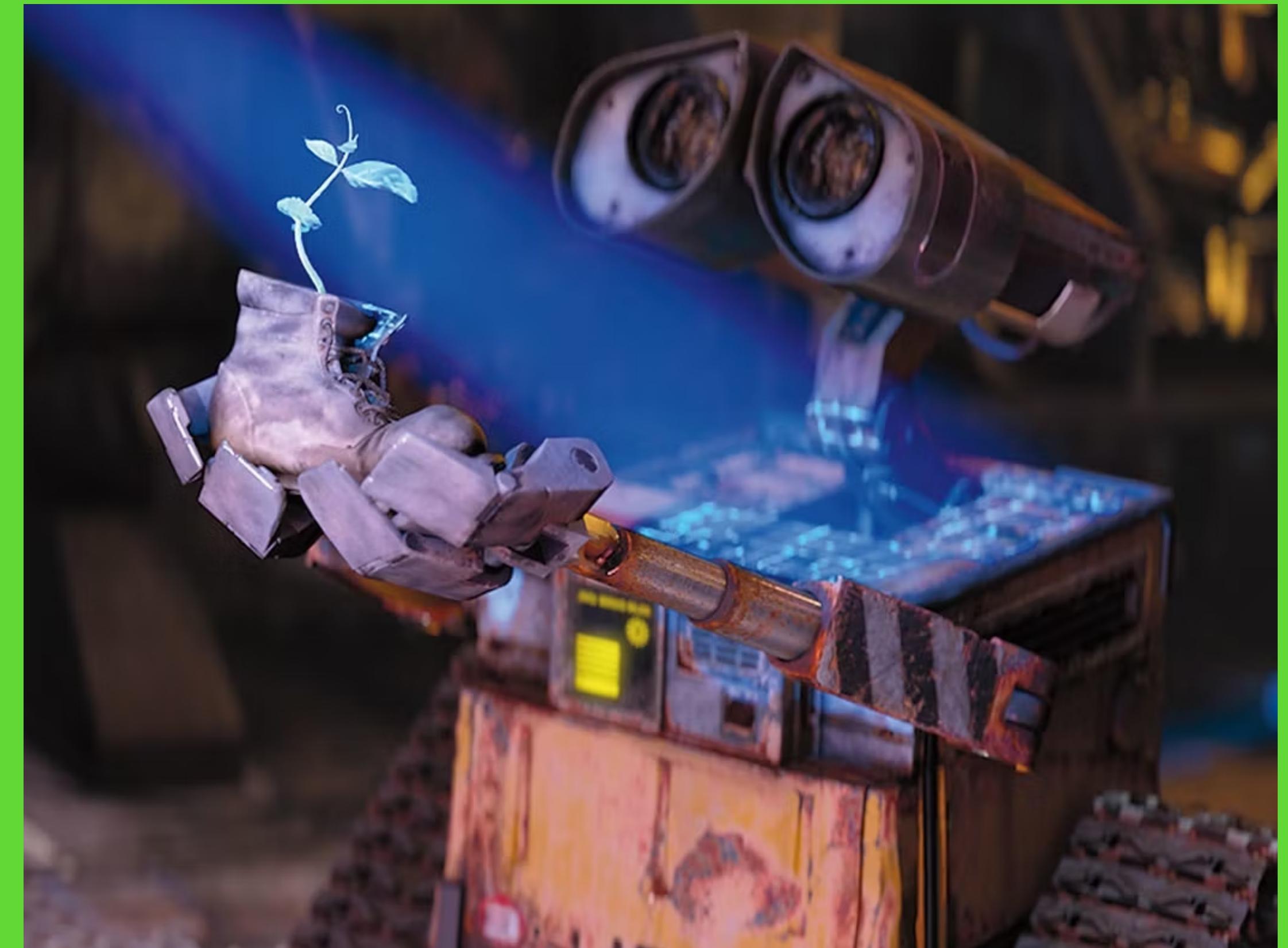


PREDICCIÓN DE LA TEMPERATURA MÁXIMA MENSUAL CON SARIMAX.

(PROJECT WALL.E)

Angel Vazquez

Bootcamp Data Science, IT Academy



OBJETIVOS

- Recopilar datos
- Selección de columnas
- Sustitución valores NaN
- Trasnformación Outliers
- Búsqueda del modelo
- Entrenamiento
- Predicción

De dónde viene

https://www.aemet.es/es/datos_abiertos/estadisticas/estadistica_meteorofenologicas

[PDF](#) Generalidades (389 KB)

[PDF](#) Parámetros (225 KB)

[PDF](#) Formatos (150 KB)

[ZIP](#) Estadística anual 2023 (859 KB)

[ZIP](#) Estadística anual 2022 (861 KB)

[ZIP](#) Estadística anual 2021 (848 KB)

[ZIP](#) Estadística anual 2020 (846 KB)

[ZIP](#) Estadística anual 2019 (846 KB)

[ZIP](#) Estadística anual 2018 (847 KB)

[ZIP](#) Estadística anual 2017 (848 KB)

[ZIP](#) Estadística anual 2016 (846 KB)

- Datos recopilados entre 2016 y 2013
- 44 variables
- 95 estaciones meteorológicas



ESTRUCTURA DEL DATAFRAME

'df_indicativos'

| | INDICATIVO | NOMBRE | PROVINCIA | ALTITUD | LATITUD | LONGITUD |
|-----|------------|-----------------------------------|------------|---------|----------|----------|
| 0 | 1387 | A CORUÑA | A CORUÑA | 58 | 432157 N | 082517 W |
| 1 | 1387E | A CORUÑA, ALVEDRO | A CORUÑA | 98 | 431825 N | 082219 W |
| 2 | 1428 | SANTIAGO DE COMPOSTELA, LABACOLLA | A CORUÑA | 370 | 425317 N | 082438 W |
| 3 | 8178D | ALBACETE, OBS. | ALBACETE | 674 | 390020 N | 015144 W |
| 4 | 8175 | ALBACETE, LOS LLANOS | ALBACETE | 702 | 385715 N | 015123 W |
| ... | ... | ... | ... | ... | ... | ... |
| 90 | 2422 | VALLADOLID | VALLADOLID | 735 | 413827 N | 044516 W |
| 91 | 2539 | VALLADOLID, VILLANUBLA | VALLADOLID | 846 | 414243 N | 045120 W |
| 92 | 2614 | ZAMORA | ZAMORA | 656 | 413056 N | 054407 W |
| 93 | 9390 | DAROCA I | ZARAGOZA | 779 | 410652 N | 012436 W |
| 94 | 9434 | ZARAGOZA, AEROPUERTO | ZARAGOZA | 249 | 413938 N | 010015 W |

95 rows × 6 columns

'fecha_df'

| | INDICATIVO | FECHA |
|------|------------|------------|
| 0 | 1387 | 2016-01-01 |
| 1 | 1387 | 2016-02-01 |
| 2 | 1387 | 2016-03-01 |
| 3 | 1387 | 2016-04-01 |
| 4 | 1387 | 2016-05-01 |
| ... | ... | ... |
| 9115 | 9434 | 2023-08-01 |
| 9116 | 9434 | 2023-09-01 |
| 9117 | 9434 | 2023-10-01 |
| 9118 | 9434 | 2023-11-01 |
| 9119 | 9434 | 2023-12-01 |

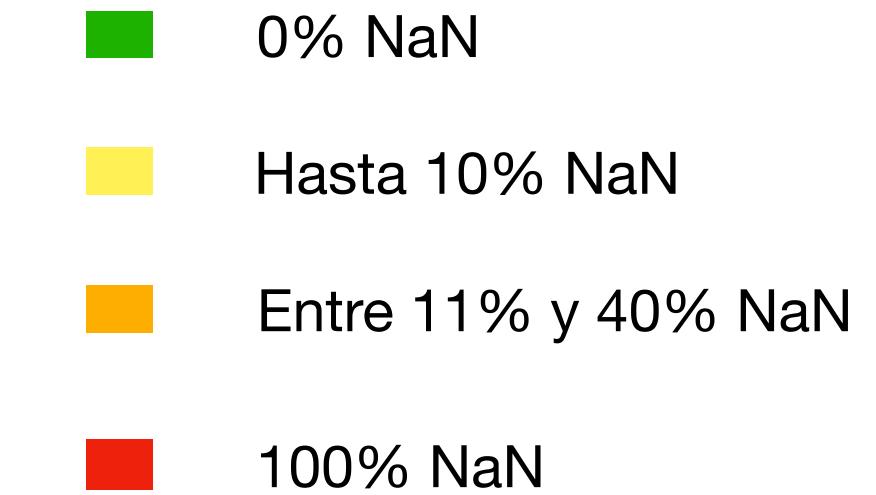
9120 rows × 2 columns

DATAFRAME RESULTANTE

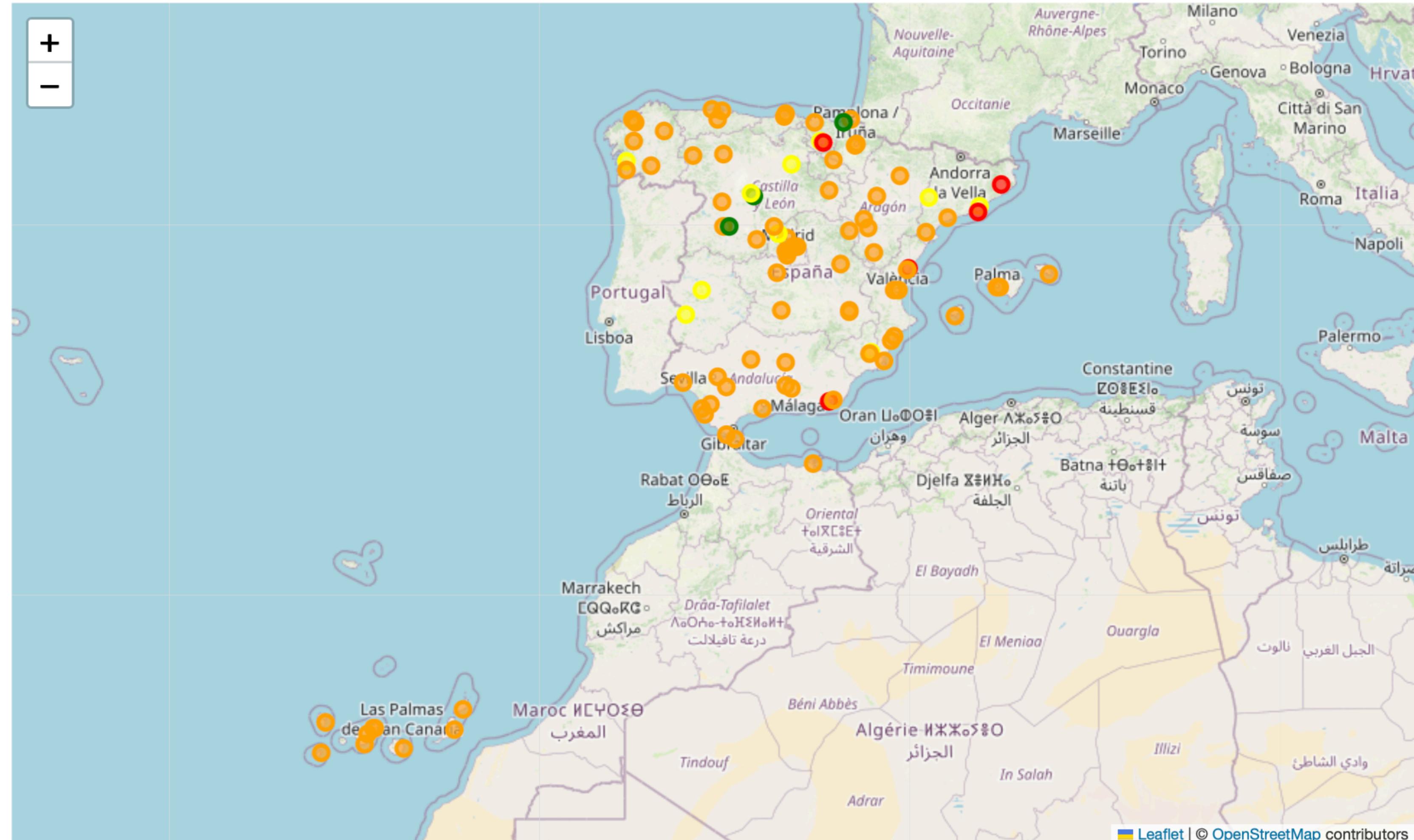
9120 filas y 25 columnas

| | INDICATIVO | FECHA | DIAS_GRA | DIAS_LLU | DIAS_NIE | DIAS_NUB | DIAS_SOL | EVAPORACION | HUMEDAD | INSOLACION% | ... | TEMP_MAX | TEMP_MEDIA_MAX |
|--|------------|------------|----------|----------|----------|----------|----------|-------------|---------|-------------|-----|----------|----------------|
| | 1387 | 2016-01-01 | 0.0 | 25.0 | 0.0 | NaN | NaN | 551.0 | 79.0 | 23.0 | ... | 20.5 | 14.6 |
| | 1387 | 2016-02-01 | 2.0 | 20.0 | 0.0 | NaN | NaN | 645.0 | 80.0 | 25.0 | ... | 16.9 | 13.3 |
| | 1387 | 2016-03-01 | 0.0 | 20.0 | 0.0 | NaN | NaN | 0.0 | 76.0 | 44.0 | ... | 17.1 | 13.9 |
| | 1387 | 2016-04-01 | 0.0 | 18.0 | 0.0 | NaN | NaN | 637.0 | 77.0 | 0.0 | ... | 20.0 | 15.0 |
| | 1387 | 2016-05-01 | 0.0 | 20.0 | 0.0 | NaN | NaN | 603.0 | 79.0 | 0.0 | ... | 25.2 | 18.4 |
| | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| | 9434 | 2023-08-01 | 0.0 | 2.0 | 0.0 | 16.0 | 15.0 | 0.0 | 0.0 | 86.0 | ... | 41.7 | 34.3 |
| | 9434 | 2023-09-01 | 1.0 | 9.0 | 0.0 | 18.0 | 7.0 | 0.0 | 0.0 | 68.0 | ... | 34.4 | 29.3 |
| | 9434 | 2023-10-01 | 0.0 | 13.0 | 0.0 | 17.0 | 7.0 | 2068.0 | 0.0 | 61.0 | ... | 33.9 | 25.6 |
| | 9434 | 2023-11-01 | 0.0 | 8.0 | 0.0 | 22.0 | 8.0 | 0.0 | 0.0 | 68.0 | ... | 24.5 | 18.7 |
| | 9434 | 2023-12-01 | 0.0 | 4.0 | 0.0 | 22.0 | 6.0 | 0.0 | 0.0 | 66.0 | ... | 20.1 | 12.8 |

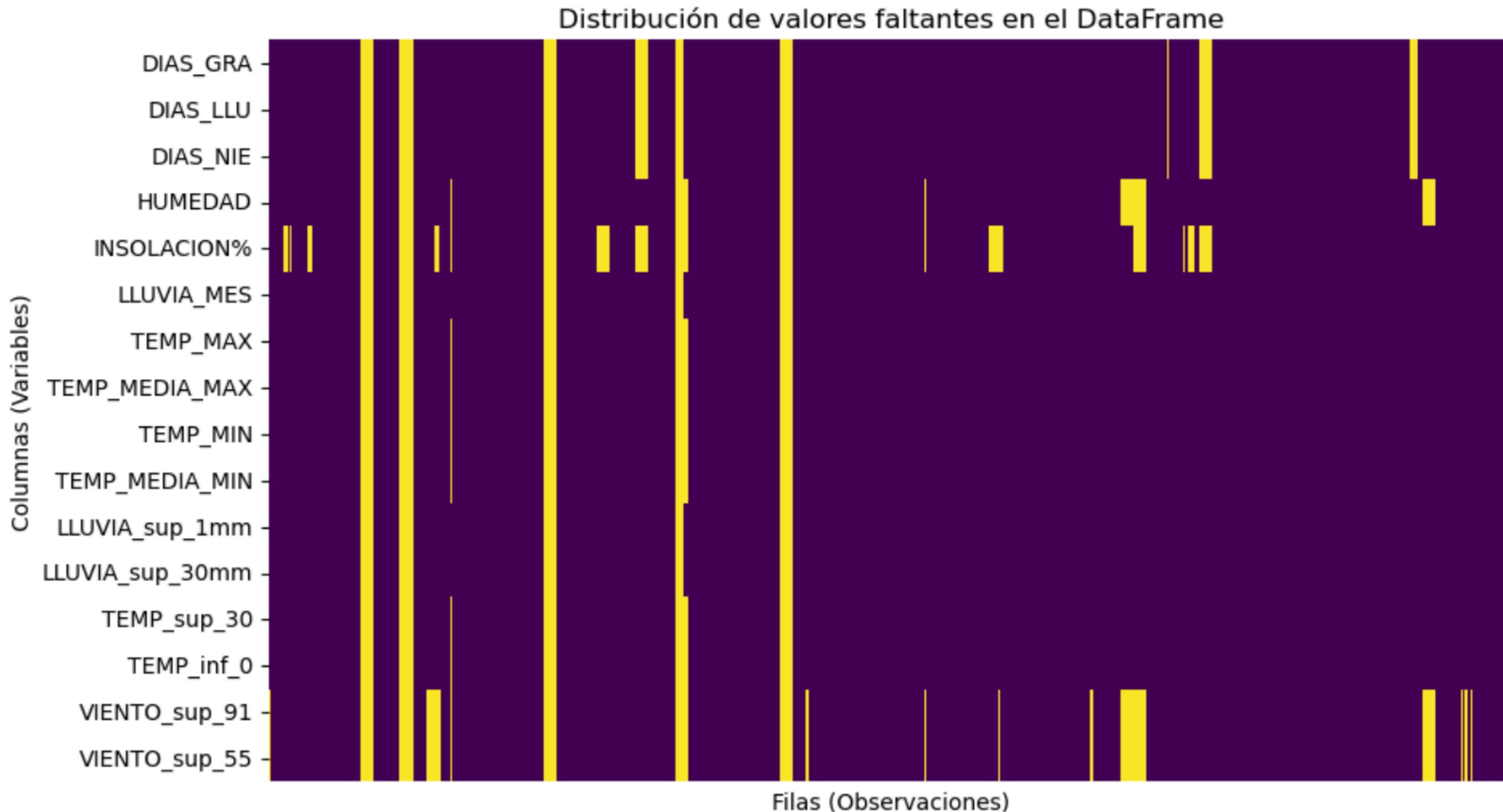
OBSERVACIÓN DE VALORES FALTANTES



| | |
|-----------------|-------|
| INDICATIVO | 0 |
| AÑO | 0 |
| MES | 0 |
| DIAS_GRA | 708 |
| DIAS_LLU | 708 |
| DIAS_NIE | 708 |
| DIAS_NUB | 2688 |
| DIAS_SOL | 2688 |
| EVAPORACION | 3096 |
| HUMEDAD | 792 |
| INSOLACION% | 1164 |
| LLUVIA_MES | 444 |
| RADIACION | 4932 |
| TEMP_10cm | 8016 |
| TEMP_20cm | 7296 |
| TEMP_50cm | 8196 |
| TEMP_MAX | 492 |
| TEMP_MEDIA_MAX | 492 |
| TEMP_MIN | 492 |
| TEMP_MEDIA_MIN | 492 |
| LLUVIA_sup_1mm | 444 |
| LLUVIA_sup_30mm | 444 |
| TEMP_sup_30 | 492 |
| TEMP_inf_0 | 492 |
| VIENTO_sup_91 | 1020 |
| VIENTO_sup_55 | 1020 |
| dtype: | int64 |



TRATAMIENTO VALORES NaN

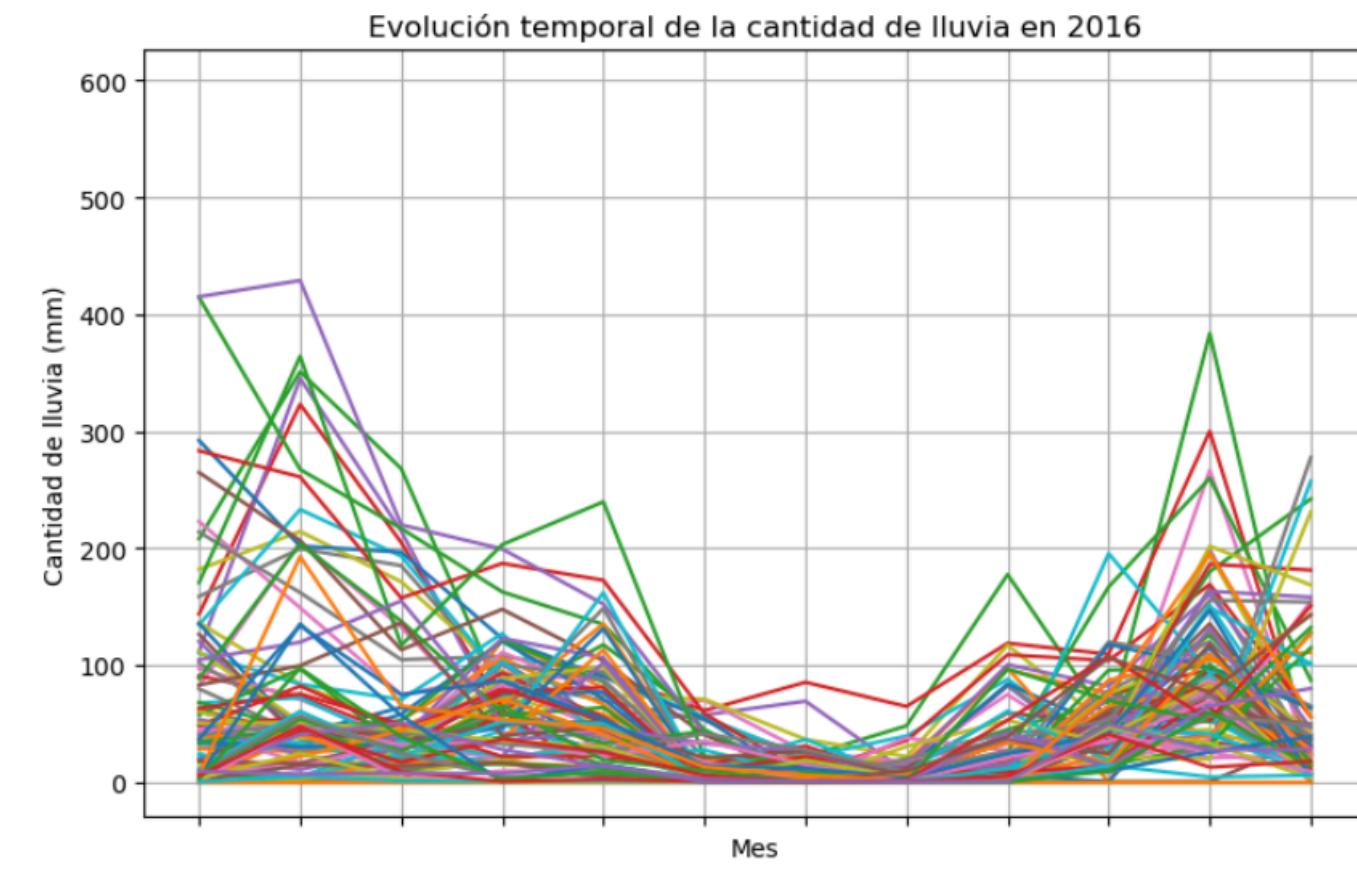


No se pueden
eliminar filas para
no alterar la serie
temporal

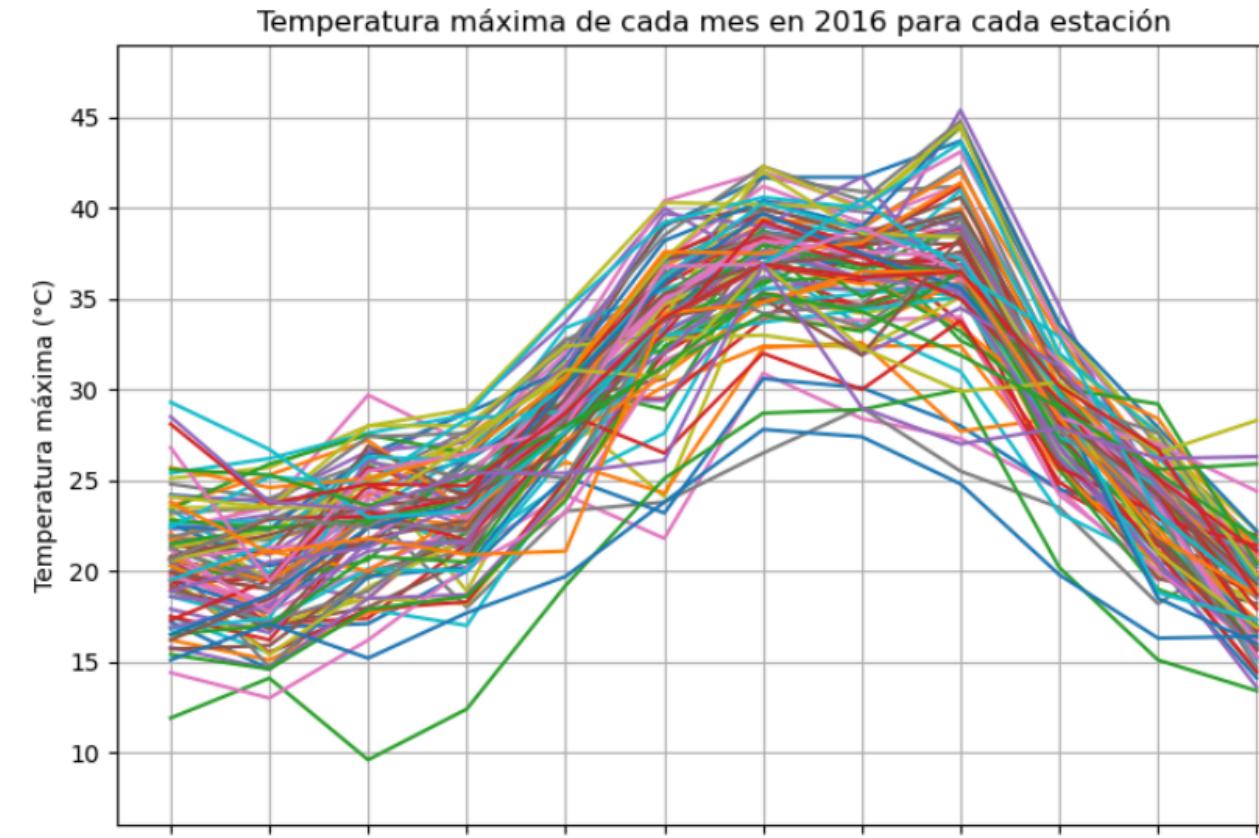
Sustitución de valores
faltantes por mediana.

SELECCIÓN VARIABLE OBJETIVO

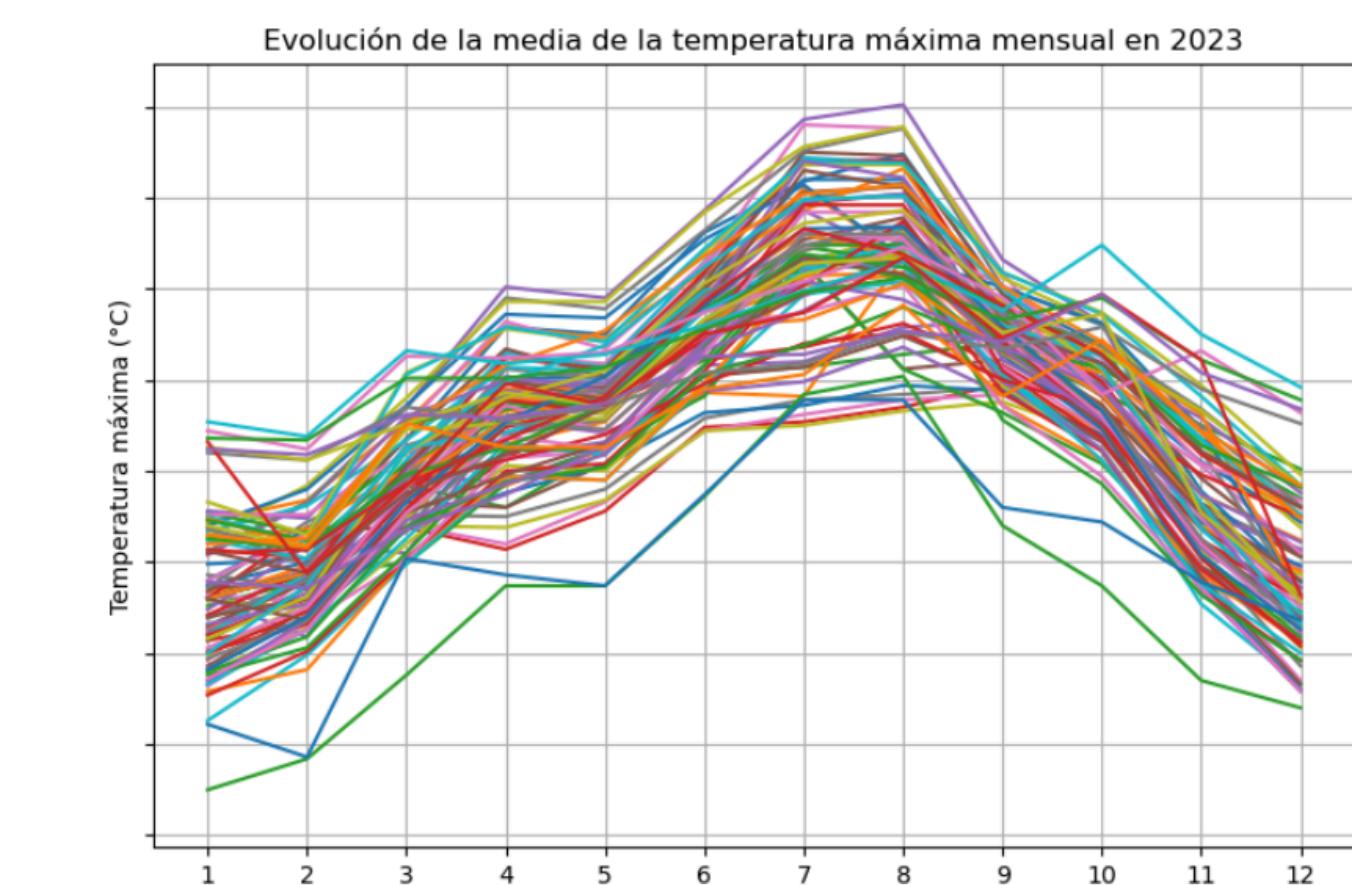
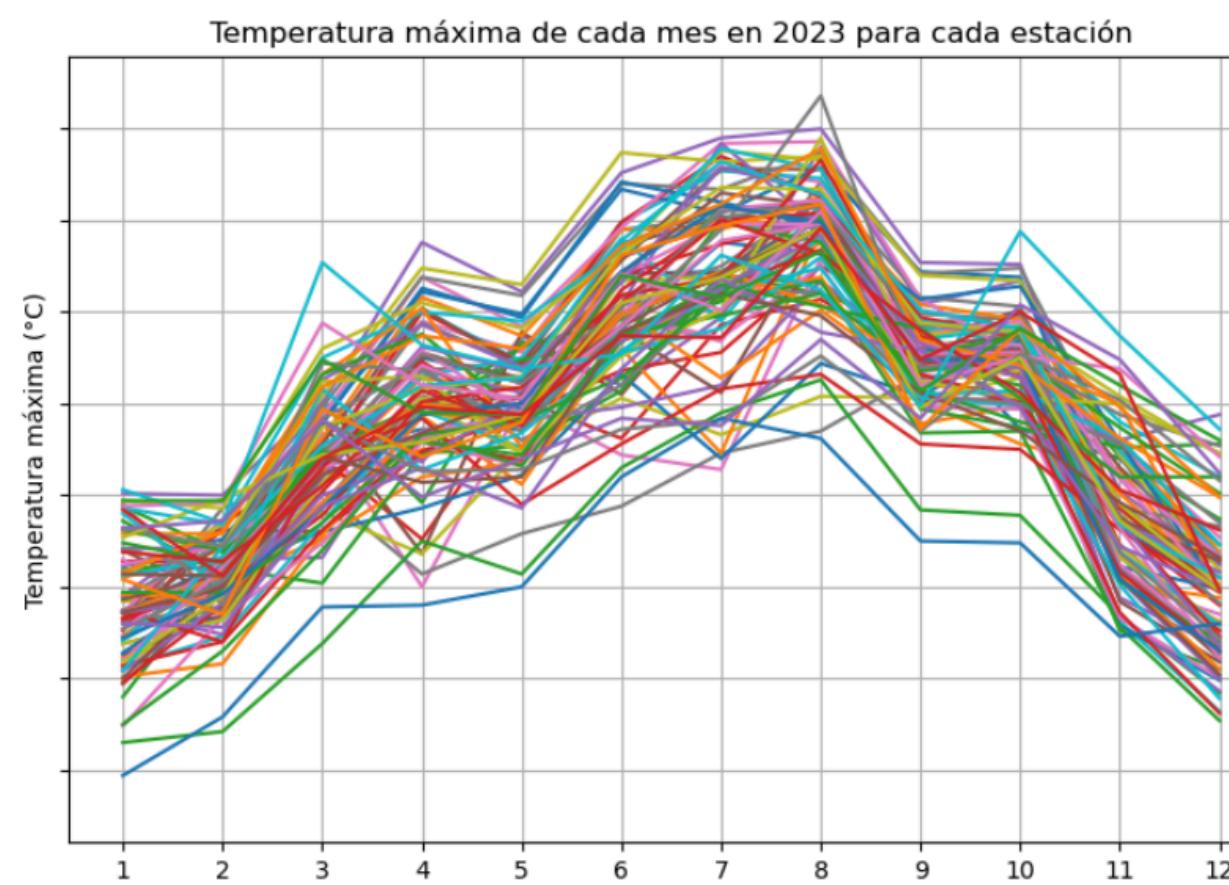
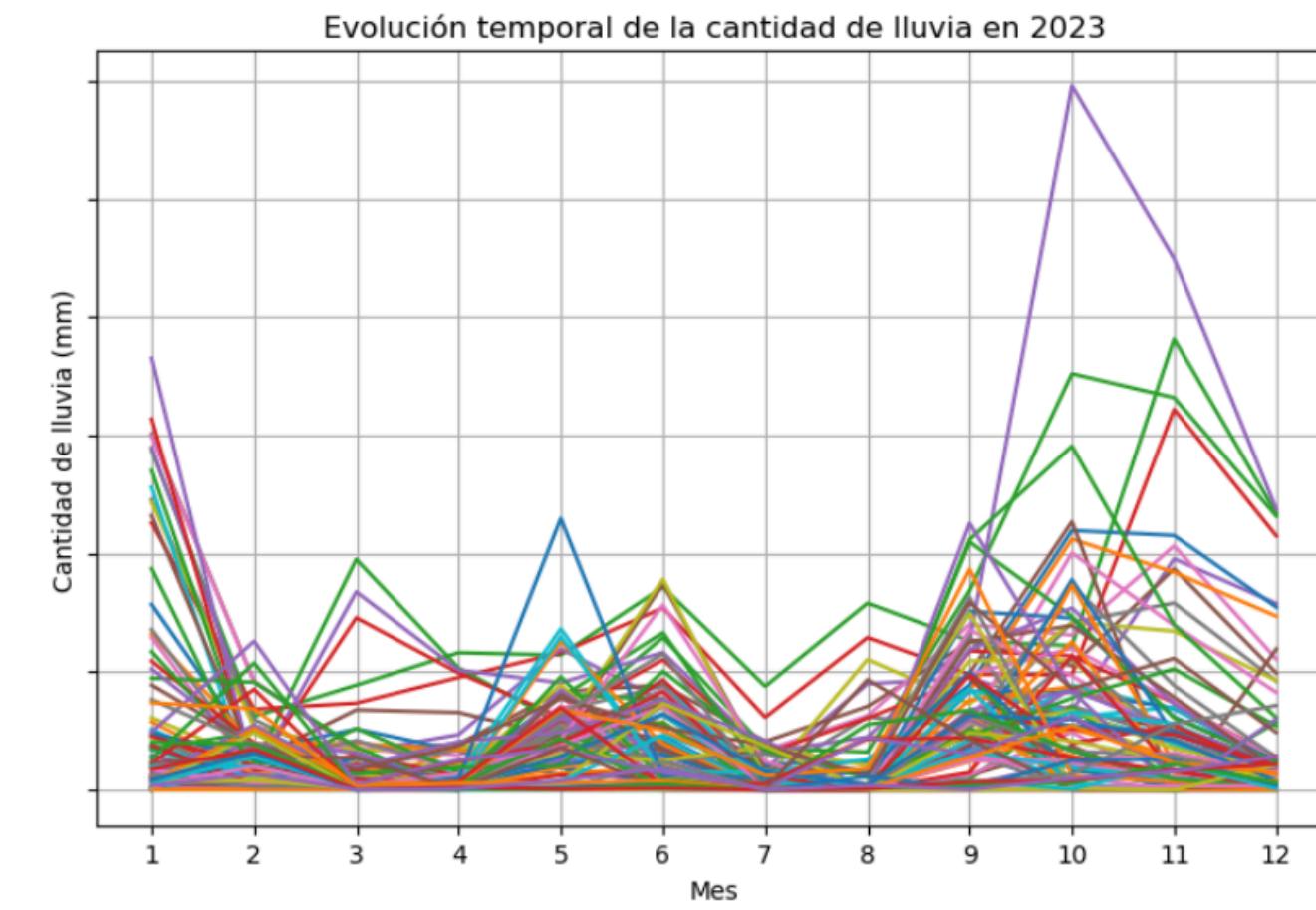
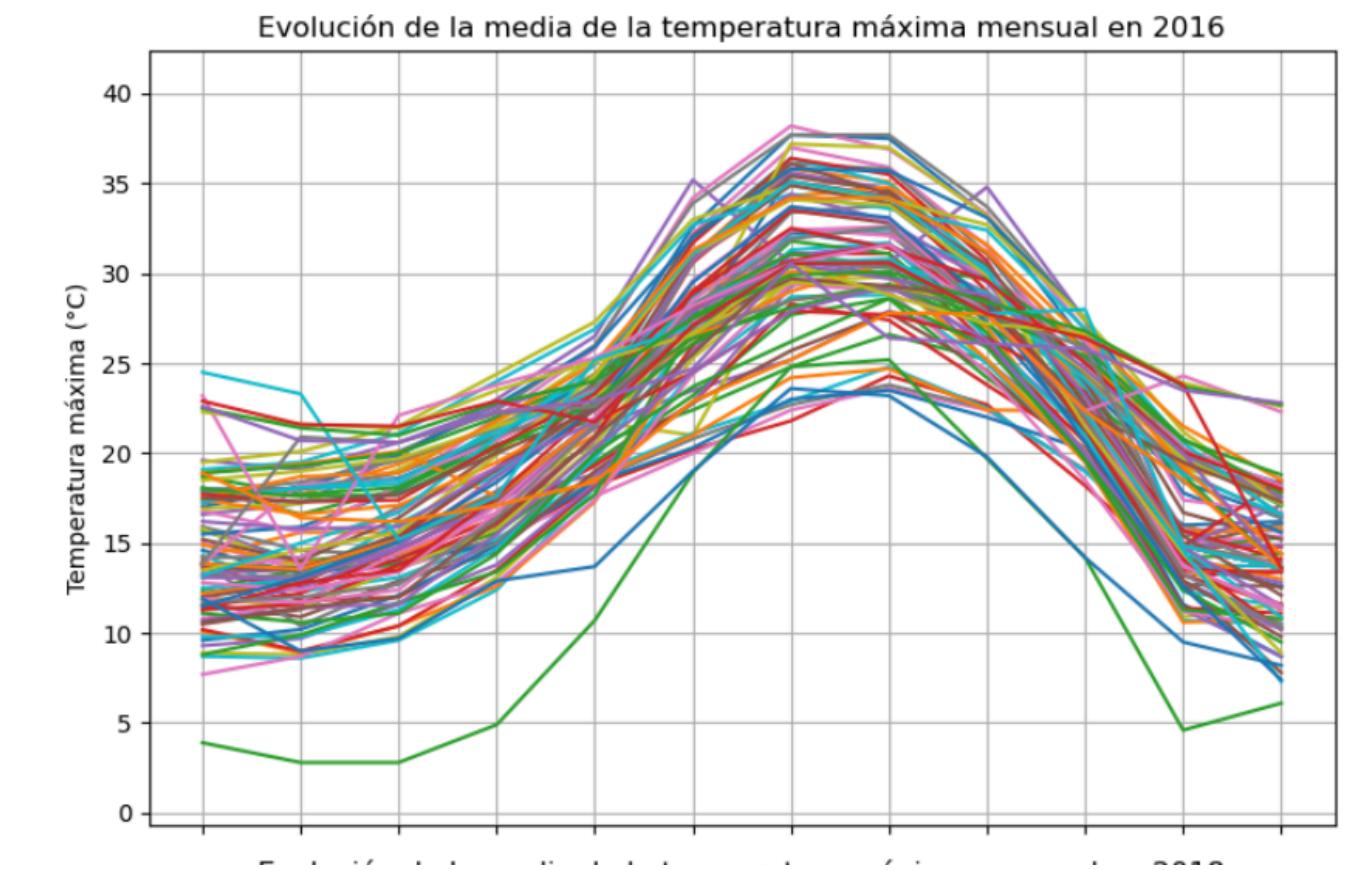
'LLUVIA_MES'



'TEMP_MAX'



'TEMP_MEDIA_MAX'



AUTO-ARIMA; BÚSQUEDA DEL MODELO.

| SARIMAX Results | | | | | | |
|-------------------------|-------------------------------------|-------------------|------------|--------|--------|--------|
| Dep. Variable: | y | No. Observations: | 9120 | | | |
| Model: | SARIMAX(1, 0, 1)x(2, 0, [1, 2], 12) | Log Likelihood | 5683.024 | | | |
| Date: | Tue, 28 May 2024 | AIC | -11350.047 | | | |
| Time: | 11:59:42 | BIC | -11293.101 | | | |
| Sample: | 0 - 9120 | HQIC | -11330.683 | | | |
| Covariance Type: | opg | | | | | |
| coef | std err | z | P> z | [0.025 | 0.975] | |
| intercept | 0.0002 | 0.000 | 0.656 | 0.512 | -0.000 | 0.001 |
| ar.L1 | 0.8224 | 0.006 | 135.186 | 0.000 | 0.810 | 0.834 |
| ma.L1 | -0.3205 | 0.008 | -39.975 | 0.000 | -0.336 | -0.305 |
| ar.S.L12 | 0.6944 | 0.060 | 11.530 | 0.000 | 0.576 | 0.812 |
| ar.S.L24 | 0.3054 | 0.060 | 5.070 | 0.000 | 0.187 | 0.423 |
| ma.S.L12 | -0.5645 | 0.062 | -9.119 | 0.000 | -0.686 | -0.443 |
| ma.S.L24 | -0.1706 | 0.051 | -3.361 | 0.001 | -0.270 | -0.071 |
| sigma2 | 0.0169 | 0.000 | 156.872 | 0.000 | 0.017 | 0.017 |
| Ljung-Box (L1) (Q): | 0.02 | Jarque-Bera (JB): | 73154.90 | | | |
| Prob(Q): | 0.89 | Prob(JB): | 0.00 | | | |
| Heteroskedasticity (H): | 1.67 | Skew: | -0.86 | | | |
| Prob(H) (two-sided): | 0.00 | Kurtosis: | 16.77 | | | |

Obtengo los parámetros óptimos del modelo

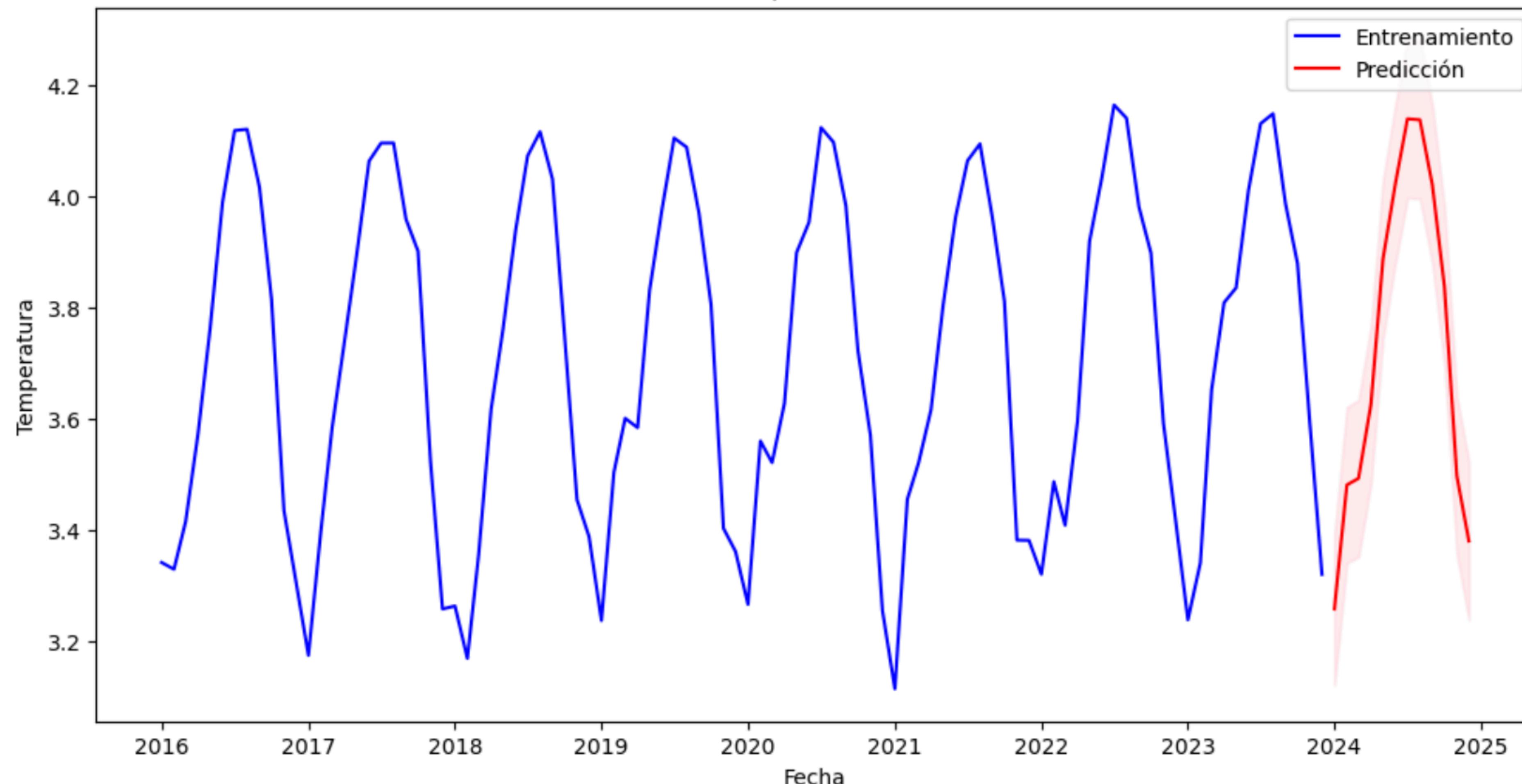
SARIMAX(1, 0, 1)x(2, 0, [1, 2], 12)

Con estos parámetros entreno el modelo

para buscar la predicción

FORECAST

Predicción de temperatura usando SARIMA



RESULTADOS

| | Predicción | CI Inf | CI Sup |
|------------|------------|----------|----------|
| 2024-01-01 | 3.257817 | 3.121459 | 3.394176 |
| 2024-02-01 | 3.480884 | 3.340949 | 3.620819 |
| 2024-03-01 | 3.492706 | 3.351461 | 3.633952 |
| 2024-04-01 | 3.626031 | 3.484300 | 3.767762 |
| 2024-05-01 | 3.887729 | 3.745818 | 4.029641 |
| 2024-06-01 | 4.023566 | 3.881587 | 4.165545 |
| 2024-07-01 | 4.139835 | 3.997831 | 4.281839 |
| 2024-08-01 | 4.138113 | 3.996100 | 4.280126 |
| 2024-09-01 | 4.020615 | 3.878599 | 4.162632 |
| 2024-10-01 | 3.838608 | 3.696590 | 3.980626 |
| 2024-11-01 | 3.497097 | 3.355078 | 3.639115 |
| 2024-12-01 | 3.380011 | 3.237992 | 3.522029 |

| | Predicción | CI Inf | CI Sup | TEMP_REAL |
|------------|------------|----------|----------|-----------|
| 2024-01-01 | 3.257817 | 3.121459 | 3.394176 | 12.977133 |
| 2024-02-01 | 3.480884 | 3.340949 | 3.620819 | 16.228828 |
| 2024-03-01 | 3.492706 | 3.351461 | 3.633952 | 16.422191 |
| 2024-04-01 | 3.626031 | 3.484300 | 3.767762 | 18.768406 |
| 2024-05-01 | 3.887729 | 3.745818 | 4.029641 | 24.389724 |
| 2024-06-01 | 4.023566 | 3.881587 | 4.165545 | 27.941099 |
| 2024-07-01 | 4.139835 | 3.997831 | 4.281839 | 31.388275 |
| 2024-08-01 | 4.138113 | 3.996100 | 4.280126 | 31.334237 |
| 2024-09-01 | 4.020615 | 3.878599 | 4.162632 | 27.858728 |
| 2024-10-01 | 3.838608 | 3.696590 | 3.980626 | 23.219609 |
| 2024-11-01 | 3.497097 | 3.355078 | 3.639115 | 16.494584 |
| 2024-12-01 | 3.380011 | 3.237992 | 3.522029 | 14.668520 |

CONCLUSIÓN

Predicción plausible de las temperaturas, pero

Malas métricas en MSE y MAE

MSE: 364.01

MAE: 18.08

Posibles soluciones:

- Mejor ajuste de los parámetros
- Reducir o ajustar la distribución de los residuos
- Considerar el uso de otros modelos menos sensible a estas características