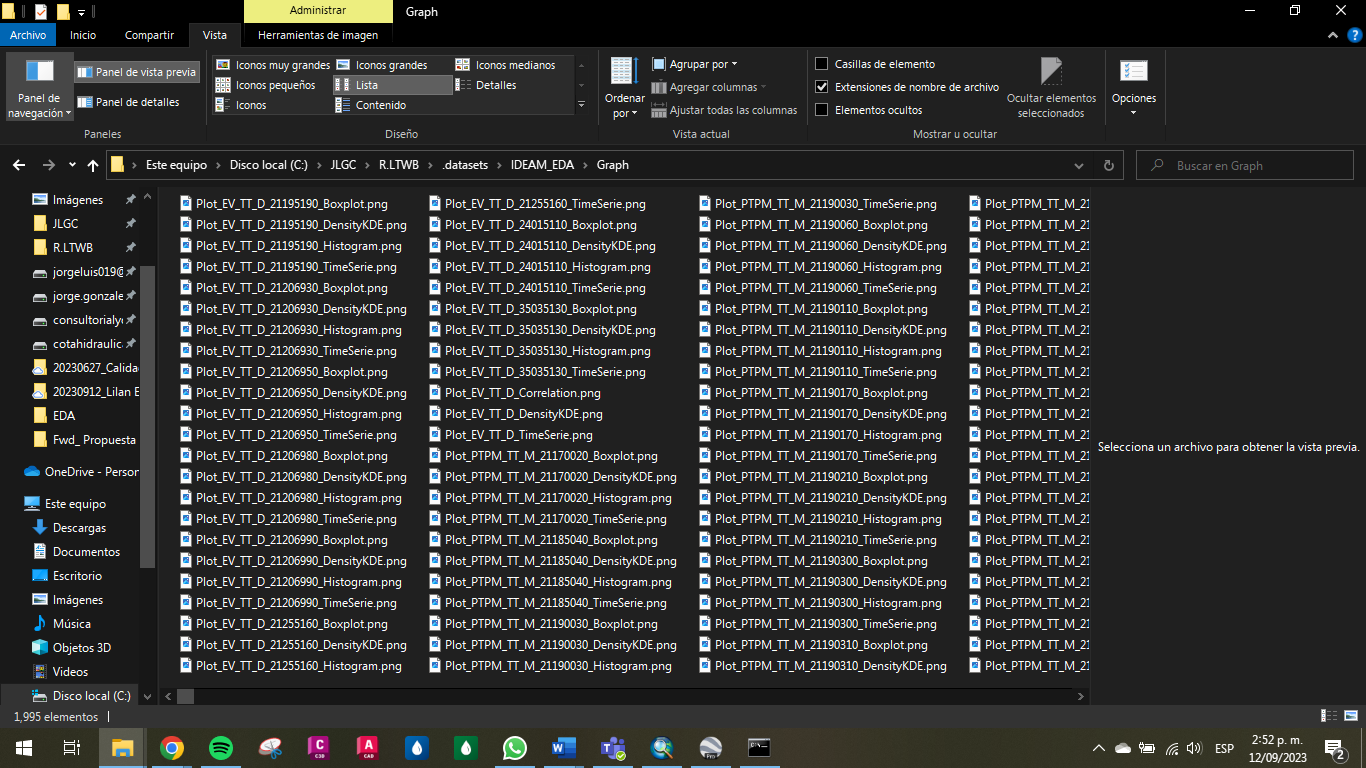


Fuente: Elaboración Propia, 2023.

Se verificó que en la carpeta /.datasets/IDEAM\_EDA se almacenaran los resultados del script en cuanto a gráficas, tablas y archivo de visualización en formato MarkDown.

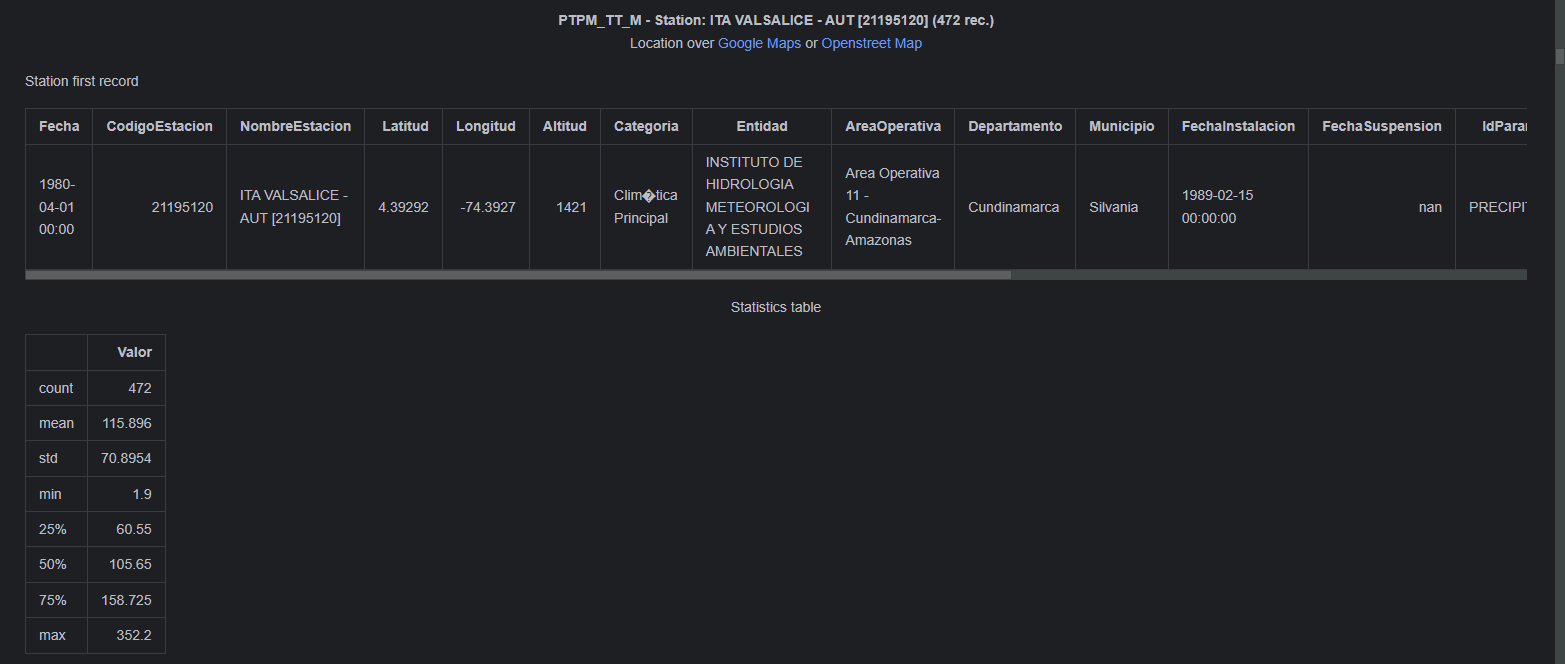
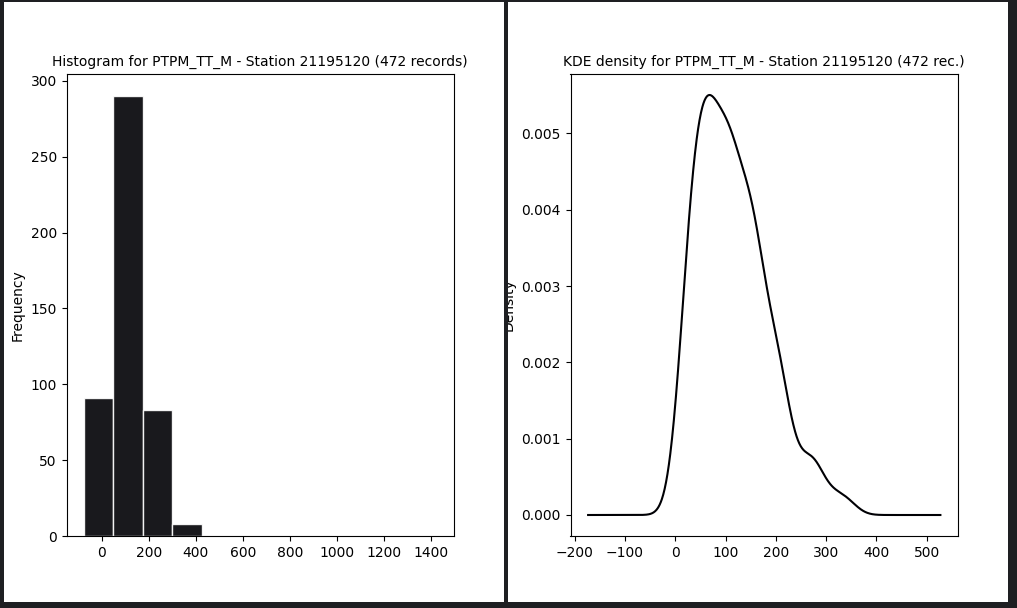
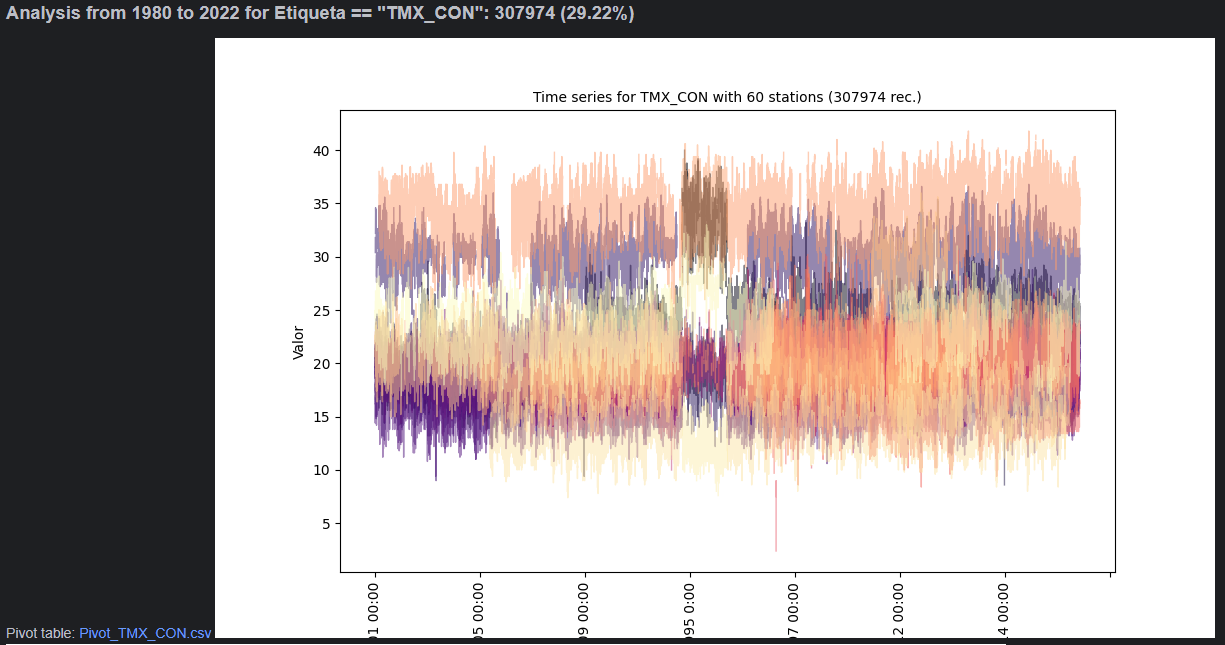
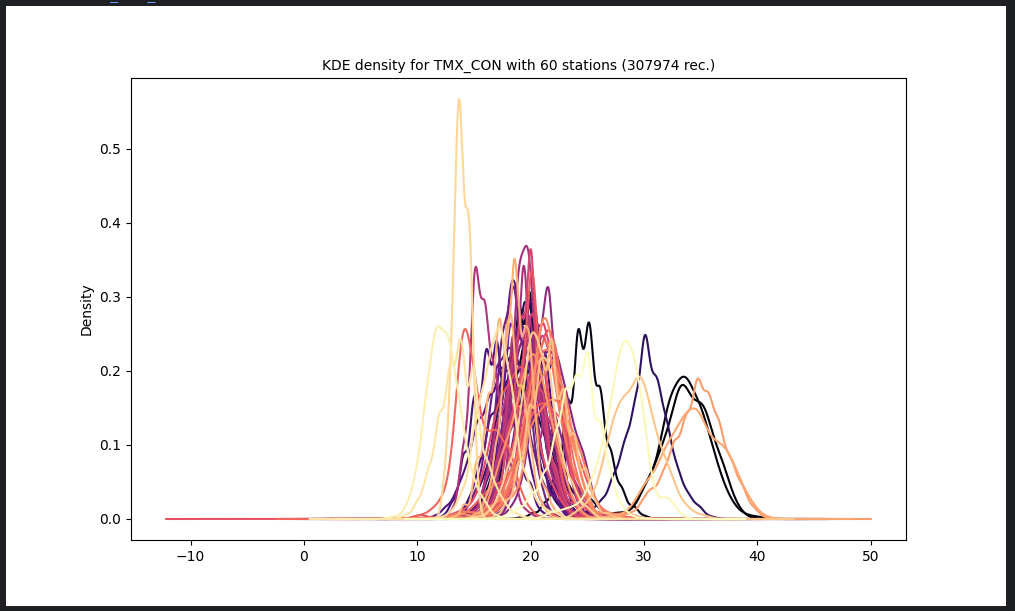
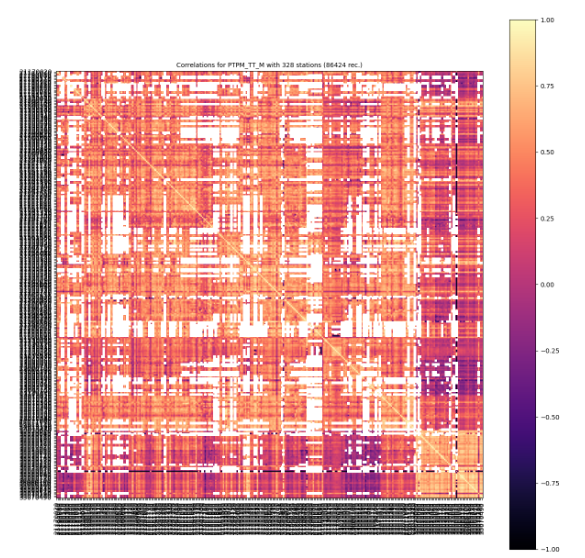
Ilustración ‑. Resultados script EDA.py





Fuente: Elaboración Propia, 2023.

Ilustración ‑. Visualización Resultados EDA.md

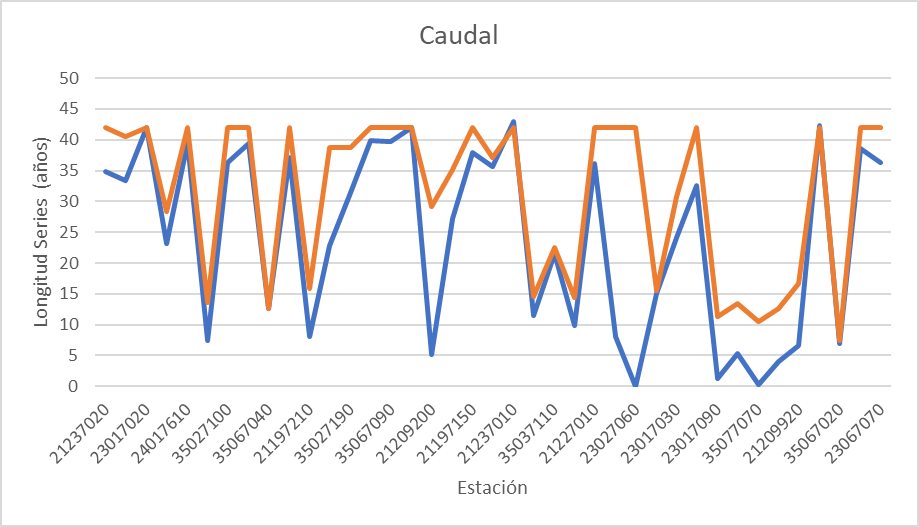
Fuente: Elaboración Propia, 2023.

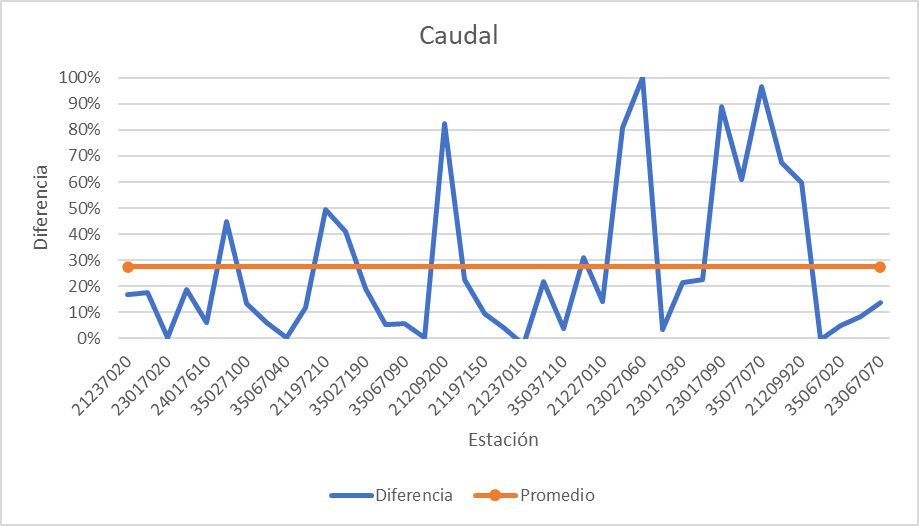
# ACTIVIDAD 2: análisis de longitud de series

Con los resultados de la ejecución del script EDA.py se puede revisar la longitud real de las series y compararla con la longitud hipotética realizada en la actividad CNEStation.

En el siguiente gráfico se observa que para las estaciones de caudal la longitud real de la serie en la mayoría de las estaciones es menor a la longitud hipotética y que el promedio de la diferencia entre valores es del 28% donde12 estaciones se encuentran por debajo de este.

Ilustración ‑. Comparación Estaciones Caudal

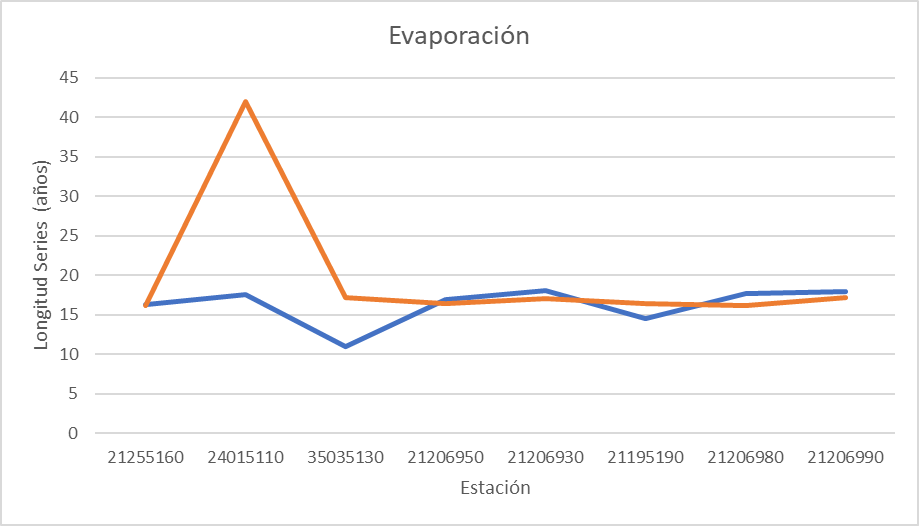


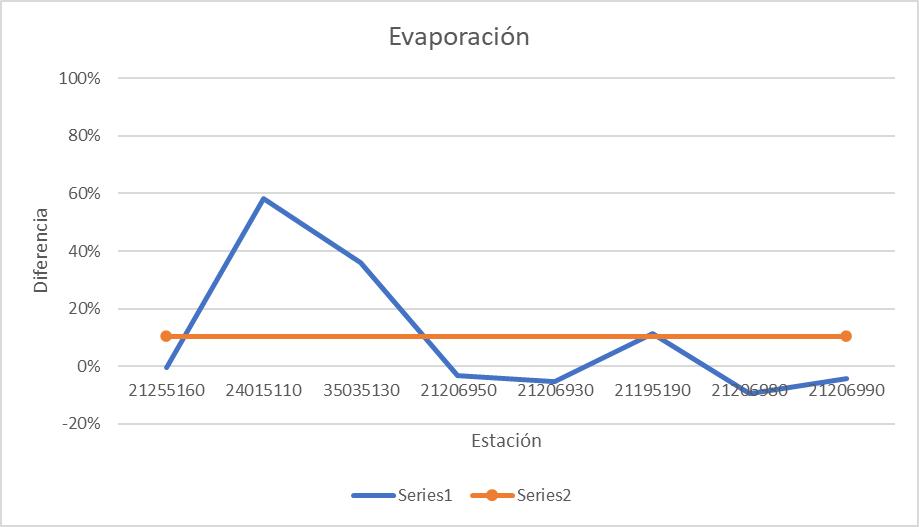


Fuente: Elaboración Propia, 2023.

En el siguiente gráfico se observa que para las estaciones de evaporación la longitud real de la serie en la mayoría de las estaciones es similar a la longitud hipotética y que el promedio de la diferencia entre valores es del 10% donde 5 estaciones se encuentran por debajo de este.

Ilustración ‑. Comparación Estaciones Evaporación

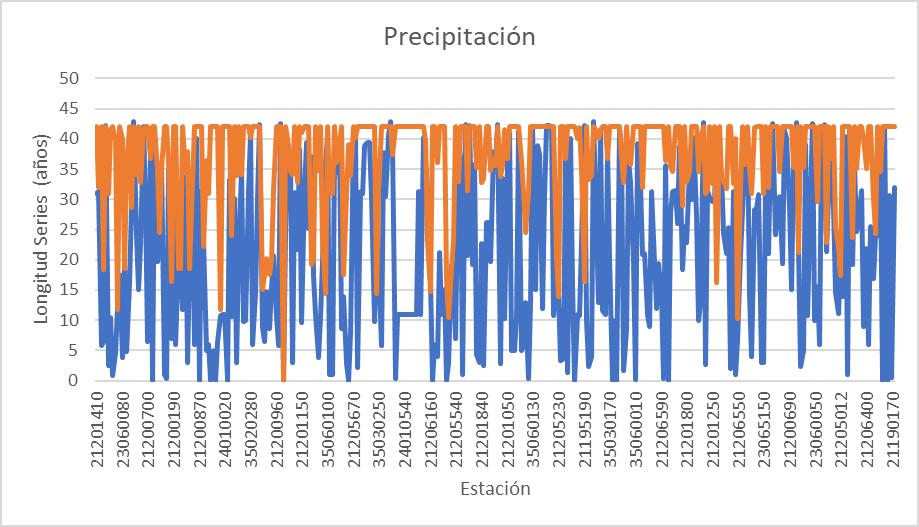


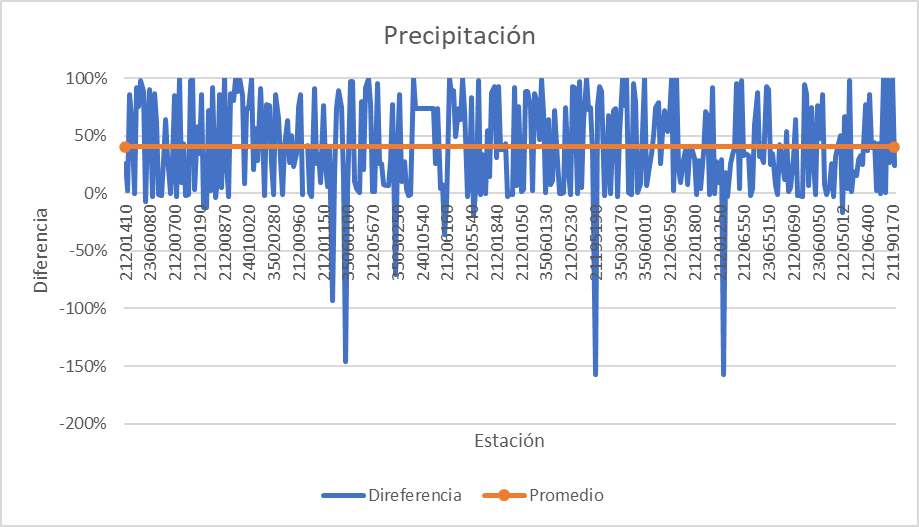


Fuente: Elaboración Propia, 2023.

En el siguiente gráfico se observa que para las estaciones de precipitación la longitud real de la serie en la mayoría de las estaciones es menor a la longitud hipotética y que el promedio de la diferencia entre valores es del 40% donde se observa que la serie parece simétrica respecto a este umbral con una gran cantidad de estaciones que se encuentran por encima identificándose que hay muchas diferencias.

Ilustración ‑. Comparación Estaciones Precipitación

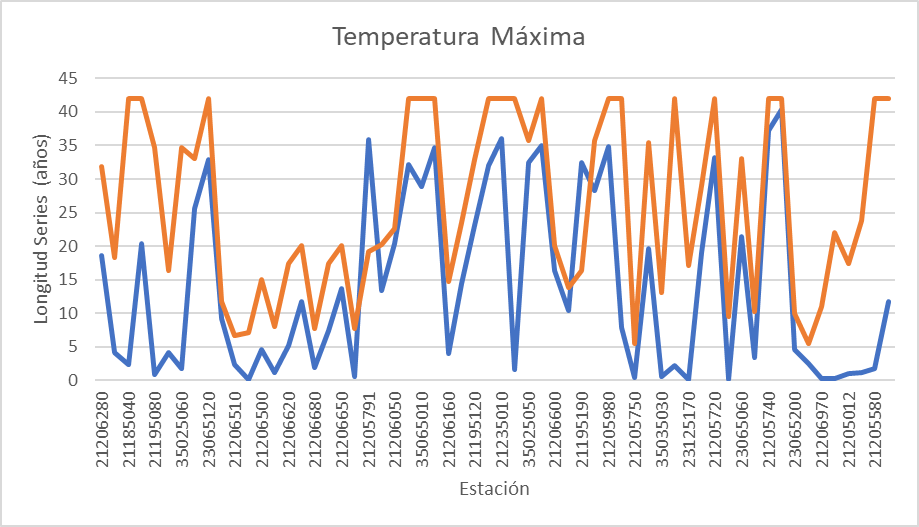
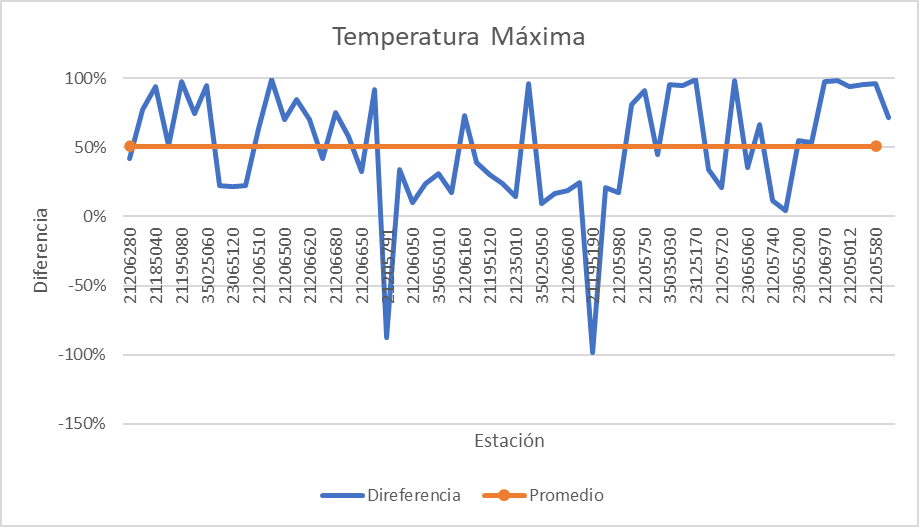




Fuente: Elaboración Propia, 2023.

En el siguiente gráfico se observa que para las estaciones de temperatura máxima la longitud real de la serie en la mayoría de las estaciones es menor a la longitud hipotética y que el promedio de la diferencia entre valores es del 51% donde 23 se encuentran por debajo de este.

Ilustración ‑. Comparación Estaciones Temperatura Máxima

Fuente: Elaboración Propia, 2023.

En el siguiente gráfico se observa que para las estaciones de temperatura mínima la longitud real de la serie en la mayoría de las estaciones es menor a la longitud hipotética y que el promedio de la diferencia entre valores es del 42% donde 30 se encuentran por debajo de este.

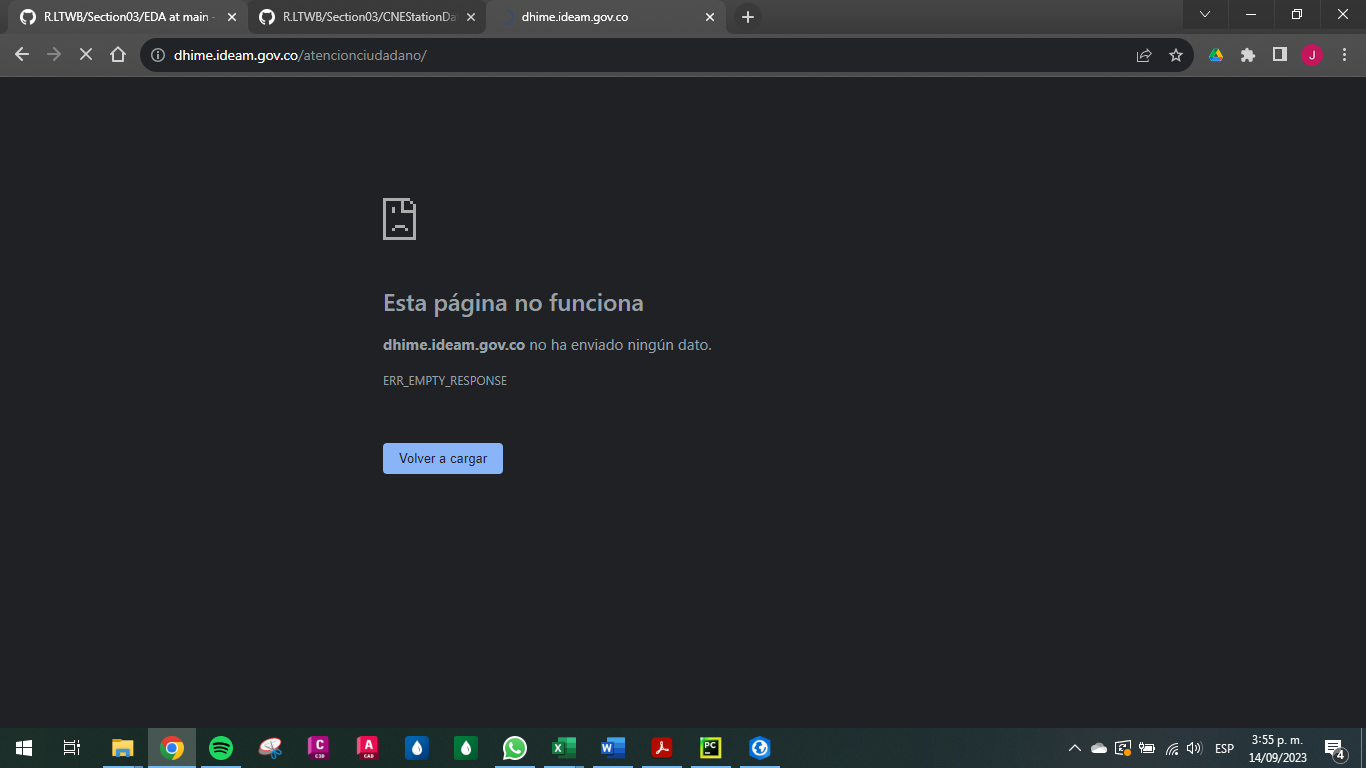
Ilustración ‑. Comparación Estaciones Temperatura Mínima

Fuente: Elaboración Propia, 2023.

# ACTIVIDAD 3: análisis de OTROS PARÁMETROS

A la fecha de desarrollo de esta actividad, la página web del IDEAM se encuentra deshabilitada por lo que no se pueden descargar los datos de otros parámetros.

Ilustración ‑. Acceso plataforma IDEAM



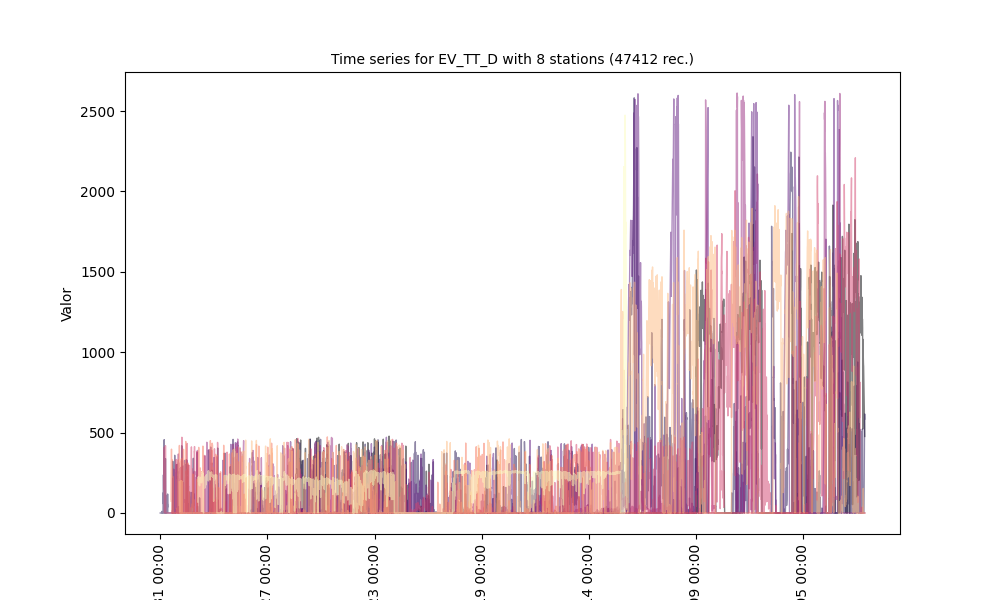
Fuente: IDEAM, 14/09/2023.

De acuerdo con los resultados del numeral anterior, se supone que en este caso también las longitudes reales de las series serán menores a las hipotéticas.

# ACTIVIDAD 4: análisis de resultados

En la siguiente ilustración se observa el resultado visual de los valores de los registros de las estaciones de evaporación.

Ilustración ‑. Registros Evaporación

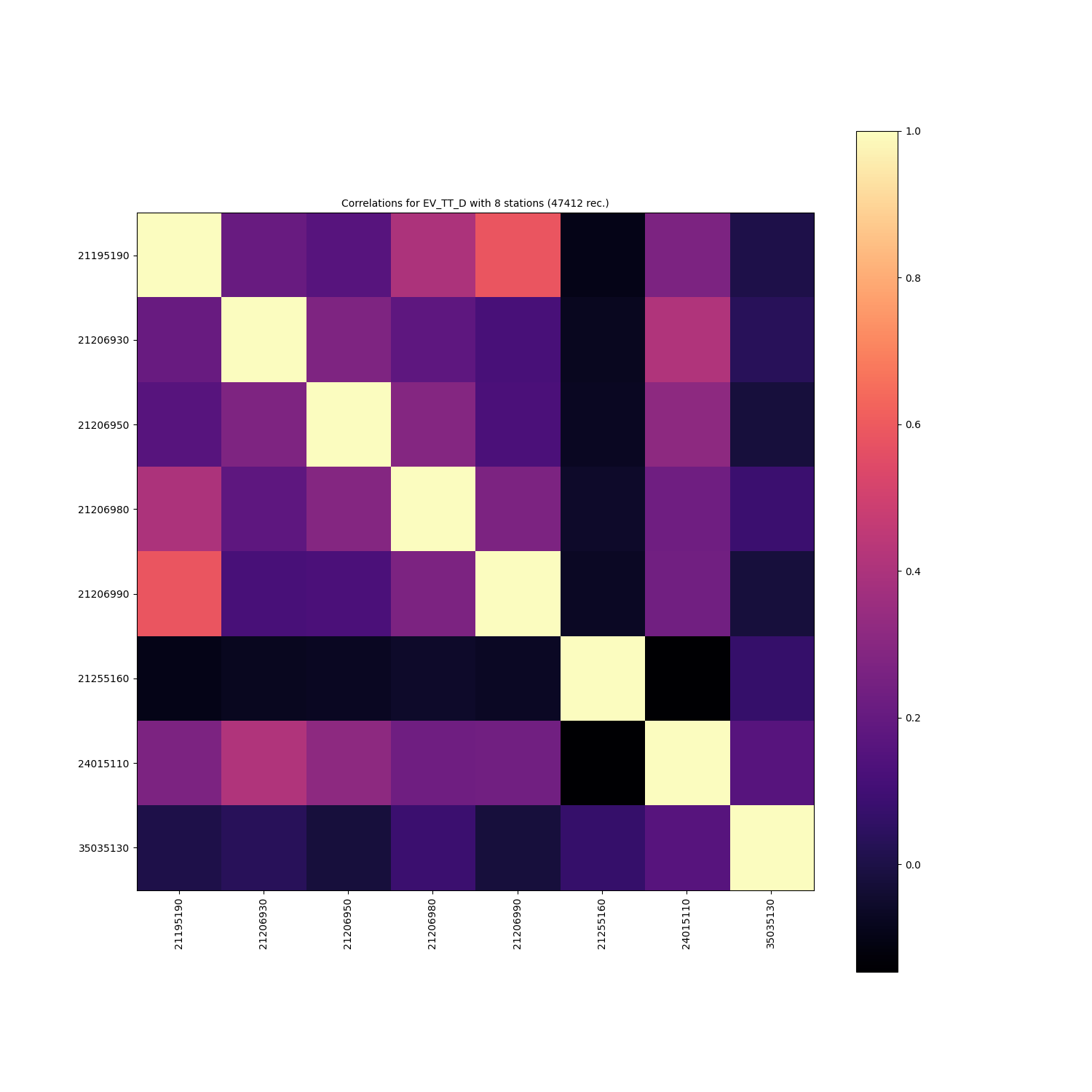


Fuente: Elaboración Propia, 2023.

Se observan dos grupos de valores, el primero todas las estaciones son similares, pero en el segundo se tienen valores diferentes respecto a la tendencia, incluso en su magnitud no parecen ser confiables para valores superiores a 500 mm de evaporación.

En cuanto a la correlación se observa que las estaciones no tienen buena correlación entre sí, con algunas más críticas las cuales son la 21255160 y la 35035130.

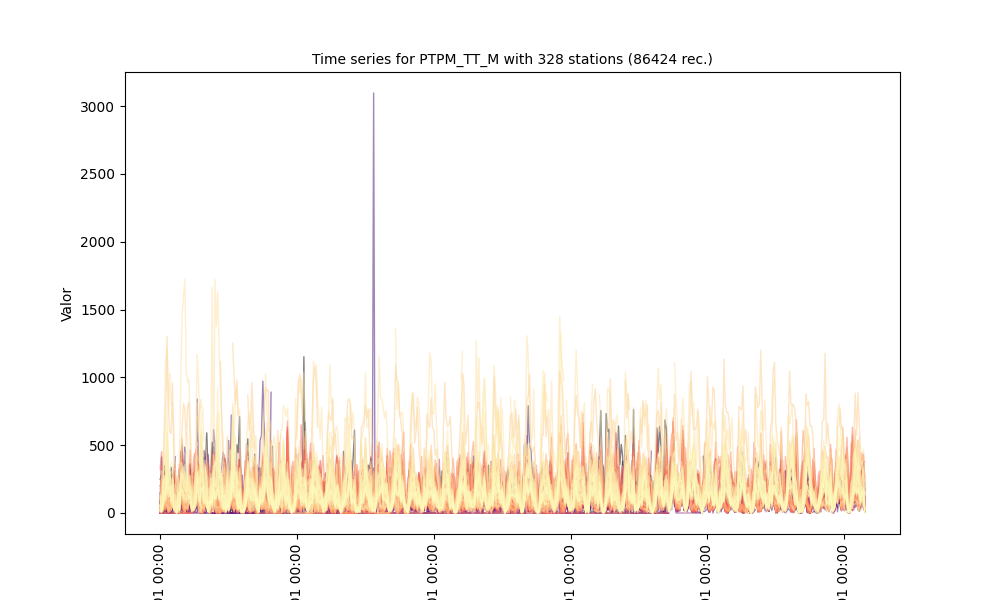
Ilustración ‑. Correlación Evaporación



Fuente: Elaboración Propia, 2023.

En la siguiente ilustración se observa el resultado visual de los valores de los registros de las estaciones de precipitación.

Ilustración ‑. Registros Precipitación

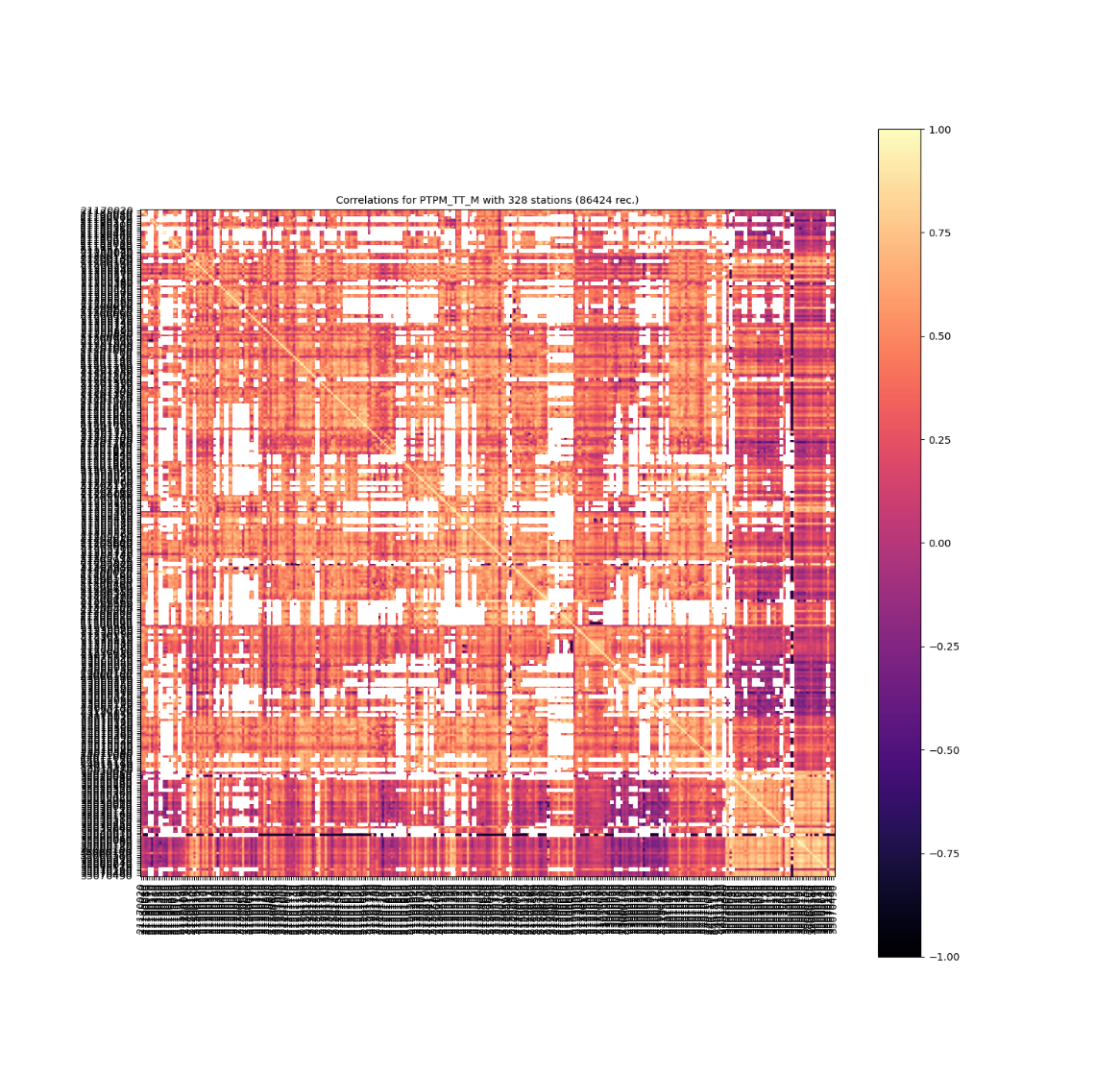


Fuente: Elaboración Propia, 2023.

Se observa que dos estaciones tienen valores por encima de manera muy notoria que las demás, una con valores claramente errados como lo es superior a 3000 mm.

En cuanto a la correlación se observa una estación que con esta comparación no debería ser utilizada para posteriores ejercicios, la cual es la 35060010, pero en términos generales la gran mayoría de estaciones tienen buena correlación.

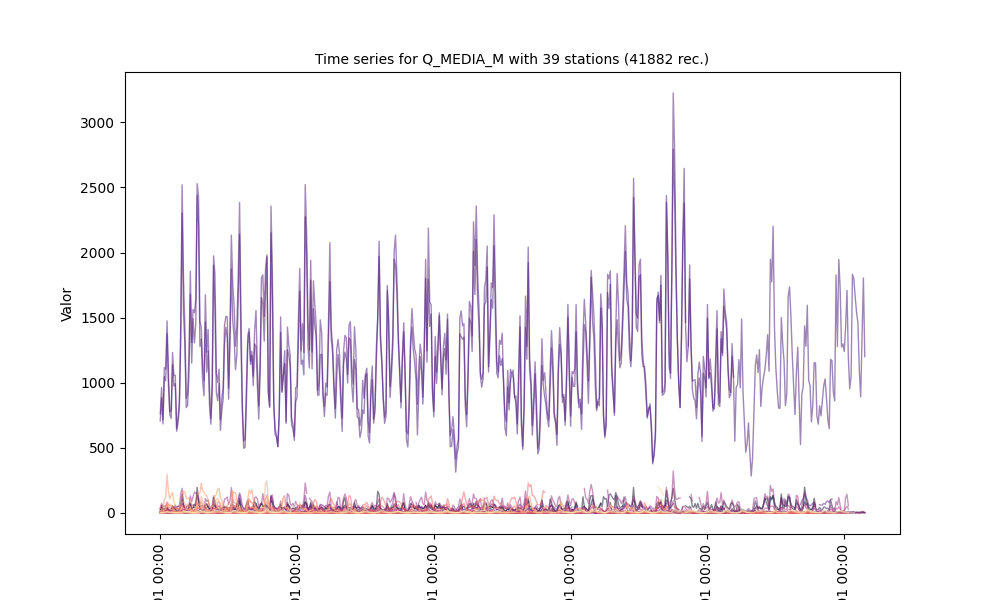
Ilustración ‑. Correlación Precipitación



Fuente: Elaboración Propia, 2023.

En la siguiente ilustración se observa el resultado visual de los valores de los registros de las estaciones de caudal.

Ilustración ‑. Registros Caudal

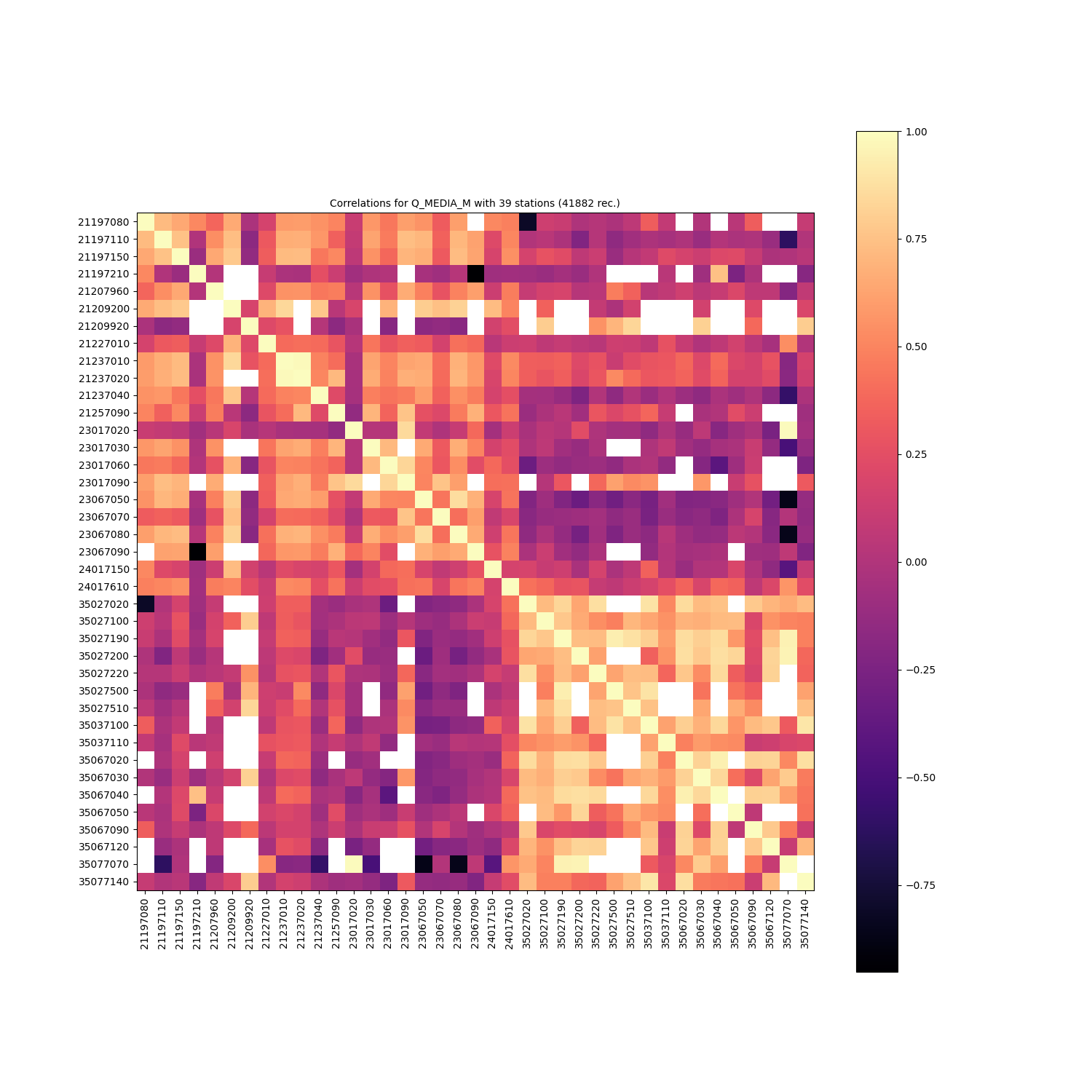


Fuente: Elaboración Propia, 2023.

Se observa que una estación tiene valores por encima de manera muy notoria que las demás, sin embargo, este no es un motivo de descarte ya que en este caso los valores dependen del tamaño del área de drenaje.

En cuanto a la correlación se observa que, se tiene una proporcionalidad entre buenas y malas correlaciones, sin embargo, al ser una variable que depende del área de drenaje no se puede dar un concepto de eliminación de alguna de ellas.

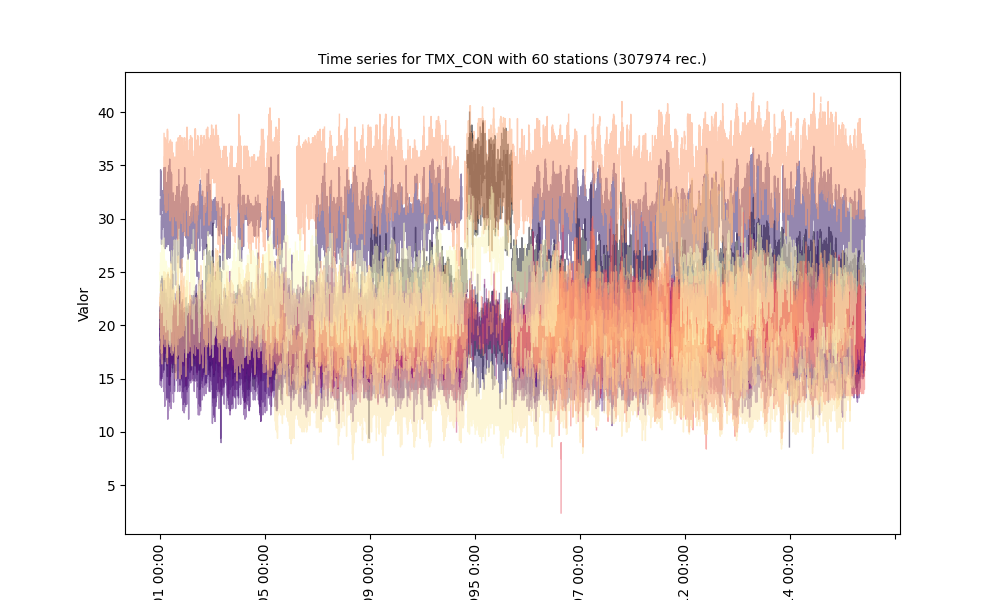
Ilustración ‑. Correlación Caudal



Fuente: Elaboración Propia, 2023.

En la siguiente ilustración se observa el resultado visual de los valores de los registros de las estaciones de temperatura máxima.

Ilustración ‑. Registros Temperatura Máxima

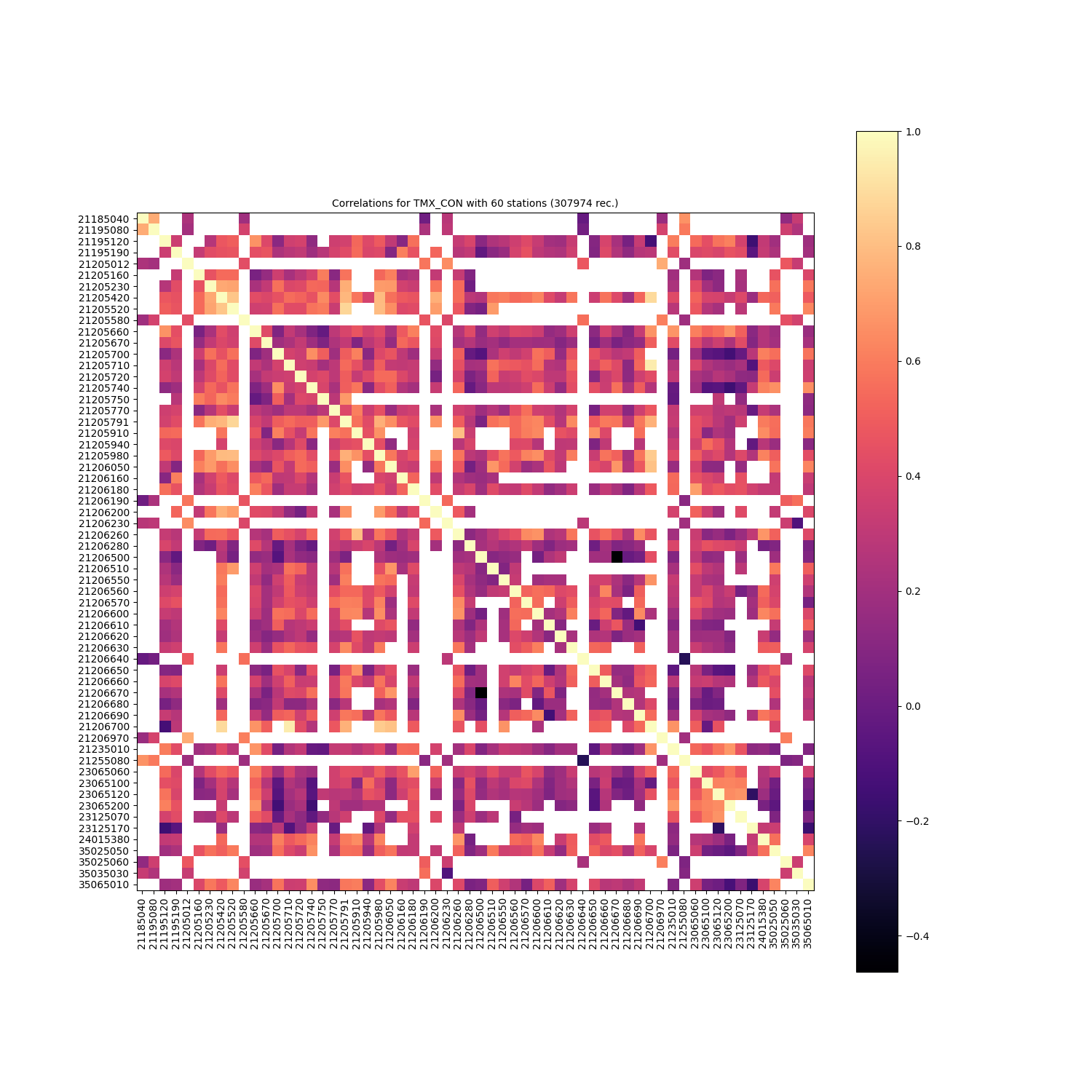


Fuente: Elaboración Propia, 2023.

Se observa en general las estaciones se encuentran en el mismo rango de registro a excepción de una que tiene un dato atípico registrado menor a 5 °C respecto a las demás.

En cuanto a la correlación se observa que la gran mayoría de estaciones no tienen buena correlación.

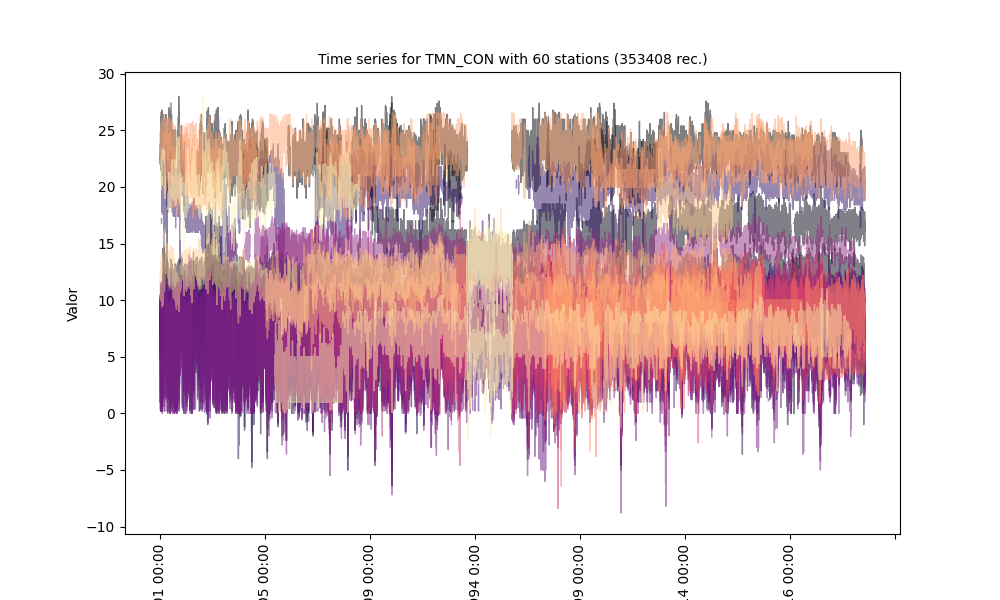
Ilustración ‑. Correlación Temperatura Máxima



Fuente: Elaboración Propia, 2023.

En la siguiente ilustración se observa el resultado visual de los valores de los registros de las estaciones de temperatura mínima.

Ilustración ‑. Registros Temperatura Mínima

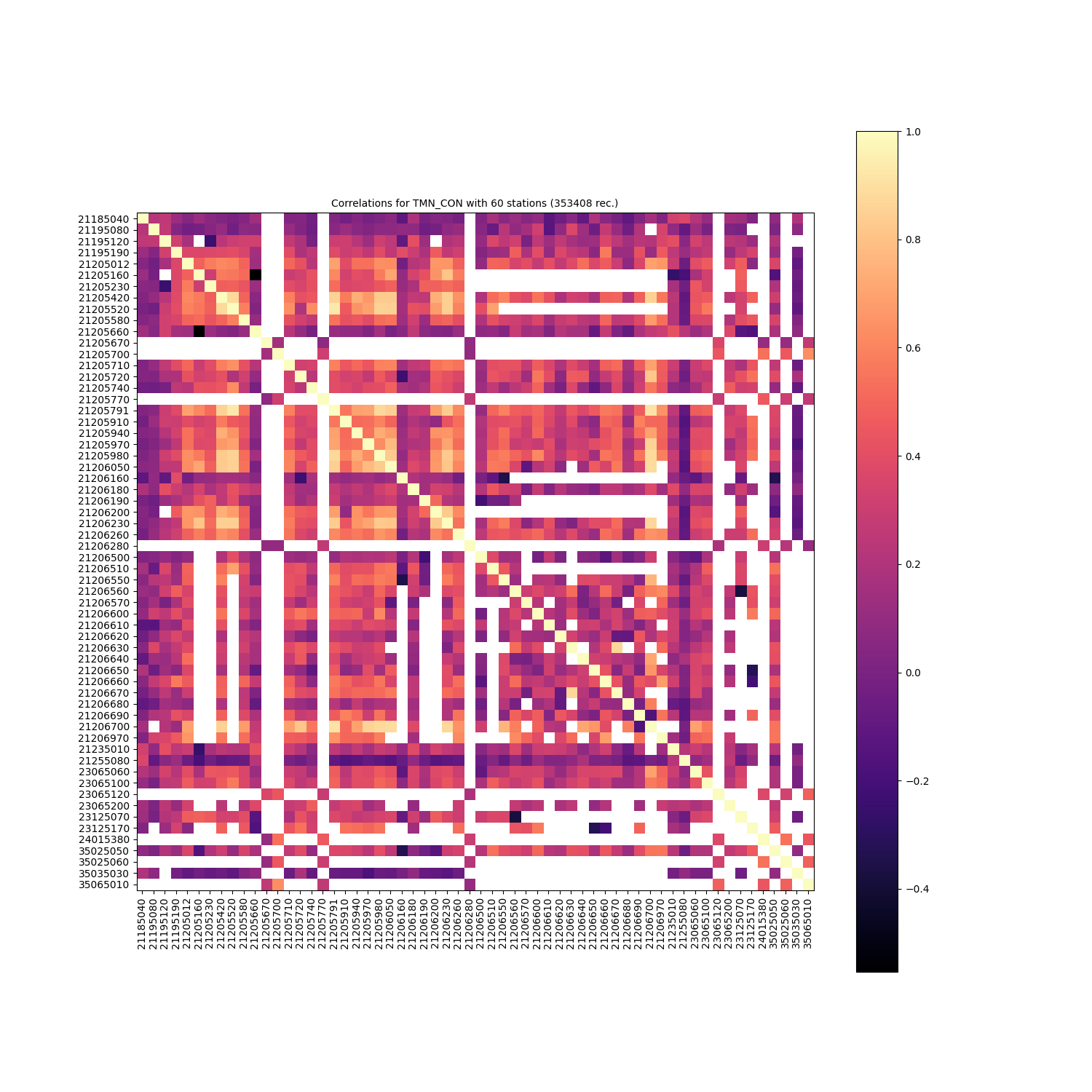


Fuente: Elaboración Propia, 2023.

Se observa que algunas estaciones tienen valores negativos los que serían errados respecto a la distribución espacial de las demás y deberían ser eliminadas.

En cuanto a la correlación se observa que la gran mayoría de estaciones no tienen buena correlación.

Ilustración ‑. Correlación Temperatura Mínima



Fuente: Elaboración Propia, 2023.

# Conclusiones

* Se realizó la ejecución del script de análisis estadística EDA.py para los datos descargados de la plataforma de información del IDEAM.
* Se realizó la comparación entre los datos de la actividad CNEStation de longitud hipotética respecto a la longitud real del dataset descargado del IDEAM.
* A pesar que el catalogo de estaciones del IDEAM indique que se tiene una longitud hipotética de varios años, para este caso la longitud real es mucho menor.
* En lo anterior es importante tener en cuenta que la longitud real calculada corresponde a la equivalencia en años de la cantidad de registros mas no significa que se tengan registros completos en ese valor, es decir, la serie puede tener vacíos interanuales.
* Se realizó el análisis cualitativo de los resultados de los valores de las series y su correlación.
* No se pude realizar la actividad con otros parámetros ya que a la fecha la plataforma de descarga del IDEAM no se encuentra activa.

# Referencias Bibliográficas

* RCFDTOOLS, 2023. Balance hidrológico de largo plazo para estimación de caudales medios usando SIG. Contenido del curso: https://github.com/rcfdtools/R.LTWB/tree/main/Section03/EDA