



# INFORME DE CAMBIO CLIMÁTICO EN ESPAÑA

Introducción del informe sobre la influencia del  
cambio climático en la actividad sísmica y volcánica

Análisis sobre el cambio climático en España en base a registros de  
temperaturas máximas

Visualización de datos registrados

Predicción climática

Ejercicio de clasificación de nuevo registro

cristina vidal

Diciembre 2022

## Informe cambio climático a partir de registro de temperaturas máximas

1. Datos previos y alcance del proyecto.
2. Plots y visualización de datos.
3. Predicción climática en base a análisis de registros históricos.
4. Clasificación de ubicación (estación meteorológica) de un nuevo registro en base a su analogía con los registros históricos.

Ya para el fin de semana, ahora estoy con react.

### 1. Datos previos y alcance del proyecto.

El presente informe se ha realizado en base a los datos proporcionados por la Agencia Estatal de Meteorología AEMET, en su página Open Data:

[AEMET OpenData - Agencia Estatal de Meteorología - AEMET. Gobierno de España](#)

Se ha recogido información de los datos disponibles que se enumeran a continuación, según el escenario 20C3M y el proyecto AEMET, con salida en formato txt:

| Variable   | Escenario  | Proyecto   | Formato   |  |
|--|--|--|---|--|
| <div> <div>Todas</div> <div>Tª máxima</div> <div>Tª mínima</div> <div>Precipitación total acumulada</div> <div>Velocidad del viento a 10m</div> <div>Velocidad máxima del viento a 10m</div> <div>Humedad relativa</div> <div>Percentil 95 de la temperatura máxima diaria</div> <div>Percentil 5 de la temperatura mínima diaria</div> <div>Descargar fichero de la precipitación diaria</div> </div> | <div> <div>Todas</div> <div>HISTORICAL</div> <div>20C3M</div> <div>CTL</div> <div>RCP4.5</div> <div>RCP8.5</div> <div>B1</div> <div>A1B</div> <div>A2</div> </div> | <div> <div>Todas</div> <div>AEMET</div> <div>ENSEMBLES</div> <div>ESCENA</div> <div>ESTCENA</div> <div>CORDEX</div> </div> | <div> <div>Todas</div> <div>TXT</div> <div>SIG</div> </div> |  |
| Variable   | Escenario  | Proyecto   | Formato   | Fichero                                    |
| Tª máxima  | 20C3M  | AEMET  | TXT   | <a href="#">Descargar fichero (4 MB)</a>   |
| Tª mínima  | 20C3M  | AEMET  | TXT   | <a href="#">Descargar fichero (4 MB)</a>   |
| Precipitación total acumulada  | 20C3M  | AEMET  | TXT   | <a href="#">Descargar fichero (3 MB)</a>   |
| Velocidad del viento a 10m   | 20C3M  | AEMET  | TXT   | <a href="#">Descargar fichero (3 MB)</a>   |
| Velocidad máxima del viento a 10m  | 20C3M  | AEMET  | TXT   | <a href="#">Descargar fichero (3 MB)</a>   |
| Humedad relativa   | 20C3M  | AEMET  | TXT   | <a href="#">Descargar fichero (4 MB)</a>   |
| Percentil 95 de la temperatura máxima diaria   | 20C3M  | AEMET  | TXT   | <a href="#">Descargar fichero (421 KB)</a> |
| Percentil 5 de la temperatura mínima diaria  | 20C3M  | AEMET  | TXT   | <a href="#">Descargar fichero (386 KB)</a> |
| Percentil 95 de la precipitación diaria  | 20C3M  | AEMET  | TXT   | <a href="#">Descargar fichero (415 KB)</a> |
| Precipitación máxima en 24h  | 20C3M  | AEMET  | TXT   | <a href="#">Descargar fichero (4 MB)</a>   |
| Nº de días con precipitación >1mm  | 20C3M  | AEMET  | TXT   | <a href="#">Descargar fichero (1 MB)</a>   |
| Nº de días con precipitación >20mm   | 20C3M  | AEMET  | TXT   | <a href="#">Descargar fichero (612 KB)</a> |
| Máximo Nº de días consecutivos con precipitación <1mm  | 20C3M  | AEMET  | TXT   | <a href="#">Descargar fichero (147 KB)</a> |

En esta primera fase del estudio, se han descargado los datos de temperatura máxima, que corresponden al valor promedio mensual, y se van a analizar utilizando técnicas de machine learning para analizar la influencia que tiene la temperatura máxima en el cambio climático.

En otras fases del estudio se repetirá este análisis con otros valores registrados como temperatura mínima, precipitación, velocidad del viento o humedad relativa.

El análisis se ha realizado mediante técnicas de machine learning, aplicadas sobre los datos recogidos en cerca de 350 estaciones meteorológicas ubicadas en España desde 1960, y en este documento se presentan los resultados más relevantes.

Ha resultado imposible asociar cada estación meteorológica con una ubicación concreta y sus coordenadas (longitud, latitud, altura) porque el archivo maestro que debe asociar cada ubicación

con una estación por medio del código, que no se corresponde con las estaciones que aparecen en la tabla de temperaturas máximas por año/mes.

Alguna estación está repetida, hay estaciones situadas en Canarias cuando la documentación indica que la cuadrícula sólo incluye la península y Baleares, resulta complicado identificar si hay valores repetidos en las estaciones y faltan muchas lecturas que se han corregido a mano de forma que se ha complicado el análisis, pero finalmente se podido realizar el estudio.

La primera visualización correcta de los datos ha sido la que se ve a continuación, se aprecia que falta un encabezado que indique lo que hay en cada una de las columnas, de modo que la primera fila aparece como encabezado.

Antes, se ha comprobado que este dataset, de temperatura máxima, estaba enlazado con el fichero maestro de localizaciones.

```
data.head()
```

|   | 196101 | 16.08 | 16.33 | 16.08.1 | 16.08.2 | 16.22 | 16.12 | 16.01 | 16.36 | 16.50 | ... | 11.79.4 | 11.79.5 | 12.54.3 | 12.54.4 | 11.36.3 | 10.84.1 | 10.26 | 13.19.2 | 13.00.1 | 12.98 |
|---|--------|-------|-------|---------|---------|-------|-------|-------|-------|-------|-----|---------|---------|---------|---------|---------|---------|-------|---------|---------|-------|
| 0 | 196102 | 15.29 | 15.45 | 15.56   | 15.56   | 15.56 | 15.31 | 14.84 | 16.13 | 16.09 | ... | 10.76   | 10.70   | 11.37   | 11.37   | 10.34   | 9.81    | 9.34  | 12.40   | 12.23   | 12.13 |
| 1 | 196103 | 15.88 | 16.35 | 16.64   | 16.64   | 17.47 | 17.03 | 15.89 | 17.33 | 18.14 | ... | 11.03   | 11.03   | 11.69   | 11.69   | 10.63   | 9.96    | 9.35  | 12.55   | 12.35   | 12.18 |
| 2 | 196104 | 16.13 | 17.15 | 16.57   | 16.57   | 17.96 | 17.82 | 16.24 | 17.27 | 17.71 | ... | 11.55   | 11.71   | 12.47   | 12.47   | 11.16   | 10.46   | 9.68  | 12.65   | 12.36   | 12.30 |
| 3 | 196105 | 18.68 | 20.01 | 20.31   | 20.31   | 23.24 | 22.28 | 18.15 | 22.17 | 23.90 | ... | 13.10   | 13.18   | 14.29   | 14.29   | 12.56   | 11.63   | 10.63 | 14.28   | 14.13   | 14.06 |
| 4 | 196106 | 22.32 | 22.93 | 24.75   | 24.75   | 27.97 | 25.82 | 20.41 | 27.83 | 30.61 | ... | 15.39   | 15.38   | 16.39   | 16.39   | 14.53   | 13.57   | 12.53 | 16.33   | 16.13   | 15.85 |

5 rows × 1446 columns

#### Primera visualización tabla Tª máxima

Identificación de Localidad a partir del Código AEMET

```
data2.shape
```

(1124, 5)

```
data2.tail()
```

|      | CodigoAEMET | Localidad  | Longitud | Latitud | Altura |
|------|-------------|--|----------|---------|--------|
| 1119 | C449G       | ANAGA-SAN ANDRES                                 | -16.1839 | 28.4967 | 20     |
| 1120 | C458A       | TACORONTE-A. S.E.A.                              | -16.4092 | 28.4864 | 327    |
| 1121 | C649I       | TELDE/AEROPUERTO DE G.CANARIA (GANDO)            | -15.3889 | 27.9292 | 24     |
| 1122 | C659P       | LAS PALMAS DE GRAN CANARIA-JUNTA DE OBRAS DEL... | -15.4167 | 28.15   | 15     |
| 1123 | C665I       | VALLESECO-CASCO                                  | -15.5736 | 28.0472 | 980    |

#### Visualización fichero maestro

Se ha añadido a la tabla un encabezado con la fecha y un número de estación meteorológica para poder identificarlas y además se ha dado a todas las celdas formato de número, a excepción de la columna FECHA.

Esta última operación se ha realizado porque se ha visto que muchas celdas tenían formato fecha y aparecían alineadas a la derecha en lugar de alinearse a la izquierda como las celdas con formato número.

Visualización tabla T<sup>3</sup> max desde spreadsheet

Una vez realizados estos cambios se ha obtenido la siguiente visualización en la que se aprecia claramente los valores que han sido añadidos.

```
data1.head()
```

|   | FECHA  | 500   | 501       | 502       | 503       | 504   | 505       | 506       | 507   | 508       | ... | 1935      | 1936      | 1937  | 1938  | 1939  | 1940  | 1941  | 1942  |
|---|--------|-------|-----------|-----------|-----------|-------|-----------|-----------|-------|-----------|-----|-----------|-----------|-------|-------|-------|-------|-------|-------|
| 0 | 196101 |       | 16.33     | 44.789,00 | 44.789,00 | 16.22 | 44.911,00 | 44.577,00 | 16.36 | 16.50     | ... | 11.79     | 11.79     | 12.54 | 12.54 | 11.36 | 10.84 | 10.26 | 13.19 |
| 1 | 196102 | 15.29 | 15.45     | 15.56     | 15.56     | 15.56 | 15.31     | 14.84     | 16.13 | 44.820,00 | ... | 10.76     | 10.70     | 11.37 | 11.37 | 10.34 | 9.81  | 9.34  | 12.40 |
| 2 | 196103 | 15.88 | 16.35     | 16.64     | 16.64     | 17.47 | 44.637,00 | 15.89     | 17.33 | 18.14     | ... | 44.631,00 | 44.631,00 | 11.69 | 11.69 | 10.63 | 9.96  | 9.35  | 12.55 |
| 3 | 196104 | 16.13 | 17.15     | 16.57     | 16.57     | 17.96 | 17.82     | 16.24     | 17.27 | 17.71     | ... | 11.55     | 11.71     | 12.47 | 12.47 | 11.16 | 10.48 | 9.68  | 12.65 |
| 4 | 196105 | 18.68 | 44.581,00 | 20.31     | 20.31     | 23.24 | 22.28     | 18.15     | 22.17 | 23.90     | ... | 44.847,00 | 13.18     | 14.29 | 14.29 | 12.56 | 11.63 | 10.63 | 14.28 |

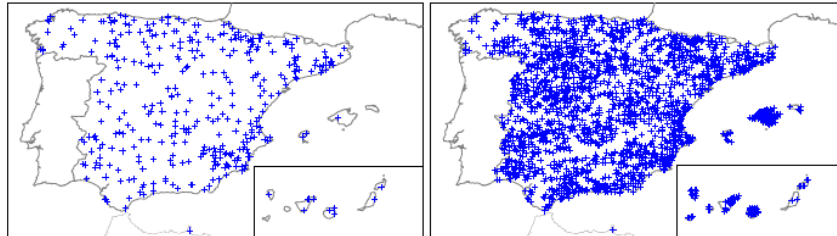
5 rows × 1446 columns

```
data1.shape
```

(480, 1446)

De hecho, si se abre con un bloc de notas (Notepad++) se ve como esos datos que se han añadidos posteriormente aparecen como datos NaN (Not a Number).

Los productos en localidades puntuales ofrecidos en **Escenarios-PNACC Datos mensuales** se refieren a 374 puntos para los productos relacionados con la temperatura y 2.321 puntos para los relacionados con la precipitación, que se distribuyen por toda la geografía de España como se refleja en los mapas de la *Figura 6*. Estos puntos se corresponden a estaciones climatológicas de AEMET que han sido seleccionadas tras un control de calidad basado en criterios de longitud temporal de la serie, número reducido de lagunas y homogeneidad (Brunet *et al.* 2008, Herrera 2011, Herrera *et al.* 2012).



Mapa de estaciones climatológicas ubicadas en España

A partir de aquí se trabaja con los datos origuinales y se va a mostrar mediante diferentes visualizaciones la influencia de la temperatura máxima en el cambio climático.

Se ha cargado el archivo original sobre data1 y se van a mostrar las posibles visualizaciones que se pueden obtener a partir de él.

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

```

```

data1.head()

```

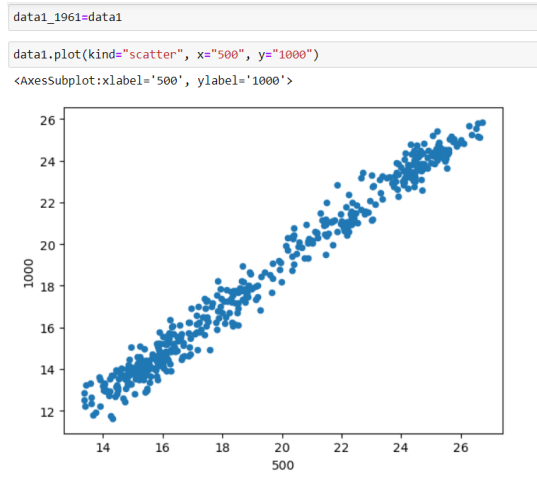
|   | FECHA  | 500   | 501   | 502   | 503   | 504   | 505   | 506   | 507   | 508   | ... | 1935  | 1936  | 1937  | 1938  | 1939  | 1940  | 1941  | 1942  | 1943  | 1944  |
|---|--------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 0 | 196101 | 16.08 | 16.33 | 16.08 | 16.08 | 16.22 | 16.12 | 16.01 | 16.36 | 16.50 | ... | 11.79 | 11.79 | 12.54 | 12.54 | 11.36 | 10.84 | 10.26 | 13.19 | 13.00 | 12.98 |
| 1 | 196102 | 15.29 | 15.45 | 15.56 | 15.56 | 15.56 | 15.31 | 14.84 | 16.13 | 16.09 | ... | 10.76 | 10.70 | 11.37 | 11.37 | 10.34 | 9.81  | 9.34  | 12.40 | 12.23 | 12.13 |
| 2 | 196103 | 15.88 | 16.35 | 16.64 | 16.64 | 17.47 | 17.03 | 15.89 | 17.33 | 18.14 | ... | 11.03 | 11.03 | 11.69 | 11.69 | 10.63 | 9.96  | 9.35  | 12.55 | 12.35 | 12.18 |
| 3 | 196104 | 16.13 | 17.15 | 16.57 | 16.57 | 17.96 | 17.82 | 16.24 | 17.27 | 17.71 | ... | 11.55 | 11.71 | 12.47 | 12.47 | 11.16 | 10.46 | 9.68  | 12.65 | 12.36 | 12.30 |
| 4 | 196105 | 18.68 | 20.01 | 20.31 | 20.31 | 23.24 | 22.28 | 18.15 | 22.17 | 23.90 | ... | 13.10 | 13.18 | 14.29 | 14.29 | 12.56 | 11.63 | 10.63 | 14.28 | 14.13 | 14.06 |

5 rows × 1446 columns

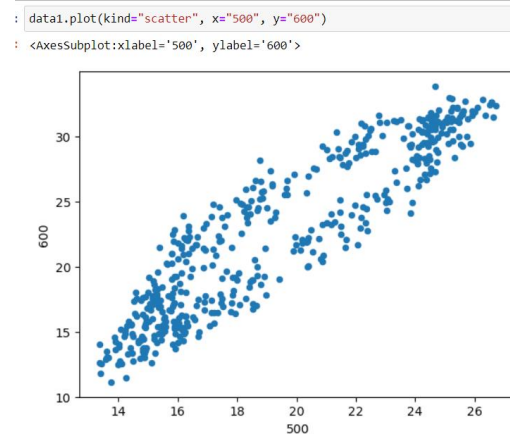
## Página 4 de 16

Representación de TODOS los puntos que recoge el dataset, en la estación 500 y la estación 1000 a lo largo del tiempo desde enero de 1961 hasta diciembre del año 2000.

Parece que hay una tendencia ascendente y relativamente similar entre ambas



Sin embargo, al comparar la estación 500 con la estación 1940 a lo largo del tiempo desde enero de 1961 hasta diciembre del año 2000 se ve claramente que la separación entre los puntos es mayor, lo cual se debe, seguramente, a que las estaciones están más separadas y tienen climas distintos aunque no se puede asegurar por el problema mencionado con el fichero maestro.



Hace falta un análisis más exhaustivo para verificar la representación gráfica del aumento de temperaturas.

Se necesita ajustar más la visualización para ver más claro el análisis, haciendo uso del Data Wranglin (cirugía de datos).

Se puede extraer un subconjunto de datos, por ejemplo, de una sola estación meteorológica y el objeto que se obtiene es de tipo series o vectores.

## Crear un subconjunto de datos

```

: Estacion500=data1_1961["500"]
: Estacion500.head()
: 0    16.08
  1    15.29
  2    15.88
  3    16.13
  4    18.68
  Name: 500, dtype: float64
: type(Estacion500)
: pandas.core.series.Series

```

También se pueden extraer varias columnas, varias estaciones meteorológicas, y el objeto obtenido sería un DataFrame.

```

subset=data1_1961[["500", "600", "700"]]
subset.head()

```

|   | 500   | 600   | 700   |
|---|-------|-------|-------|
| 0 | 16.08 | 18.09 | 18.86 |
| 1 | 15.29 | 14.18 | 16.10 |
| 2 | 15.88 | 18.35 | 19.96 |
| 3 | 16.13 | 21.63 | 22.58 |
| 4 | 18.68 | 24.34 | 24.84 |

```

type(subset)
pandas.core.frame.DataFrame

```

Otra opción para esto es especificar las columnas deseadas es utilizar un solo paso mediante la opción `desired_columns`.

Así, se haría una comparación entre estaciones, en principio, distribuidas por toda la malla.

```

: desired_columns=["500", "1000", "1500", "1944"]
subset=data1[desired_columns]
subset.head()

```

|   | 500   | 1000  | 1500  | 1944  |
|---|-------|-------|-------|-------|
| 0 | 16.08 | 14.53 | 12.42 | 12.98 |
| 1 | 15.29 | 13.99 | 12.15 | 12.13 |
| 2 | 15.88 | 14.38 | 14.66 | 12.18 |
| 3 | 16.13 | 15.52 | 16.75 | 12.30 |
| 4 | 18.68 | 17.78 | 22.48 | 14.06 |

Para reducir el peso del archivo se puede escoger una serie de columnas de trabajo y eliminar las demás, lo cual tendría más sentido si se conociera la ubicación exacta de las estaciones, pero aún así se ha hecho escogiendo únicamente las estaciones por centenas.

```

desired_columns=["500","600","700","800","900","1000","1100","1200","1300","1400","1500","1600","1700","1800","1900","2000","2100"]
desired_columns=

```

```

['500',
'600',
'700',
'800',
'900',
'1000',
'1100',
'1200',
'1300',
'1400',
'1500',
'1600',
'1700',
'1800',
'1900',
'2000',
'2100',
'2200',
'2300',
'2400',
'2500',
'2600',
'2700',
'2800',
'2900']

```







### 3. Predicción climática en base a análisis de registros históricos.

Para hacer una predicción climática de evolución de las temperaturas máximas en base únicamente a los registros de temperatura previos a lo largo de 40 años, con carácter geográfico nacional, de ha creado una estación meteorológica ficticia sobre la que se va a hacer la predicción.

Esta estación ficticia se ha obtenido mediante el cálculo del promedio de temperaturas de todas las estaciones meteorológicas, pues el ámbito de este estudio es nacional sin entrar en particularidades propias de cada región.

Ya para el fin de semana, ahora estoy con react.

### 4. Clasificación de ubicación (estación meteorológica) de un nuevo registro en base a su analogía con los registros históricos.

Otro objetivo del estudio es comprobar la capacidad que se tiene de clasificar un nuevo registro de temperaturas como perteneciente a una zona concreta, por su estación meteorológica asociada.

Cabe destacar la complejidad de esta clasificación por los motivos indicados en el apartado 1 Datos Previos y alcance del proyecto, ya que ha resultado imposible asociar las estaciones meteorológicas con ubicaciones concretas de la geografía española por desplazamientos en la ubicación de las estaciones, cambios en el entorno, etc....

No obstante, se ha realizado el estudio tratando de identificar la o las estaciones que más se asemejan al nuevo registro y para ello se han métodos de clustering, que establecen categorías de las entradas del dataset en clusters (grupos) en base a su similitud.

Así, se han establecido unos grupos de estaciones meteorológicas que corresponden a las mismas zonas geográficas, calculándose para cada cluster las distancias con los demás clusters.

Estas distancias se pueden calcular en base a 3 modelos, Manhattan distance, Euclidean distance y Chebichev distance, en este proyecto se ha optado por trabajar con la distancia euclídea pero se podría haber trabajado con cualquiera de las otras.

Para calcular la distancia euclídea entre columnas voy a quitar la primera columna de fechas para que no calcule distancias con esa columna:

```
: sinfecha=data20.columns.values.tolist()[1:]
sinfecha
```

```
['500',
 '501',
 '502',
 '503',
 '504',
 '505',
 '506',
 '507',
 '508',
 '509',
 '510',
 '511',
 '512',
 '513',
 '514',
 '515',
 '516',
 '517',
 '518',
 '519']
```

Ya continuación calculo las distancias:

Distancia de Manhattan o de manzana: dd1

Distancia Euclidea:  $dd_2$ 

- Cuanto mayor es el índice, más pequeña es la distancia y mayor es el número de clusters

En este caso se ha utilizado la distancia euclídea, que es el segmento que une los centros (centroides) de los diferentes registros y evidentemente, en la diagonal la distancia de un punto a sí mismo es cero.

```
dd1=distance_matrix(data20[sinfeca],data20[sinfeca],p=1)
dd2=distance_matrix(data20[sinfeca],data20[sinfeca],p=2)
dd10=distance_matrix(data20[sinfeca],data20[sinfeca],p=10)
```

```
dd1
array([[ 0., 1004.42, 2356.08, ..., 8947.19, 1364.48, 1978.3 ],
       [ 1004.42, 0., 3129.68, ..., 9807.51, 2222.94, 1595.06],
       [ 2356.08, 3129.68, 0., ..., 6678.83, 1767.42, 4279. ],
       ...,
       [ 8947.19, 9807.51, 6678.83, ..., 0., 7584.69, 10897.97],
       [ 1364.48, 2222.94, 1767.42, ..., 7584.69, 0., 3313.28],
       [ 1978.3, 1595.06, 4279. , ..., 10897.97, 3313.28, 0. ]])
```

```
dd1.shape
```

(480, 480)

```
: dd1=distance_matrix(data20[sinfecha],data20[sinfecha],p=1)
dd2=distance_matrix(data20[sinfecha],data20[sinfecha],p=2)
```

```

: dd2
: array([[ 0.,          38.96161958,  70.27533991, ..., 240.38073238,
[ 44.4569837 ,  56.5490053 ],
[ 38.96161958,  0.,          88.96746484, ..., 263.34049651,
[ 73.7815206 ,  47.27539952],
[ 70.27533991,  88.96746484,  0.,          ..., 181.73348591,
[ 54.63110652, 120.2409639 ],
...,
[240.38073238, 263.34049651, 181.73348591, ..., 0.,
206.39178109, 292.49225169],
[ 44.4569837 ,  73.7815206 ,  54.63110652, ..., 206.39178109,
[ 0.,          90.35050968],
[ 56.5490053 ,  47.27539952, 120.2409639 , ..., 292.49225169,
[ 90.35050968,  0.,          11]])

```

: dd10

```
: array([[ 0.,      5.39573019,  6.03781562, ..., 14.46059572,
          4.57319196,  3.86360627],
 [ 5.39573019,  0.,      6.65502262, ..., 15.826577 ,
          7.60148769,  3.70765741],
 [ 6.03781562,  6.65502262,  0.,      ..., 11.57316017,
          4.66075507,  8.28161328],
 ...,
 [14.46059572, 15.826577 , 11.57316017, ..., 0.,
          12.93170454, 17.42523828],
 [ 4.57319196,  7.60148769,  4.66075507, ..., 12.93170454,
          0.,      6.2697782 ],
 [ 3.86360627,  3.70765741,  8.28161328, ..., 17.42523828,
          6.2697782 ,  0.      ]])
```

Se puede ver que el tamaño de estas matrices es de 480, lo que coincide con el número de localidades que aparecen en el fichero maestro y resulta coherente porque son las estaciones desde las que se han tomado los registros.

También se observa que cuanto mayor es el número de clusters, menor es la distancia entre ellos y se ha querido hacer una representación gráfica en 3D de algunas de estas distancias, pero no ha salido el dibujo, posiblemente falte algún componente por instalar en el equipo donde se han realizado los cálculos o el número de elementos es excesivo.

```
: import matplotlib.pyplot as plt
  from mpl_toolkits.mplot3d import Axes3D

: fig=plt.figure()
  <Figure size 640x480 with 0 Axes>

: ax=fig.add_subplot(111, projection="3d")

: ax.scatter( xs=data20["1921"], ys=data20["500"], zs=data20["582"])
  <mpl_toolkits.mplot3d.art3d.Path3DCollection at 0x2dff9eb1430>

:
```

Como el tamaño de este dataset es excesivamente grande, 1446 columnas, se han agrupado los datos en torno a algunas columnas, esto tendría más sentido si se conocieran las ubicaciones de las estaciones, no obstante, se ha clasificado revisando los datos de la tabla original descargada de AEMET, concretamente los registros de temperaturas en agosto de un año intermedio, por ejemplo 1973.

Con esta revisión de los datos originales, se han establecido unos baremos de temperatura, escogiendo unas estaciones representativas:

| Rango temperatura | Nº Estación meteorológica | Temperatura agosto 1973 |
|-------------------|---------------------------|-------------------------|
| 15º               | 1921                      | 15.67                   |
| 20º               | 1772                      | 20.62                   |
| 25º               | 500                       | 25.55                   |
| 30º               | 522                       | 30.02                   |
| 35º               | 529                       | 335.66                  |
| 40º               | 582                       | 38.23                   |

Se ha obtenido el resultado que se ve a continuación que no es demasiado bueno porque hace 479 grupos, uno por cada fila y muestra en la zona superior de cada grupo las temperaturas de 6 estaciones meteorológicas seleccionadas previamente.

```
grouped_columns=data20.groupby(["1921","1772","500","522","529","582"])
```

```
len(grouped_columns)
```

```
479
```

```
for names, groups in grouped_columns:
    print(names)
    print(groups)
```

```
(7.24, 1.78, 13.91, 13.05, 13.19, 14.58)
FECHA      500      501      502      503      504      505      506      507      508 \
229  198002  13.91  14.08  13.92  13.92  14.08  14.02  13.85  14.27  14.19

...  1935  1936  1937  1938  1939  1940  1941  1942  1943  1944
229  ...  8.97  8.94  9.22  9.22  8.75  8.46  8.24  10.85  10.6  10.39

[1 rows x 1446 columns]
(7.33, -1.22, 13.75, 11.22, 12.3, 12.81)
FECHA      500      501      502      503      504      505      506      507      508 \
240  198101  13.75  13.58  13.66  13.66  13.0  13.08  13.85  13.81  13.38

...  1935  1936  1937  1938  1939  1940  1941  1942  1943  1944
240  ...   9.0  9.03  9.22  9.22  8.95  8.67  8.55  10.78  10.41  10.13

[1 rows x 1446 columns]
(7.71, 3.82, 14.39, 13.88, 14.02, 15.67)
FECHA      500      501      502      503      504      505      506      507      508 \
434  199703  14.39  14.67  14.53  14.53  14.83  14.78  14.35  14.97  14.96
```

Se ha consultado el método que debe utilizarse para hacer la agrupación óptima en la siguiente utilizando la librería numpy.

### [Clustering jerárquico \(scipy.cluster.hierarchy\) — Manual de SciPy v1.9.3](#)

Finalmente se ha optado por hacer un cluster jerárquico en el sea el algoritmo el que calcule el número de clusters óptimo para generar un dendrograma, que es una representación gráfica de los datos en forma de árbol, que organiza los datos en subcategorías que se van dividiendo hasta llegar al nivel de detalle deseado.

#### a. Clustering jerárquico.

El proceso de agrupación jerárquica implica agrupar subclusters en clusters más grandes de manera ascendente, o bien dividir un cluster más grande en subgrupos mas pequeños de arriba hacia abajo.

Este proceso se ha conseguido mediante el algoritmo de agrupación linkage, pero al realizarlo sobre el fichero maestro ha aparecido un error que se ha solucionado modificando este fichero maestro, lo cual pone en duda la eficacia del método para ficheros con datos excesivamente desordenados.

```
: linked = linkage(data5["CodigoAEMET"], 'ward')
linked

-----
ValueError                                Traceback (most recent call last)
Input In [124], in <cell line: 1>()
----> 1 linked = linkage(data5["CodigoAEMET"], 'ward')
      2 linked

File ~\anaconda3\envs\Ciberseguridad\lib\site-packages\scipy\cluster\hierarchy.py:1046, in linkage(y, method, metric, optimal_ordering)
    1043 if method not in _LINKAGE_METHODS:
    1044     raise ValueError("Invalid method: {}".format(method))
-> 1046 y = _convert_to_double(np.asarray(y, order='c'))
    1048 if y.ndim == 1:
    1049     distance.is_valid_y(y, throw=True, name='y')

File ~\anaconda3\envs\Ciberseguridad\lib\site-packages\scipy\cluster\hierarchy.py:1572, in _convert_to_double(X)
    1570 def _convert_to_double(X):
    1571     if X.dtype != np.double:
-> 1572         X = X.astype(np.double)
    1573     if not X.flags.contiguous:
    1574         X = X.copy()

ValueError: could not convert string to float: '0016B'
```

*Error señalado al no convertir a float un dato string*

```
] : data5=pd.read_csv("../Dataset/Fichero_Maestro_modificado.csv")
```

```
] : data5
```

```
] :
```

|     | CodigoAEMET | Localidad   | Longitud | Latitud | Altura |
|-----|-------------|---|----------|---------|--------|
| 0   | 1           | EL PERELLO  | 71.667   | 408.667 | 142.0  |
| 1   | 0016B       | NaN   | NaN      | NaN     | NaN    |
| 2   | 17          | VILASECA DE SOLCINA                               | 1.145    | 411.117 | 53.0   |
| 3   | 19          | NaN   | NaN      | NaN     | NaN    |
| 4   | 0020O       | NaN   | NaN      | NaN     | NaN    |
| ... | ...         | ...   | ...      | ...     | ...    |
| 369 | C449G       | ANAGA-SAN ANDRES                                  | -161.839 | 284.967 | 20.0   |
| 370 | C458A       | TACORONTE-A. S.E.A.                               | -164.092 | 284.864 | 327.0  |
| 371 | C649I       | TELDE/AEROPUERTO DE G.CANARIA (GANDO)             | -153.889 | 279.292 | 24.0   |
| 372 | C659P       | LAS PALMAS DE GRAN CANARIA-JUNTA DE OBRAS DEL ... | -154.167 | 28.150  | 15.0   |
| 373 | C665I       | VALLESECO-CASCO                                   | -155.736 | 280.472 | 980.0  |

374 rows × 5 columns

### Fichero maestro con datos float y string en su columna CodigoAEMET

```
: data2=pd.read_csv("../Dataset/Fichero_Maestro_modificado.csv")
```

```
: data2
```

```
:
```

|     | CodigoAEMET | Localidad   | Longitud | Latitud | Altura |
|-----|-------------|---|----------|---------|--------|
| 0   | 1           | EL PERELLO  | 71.667   | 408.667 | 142.0  |
| 1   | 160000      | NaN   | NaN      | NaN     | NaN    |
| 2   | 17          | VILASECA DE SOLCINA                               | 1.145    | 411.117 | 53.0   |
| 3   | 19          | NaN   | NaN      | NaN     | NaN    |
| 4   | 200000      | NaN   | NaN      | NaN     | NaN    |
| ... | ...         | ...   | ...      | ...     | ...    |
| 369 | 449000      | ANAGA-SAN ANDRES                                  | -161.839 | 284.967 | 20.0   |
| 370 | 458000      | TACORONTE-A. S.E.A.                               | -164.092 | 284.864 | 327.0  |
| 371 | 649000      | TELDE/AEROPUERTO DE G.CANARIA (GANDO)             | -153.889 | 279.292 | 24.0   |
| 372 | 659000      | LAS PALMAS DE GRAN CANARIA-JUNTA DE OBRAS DEL ... | -154.167 | 28.150  | 15.0   |
| 373 | 665000      | VALLESECO-CASCO                                   | -155.736 | 280.472 | 980.0  |

374 rows × 5 columns

### Fichero maestro modificado

En la modificación se ha visto que ciertos valores son NaN (Not a Number), valores que provocan errores en la agrupación de los clusters.

```
data2.dropna(axis=0, how="any")
```

|     | CodigoAEMET | Localidad   | Longitud  | Latitud | Altura |
|-----|-------------|---|-----------|---------|--------|
| 0   | 1           | EL PERELLO  | 71.6670   | 408.667 | 142.0  |
| 2   | 17          | VILASECA DE SOLCINA                               | 1.1450    | 411.117 | 53.0   |
| 5   | 25          | SARRAL  | 1.2500    | 41.450  | 400.0  |
| 8   | 111         | CABRIANES   | 1.8953    | 417.986 | 246.0  |
| 9   | 158000      | MONTSERRAT  | 1.8403    | 415.956 | 735.0  |
| ... | ...         | ...   | ...       | ...     | ...    |
| 369 | 449000      | ANAGA-SAN ANDRES                                  | -161.8390 | 284.967 | 20.0   |
| 370 | 458000      | TACORONTE-A. S.E.A.                               | -164.0920 | 284.864 | 327.0  |
| 371 | 649000      | TELDE/AEROPUERTO DE G.CANARIA (GANDO)             | -153.8890 | 279.292 | 24.0   |
| 372 | 659000      | LAS PALMAS DE GRAN CANARIA-JUNTA DE OBRAS DEL ... | -154.1670 | 28.150  | 15.0   |
| 373 | 665000      | VALLESECO-CASCO                                   | -155.7360 | 280.472 | 980.0  |

350 rows × 5 columns

Para implementar el clustering jerárquico en Python se ha utilizado la muestra de temperaturas de agosto desde 1961 hasta 2000 y se ha representado en un dendrograma en el cual se aprecian las distancias (vertical) que permitirían unir a un cluster con los demás.

## Clustering jerárquico

```

: import matplotlib.pyplot as plt
: from scipy.cluster.hierarchy import dendrogram, linkage
: import numpy as np
: import pandas as pd

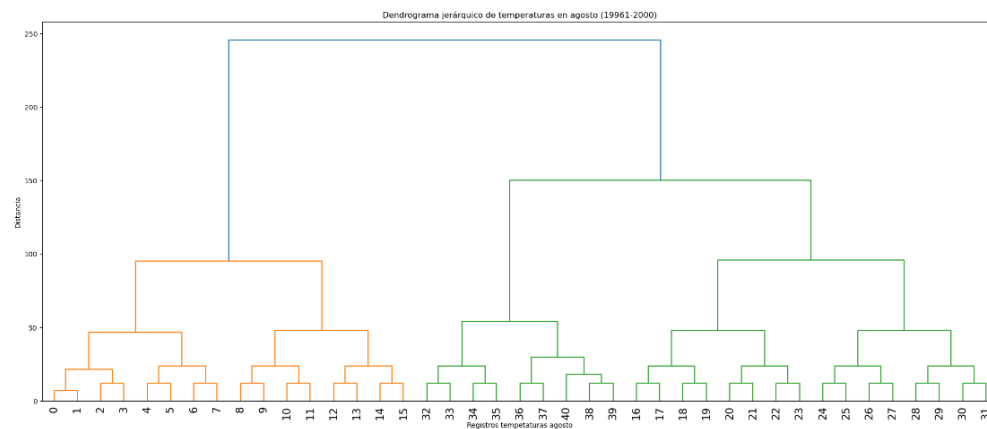
: X = [[i] for i in [0, 7, 19, 31, 43, 55, 67, 79, 91, 103, 115, 127, 139, 151, 163, 175, 187, 199, 211, 223, 235, 247, 259, 271, 283, 295, 307, 319, 331, 343, 355, 367, 379, 391, 403, 415, 427, 439, 451, 463, 475, 487, 499, 511, 523, 535, 547, 559, 571, 583, 595, 607, 619, 631, 643, 655, 667, 679, 691, 703, 715, 727, 739, 751, 763, 775, 787, 799, 811, 823, 835, 847, 859, 871, 883, 895, 907, 919, 931, 943, 955, 967, 979, 991, 1003, 1015, 1027, 1039, 1051, 1063, 1075, 1087, 1099, 1111, 1123, 1135, 1147, 1159, 1171, 1183, 1195, 1207, 1219, 1231, 1243, 1255, 1267, 1279, 1291, 1303, 1315, 1327, 1339, 1351, 1363, 1375, 1387, 1399, 1411, 1423, 1435, 1447, 1459, 1471, 1483, 1495, 1507, 1519, 1531, 1543, 1555, 1567, 1579, 1591, 1603, 1615, 1627, 1639, 1651, 1663, 1675, 1687, 1699, 1711, 1723, 1735, 1747, 1759, 1771, 1783, 1795, 1807, 1819, 1831, 1843, 1855, 1867, 1879, 1891, 1903, 1915, 1927, 1939, 1951, 1963, 1975, 1987, 1999, 2011, 2023, 2035, 2047, 2059, 2071, 2083, 2095, 2107, 2119, 2131, 2143, 2155, 2167, 2179, 2191, 2203, 2215, 2227, 2239, 2251, 2263, 2275, 2287, 2299, 2311, 2323, 2335, 2347, 2359, 2371, 2383, 2395, 2407, 2419, 2431, 2443, 2455, 2467, 2479, 2491, 2503, 2515, 2527, 2539, 2551, 2563, 2575, 2587, 2599, 2611, 2623, 2635, 2647, 2659, 2671, 2683, 2695, 2707, 2719, 2731, 2743, 2755, 2767, 2779, 2791, 2803, 2815, 2827, 2839, 2851, 2863, 2875, 2887, 2899, 2911, 2923, 2935, 2947, 2959, 2971, 2983, 2995, 3007, 3019, 3031, 3043, 3055, 3067, 3079, 3091, 3103, 3115, 3127, 3139, 3151, 3163, 3175, 3187, 3199, 3211, 3223, 3235, 3247, 3259, 3271, 3283, 3295, 3307, 3319, 3331, 3343, 3355, 3367, 3379, 3391, 3403, 3415, 3427, 3439, 3451, 3463, 3475, 3487, 3499, 3511, 3523, 3535, 3547, 3559, 3571, 3583, 3595, 3607, 3619, 3631, 3643, 3655, 3667, 3679, 3691, 3703, 3715, 3727, 3739, 3751, 3763, 3775, 3787, 3799, 3811, 3823, 3835, 3847, 3859, 3871, 3883, 3895, 3907, 3919, 3931, 3943, 3955, 3967, 3979, 3991, 4003, 4015, 4027, 4039, 4051, 4063, 4075, 4087, 4099, 4111, 4123, 4135, 4147, 4159, 4171, 4183, 4195, 4207, 4219, 4231, 4243, 4255, 4267, 4279, 4291, 4303, 4315, 4327, 4339, 4351, 4363, 4375, 4387, 4399, 4411, 4423, 4435, 4447, 4459, 4471, 4483, 4495, 4507, 4519, 4531, 4543, 4555, 4567, 4579, 4591, 4603, 4615, 4627, 4639, 4651, 4663, 4675, 4687, 4699, 4711, 4723, 4735, 4747, 4759, 4771, 4783, 4795, 4807, 4819, 4831, 4843, 4855, 4867, 4879, 4891, 4903, 4915, 4927, 4939, 4951, 4963, 4975, 4987, 4999, 5011, 5023, 5035, 5047, 5059, 5071, 5083, 5095, 5107, 5119, 5131, 5143, 5155, 5167, 5179, 5191, 5203, 5215, 5227, 5239, 5251, 5263, 5275, 5287, 5299, 5311, 5323, 5335, 5347, 5359, 5371, 5383, 5395, 5407, 5419, 5431, 5443, 5455, 5467, 5479, 5491, 5503, 5515, 5527, 5539, 5551, 5563, 5575, 5587, 5599, 5611, 5623, 5635, 5647, 5659, 5671, 5683, 5695, 5707, 5719, 5731, 5743, 5755, 5767, 5779, 5791, 5803, 5815, 5827, 5839, 5851, 5863, 5875, 5887, 5899, 5911, 5923, 5935, 5947, 5959, 5971, 5983, 5995, 6007, 6019, 6031, 6043, 6055, 6067, 6079, 6091, 6103, 6115, 6127, 6139, 6151, 6163, 6175, 6187, 6199, 6211, 6223, 6235, 6247, 6259, 6271, 6283, 6295, 6307, 6319, 6331, 6343, 6355, 6367, 6379, 6391, 6403, 6415, 6427, 6439, 6451, 6463, 6475, 6487, 6499, 6511, 6523, 6535, 6547, 6559, 6571, 6583, 6595, 6607, 6619, 6631, 6643, 6655, 6667, 6679, 6691, 6703, 6715, 6727, 6739, 6751, 6763, 6775, 6787, 6799, 6811, 6823, 6835, 6847, 6859, 6871, 6883, 6895, 6907, 6919, 6931, 6943, 6955, 6967, 6979, 6991, 7003, 7015, 7027, 7039, 7051, 7063, 7075, 7087, 7099, 7111, 7123, 7135, 7147, 7159, 7171, 7183, 7195, 7207, 7219, 7231, 7243, 7255, 7267, 7279, 7291, 7303, 7315, 7327, 7339, 7351, 7363, 7375, 7387, 7399, 7411, 7423, 7435, 7447, 7459, 7471, 7483, 7495, 7507, 7519, 7531, 7543, 7555, 7567, 7579, 7591, 7603, 7615, 7627, 7639, 7651, 7663, 7675, 7687, 7699, 7711, 7723, 7735, 7747, 7759, 7771, 7783, 7795, 7807, 7819, 7831, 7843, 7855, 7867, 7879, 7891, 7903, 7915, 7927, 7939, 7951, 7963, 7975, 7987, 7999, 8011, 8023, 8035, 8047, 8059, 8071, 8083, 8095, 8107, 8119, 8131, 8143, 8155, 8167, 8179, 8191, 8203, 8215, 8227, 8239, 8251, 8263, 8275, 8287, 8299, 8311, 8323, 8335, 8347, 8359, 8371, 8383, 8395, 8407, 8419, 8431, 8443, 8455, 8467, 8479, 8491, 8503, 8515, 8527, 8539, 8551, 8563, 8575, 8587, 8599, 8611, 8623, 8635, 8647, 8659, 8671, 8683, 8695, 8707, 8719, 8731, 8743, 8755, 8767, 8779, 8791, 8803, 8815, 8827, 8839, 8851, 8863, 8875, 8887, 8899, 8911, 8923, 8935, 8947, 8959, 8971, 8983, 8995, 9007, 9019, 9031, 9043, 9055, 9067, 9079, 9091, 9103, 9115, 9127, 9139, 9151, 9163, 9175, 9187, 9199, 9211, 9223, 9235, 9247, 9259, 9271, 9283, 9295, 9307, 9319, 9331, 9343, 9355, 9367, 9379, 9391, 9403, 9415, 9427, 9439, 9451, 9463, 9475, 9487, 9499, 9511, 9523, 9535, 9547, 9559, 9571, 9583, 9595, 9607, 9619, 9631, 9643, 9655, 9667, 9679, 9691, 9703, 9715, 9727, 9739, 9751, 9763, 9775, 9787, 9799, 9811, 9823, 9835, 9847, 9859, 9871, 9883, 9895, 9907, 9919, 9931, 9943, 9955, 9967, 9979, 9991, 10003, 10015, 10027, 10039, 10051, 10063, 10075, 10087, 10099, 10111, 10123, 10135, 10147, 10159, 10171, 10183, 10195, 10207, 10219, 10231, 10243, 10255, 10267, 10279, 10291, 10303, 10315, 10327, 10339, 10351, 10363, 10375, 10387, 10399, 10411, 10423, 10435, 10447, 10459, 10471, 10483, 10495, 10507, 10519, 10531, 10543, 10555, 10567, 10579, 10591, 10603, 10615, 10627, 10639, 10651, 10663, 10675, 10687, 10699, 10711, 10723, 10735, 10747, 10759, 10771, 10783, 10795, 10807, 10819, 10831, 10843, 10855, 10867, 10879, 10891, 10903, 10915, 10927, 10939, 10951, 10963, 10975, 10987, 10999, 11011, 11023, 11035, 11047, 11059, 11071, 11083, 11095, 11107, 11119, 11131, 11143, 11155, 11167, 11179, 11191, 11203, 11215, 11227, 11239, 11251, 11263, 11275, 11287, 11299, 11311, 11323, 11335, 11347, 11359, 11371, 11383, 11395, 11407, 11419, 11431, 11443, 11455, 11467, 11479, 11491, 11503, 11515, 11527, 11539, 11551, 11563, 11575, 11587, 11599, 11611, 11623, 11635, 11647, 11659, 11671, 11683, 11695, 11707, 11719, 11731, 11743, 11755, 11767, 11779, 11791, 11803, 11815, 11827, 11839, 11851, 11863, 11875, 11887, 11899, 11911, 11923, 11935, 11947, 11959, 11971, 11983, 11995, 12007, 12019, 12031, 12043, 12055, 12067, 12079, 12091, 12103, 12115, 12127, 12139, 12151, 12163, 12175, 12187, 12199, 12211, 12223, 12235, 12247, 12259, 12271, 12283, 12295, 12307, 12319, 12331, 12343, 12355, 12367, 12379, 12391, 12403, 12415, 12427, 12439, 12451, 12463, 12475, 12487, 12499, 12511, 12523, 12535, 12547, 12559, 12571, 12583, 12595, 12607, 12619, 12631, 12643, 12655, 12667, 12679, 12691, 12703, 12715, 12727, 12739, 12751, 12763, 12775, 12787, 12799, 12811, 12823, 12835, 12847, 12859, 12871, 12883, 12895, 12907, 12919, 12931, 12943, 12955, 12967, 12979, 12991, 13003, 13015, 13027, 13039, 13051, 13063, 13075, 13087, 13099, 13111, 13123, 13135, 13147, 13159, 13171, 13183, 13195, 13207, 13219, 13231, 13243, 13255, 13267, 13279, 13291, 13303, 13315, 13327, 13339, 13351, 13363, 13375, 13387, 13399, 13411, 13423, 13435, 13447, 13459, 13471, 13483, 13495, 13507, 13519, 13531, 13543, 13555, 13567, 13579, 13591, 13603, 13615, 13627, 13639, 13651, 13663, 13675, 13687, 13699, 13711, 13723, 13735, 13747, 13759, 13771, 13783, 13795, 13807, 13819, 13831, 13843, 13855, 13867, 13879, 13891, 13903, 13915, 13927, 13939, 13951, 13963, 13975, 13987, 13999, 14011, 14023, 14035, 14047, 14059, 14071, 14083, 14095, 14107, 14119, 14131, 14143, 14155, 14167, 14179, 14191, 14203, 14215, 14227, 14239, 14251, 14263, 14275, 14287, 14299, 14311, 14323, 14335, 14347, 14359, 14371, 14383, 14395, 14407, 14419, 14431, 14443, 14455, 14467, 14479, 14491, 14503, 14515, 14527, 14539, 14551, 14563, 14575, 14587, 14599, 14611, 14623, 14635, 14647, 14659, 14671, 14683, 14695, 14707, 14719, 14731, 14743, 14755, 14767, 14779, 14791, 14803, 14815, 14827, 14839, 14851, 14863, 14875, 14887, 14899, 14911, 14923, 14935, 14947, 14959, 14971, 14983, 14995, 15007, 15019, 15031, 15043, 15055, 15067, 15079, 15091, 15103, 15115, 15127, 15139, 15151, 15163, 15175, 15187, 15199, 15211, 15223, 15235, 15247, 15259, 15271, 15283, 15295, 15307, 15319, 15331, 15343, 15355, 15367, 15379, 15391, 15403, 15415, 15427, 15439, 15451, 15463, 15475, 15487, 15499, 15511, 15523, 15535, 15547, 15559, 15571, 15583, 15595, 15607, 15619, 15631, 15643, 15655, 15667, 15679, 15691, 15703, 15715, 15727, 15739, 15751, 15763, 15775, 15787, 15799, 15811, 15823, 15835, 15847, 15859, 15871, 15883, 15895, 15907, 15919, 15931, 15943, 15955, 15967, 15979, 15991, 16003, 16015, 16027, 16039, 16051, 16063, 16075, 16087, 16099, 16111, 16123, 16135, 16147, 16159, 16171, 16183, 16195, 16207, 16219, 16231, 16243, 16255, 16267, 16279, 16291, 16303, 16315, 16327, 16339, 16351, 16363, 16375, 16387, 16399, 16411, 16423, 16435, 16447, 16459, 16471, 16483, 16495, 16507, 16519, 16531, 16543, 16555, 16567, 16579, 16591, 16603, 16615, 16627, 16639, 16651, 16663, 16675, 16687, 16699, 16711, 16723, 16735, 16747, 16759, 16771, 16783, 16795, 16807, 16819, 16831, 16843, 16855, 16867, 16879, 16891, 16903, 16915, 16927, 16939, 16951, 16963, 16975, 16987, 16999, 17011, 17023, 17035, 17047, 17059, 17071, 17083, 17095, 17107, 17119, 17131, 17143, 17155, 17167, 17179, 17191, 17203, 17215, 17227, 17239, 17251, 17263, 17275, 17287, 17299, 17311, 17323, 17335, 17347, 17359, 17371, 17383, 17395, 17407, 17419, 17431, 17443, 17455, 17467, 17479, 17491, 17503, 17515, 17527, 17539, 17551, 17563, 17575, 17587, 17599, 17611, 17623, 17635, 17647, 17659, 17671, 17683, 17695, 17707, 17719, 17731, 17743, 17755, 17767, 17779, 17791, 17803, 17815, 17827, 17839, 17851, 17863, 17875, 17887, 17899, 17911, 17923, 17935, 17947, 17959, 17971, 17983, 17995, 18007, 18019, 18031, 18043, 18055, 18067, 18079, 18091, 18103, 18115, 18127, 18139, 18151, 18163, 18175, 18187, 18199, 18211, 18223, 18235, 18247, 18259, 18271, 18283, 18295, 18307, 18319, 18331, 18343, 18355, 18367, 18379, 18391, 18403, 18415, 18427, 18439, 18451, 18463, 18475, 18487, 18499, 18511, 18523, 18535, 18547, 18559, 18571, 18583, 18595, 18607, 18619, 18631, 18643, 18655, 18667, 18679, 18691, 18703, 18715, 18727, 18739, 18751, 18763, 18775, 18787, 18799, 18811, 18823, 18835, 18847, 18859, 18871, 18883, 18895, 18907, 18919, 18931, 18943, 18955, 18967, 18979, 18991, 19003, 19015, 19027, 19039, 19051, 19063, 19075, 19087, 19099, 19111, 19123, 19135, 19147, 19159, 19171, 19183, 19195, 19207, 19219, 19231, 19243, 19255, 19267, 19279, 19291, 19303, 19315, 19327, 19339, 19351, 19363, 19375, 19387, 19399, 19411, 19423, 19435, 19447, 19459, 19471, 19483, 19495, 19507, 19519, 19531, 19543, 19555, 19567, 19579, 19591, 19603, 19615, 19627, 19639, 19651, 19663, 19675, 19687, 19699, 19711, 19723, 19735, 19747, 19759, 19771, 19783, 19795, 19807, 19819, 19831, 19843, 19855, 19867, 19879, 19891, 19903, 19915, 19927, 19939, 19951, 19963, 19975, 19987, 19999, 20011, 20023, 20035, 20047, 20059, 20071, 20083, 20095, 20107, 20119, 20131, 20143, 20155, 20167, 20179, 20191, 20203, 20215, 20227, 20239, 20251, 20263, 20275, 20287, 20299, 20311, 20323, 20335, 20347, 20359, 20371, 20383, 20395, 20407, 20419, 20431, 20443, 20455, 20467, 20479, 20491, 20503, 20515, 20527, 20539, 20551, 20563, 20575, 20587, 20599, 20611, 20623, 20635, 20647, 20659, 20671, 20683, 20695, 20707, 20719, 20731, 20743, 20755, 20767, 20779, 20791, 20803, 20815, 20827, 20839, 20851, 20863, 20875, 20887, 20899, 20911, 20923, 20935, 20947, 20959, 20971, 20983, 20995, 21007, 21019, 21031, 21043, 21055, 21067, 21079, 21091, 21103, 21115, 21127, 21139, 21151, 21163, 21175, 21187, 21199, 21211, 21223, 21235, 21247, 21259, 21271, 21283, 21295, 21307, 21319, 21331, 21343, 21355, 21367, 21379, 21391, 21403, 21415, 21427, 21439, 21451, 21463, 21475, 21487, 21499, 21511, 21523, 21535, 21547, 21559, 21571, 21583, 21595, 21607, 21619, 21631, 21643, 21655, 21667, 21679, 21691, 21703, 21715, 21727, 21739, 21751, 21763, 21775, 21787, 21799, 21811, 21823, 21835, 21847, 21859, 21871, 21883, 21895, 21907, 21919, 21931, 21943, 21955, 21967, 21979, 21991, 22003, 22015, 22027, 22039, 22051, 22063, 22075, 22087, 22099, 22111, 22123, 22135, 22147, 22159, 22171, 22183, 22195, 22207, 22219, 22231, 22243, 22255, 22267, 22279, 22291, 22303, 22315, 22327, 22339, 22351, 22363, 22375, 22387, 22399, 22411, 22423, 22435, 22447, 22459, 22471, 22483, 22495, 22
```

```

Z = linkage(X, 'average')
plt.figure(figsize=(25,10))
plt.title("Dendrograma jerárquico de temperaturas en agosto (19961-2000)")
plt.xlabel("Registros tempetaturas agosto")
plt.ylabel("Distancia")
dendrogram(Z, leaf_rotation=90., leaf_font_size=15.5)
plt.show

```

```
<function matplotlib.pyplot.show(close=None, block=None)>
```



Se observa que los primeros 15 meses de agosto se separan bastante de los 15 últimos porque para que junten ambas ramas del dendrograma necesitan mucha distancia vertical.

Los clusters también se pueden representar mediante otro tipo de diagramas mediante random y se han representado en el eje x las estaciones meteorológicas (desde la 0 hasta la 480 tomadas cada 7 unidades) y en el eje y las temperaturas promedio recogidas en estas estaciones.

Se ha encontrado un obstáculo y es que las temperaturas recogidas vienen en formato de número pero con punto y no con coma, lo cual ha supuesto una dificultad añadida a la hora de trabajar con este dataset, pero se ha resuelto visualizándolas estadísticas de columna y colocando en el eje y la media de los valores más frecuentes de las columnas escogidas.

Con las estaciones meteorológicas y las medias de los valores más frecuentes (valores enteros) se han creado unas matrices a partir de las cuales se van a obtener unas visualizaciones



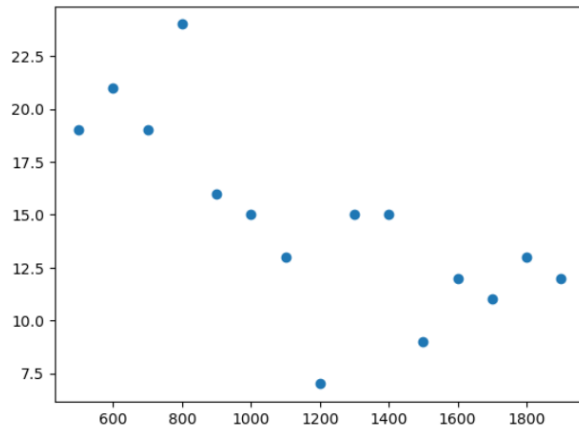
```
x = [500,600,700,800,900,1000,1100,1200,1300,1400,1500,1600,1700,1800,1900]
```

```
y = [19.54,21.03,18.74,24.24,16.16,14.71,13.53,6.59,15.52,14.71,9.33,11.71,11.03,12.59,11.58]
```

```
yentero=[19,21,19,24,16,15,13,7,15,15,9,12,11,13,12]
```

```
plt.scatter(x,yentero)
```

```
<matplotlib.collections.PathCollection at 0x2df88b055e0>
```



Lo cierto es que visualización no queda muy vistosa por la muestra reducida de datos con los que se han trabajado, pero se va ampliar la muestra, escogiendo todas las estaciones (500-1944) y todos los registros de agosto de 1973.

Ya para el fin de semana, ahora estoy con react.

### b. Árboles de clasificación

Para escoger una temperatura al azar y clasificarla en la estación meteorológica más probable.

Ya para el fin de semana, ahora estoy con react.

