

Informe cambio climático a partir de registro de temperaturas máximas

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	
Todos Tª máxima Tª mínima Precipitación total acumulada Velocidad del viento a 10m Velocidad máxima del viento a 10m Humedad relativa Percentil 95 de la temperatura máxima diaria Percentil 5 de la temperatura mínima diaria Descargar CSV de los datos de las estaciones	Todos HISTORICAL 20C3M C1L RCP4 5 RCP8 5 B1 A1B A2	Todos AEMET ENSEMBLES ESCENA ESTCENA CORDEX	Todos TXT SIG	

Variable	Escenario	Proyecto	Formato	Fichero
Tª máxima	20C3M	AEMET	TXT	Descargar fichero (4 MB)
Tª mínima	20C3M	AEMET	TXT	Descargar fichero (4 MB)
Precipitación total acumulada	20C3M	AEMET	TXT	Descargar fichero (3 MB)
Velocidad del viento a 10m	20C3M	AEMET	TXT	Descargar fichero (3 MB)
Velocidad máxima del viento a 10m	20C3M	AEMET	TXT	Descargar fichero (3 MB)
Humedad relativa	20C3M	AEMET	TXT	Descargar fichero (4 MB)
Percentil 95 de la temperatura máxima diaria	20C3M	AEMET	TXT	Descargar fichero (421 KB)
Percentil 5 de la temperatura mínima diaria	20C3M	AEMET	TXT	Descargar fichero (386 KB)
Percentil 95 de la precipitación diaria	20C3M	AEMET	TXT	Descargar fichero (415 KB)
Precipitación máxima en 24h	20C3M	AEMET	TXT	Descargar fichero (4 MB)
Nº de días con precipitación <1mm	20C3M	AEMET	TXT	Descargar fichero (1 MB)
Nº de días con precipitación >20mm	20C3M	AEMET	TXT	Descargar fichero (612 KB)
Máximo Nº de días consecutivos con precipitación <1mm	20C3M	AEMET	TXT	Descargar fichero (147 KB)

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 Codigo 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	C685I	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.

TXMM_RCA_ECHAMS_20C3M_196101_200012

Archivo Editar Ver Insertar Formato Datos Herramientas Extensiones Ayuda Última modificación hace 4 minutos

100% 21/09/2022

Problemas...

B I X

123

Compartir

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
332	19808	24.77	25.06	27.07	27.07	29.59	27.63	22.90	29.84	31.98	31.98	32.42	31.59	25.92	22.77	27.51	33.24	33.96
333	19809	23.90	24.64	25.25	25.25	27.84	26.62	22.87	26.81	28.78	28.78	29.39	29.09	25.17	22.65	24.90	29.04	29.86
334	19810	20.18	20.67	20.68	20.68	21.17	20.87	19.82	21.72	21.95	21.95	21.28	20.48	19.68	19.29	21.36	22.90	22.82
335	19801	18.40	18.54	18.74	18.74	18.96	18.69	17.96	19.37	19.52	19.52	18.40	17.46	17.29	17.37	19.07	20.19	19.94
336	19812	16.90	16.78	16.98	16.98	16.73	16.64	16.36	17.34	17.34	15.79	14.96	15.40	15.79	17.32	17.84	17.71	17.71
337	19801	15.59	15.67	15.64	15.64	15.53	15.64	15.39	16.09	15.92	15.92	14.61	14.05	14.56	14.84	16.08	16.56	16.22
338	19802	13.67	13.71	13.50	13.50	13.17	13.34	13.79	13.89	13.36	13.36	12.11	11.51	12.00	12.62	14.01	14.04	13.39
339	19803	15.09	15.76	15.23	15.23	15.76	15.78	15.63	15.68	15.68	14.48	14.31	14.48	15.50	16.23	15.85		
340	19804	16.81	17.03	16.34	16.34	17.62	17.50	15.62	17.09	17.34	17.34	14.87	17.04	16.59	15.67	15.20	16.20	17.75
341	19805	16.95	18.28	18.05	18.05	20.08	19.48	16.78	19.59	20.73	20.73	20.66	19.69	17.84	16.44	17.98	21.44	21.23
342	19806	21.46	22.80	23.49	23.49	26.68	25.30	20.68	25.74	28.33	28.33	28.85	27.51	23.52	20.58	23.21	28.86	29.86
343	19807	24.10	25.56	26.45	26.45	29.55	28.07	23.34	29.07	31.86	31.86	32.70	31.53	26.49	23.48	26.17	32.68	33.92
344	19808	24.35	25.72	26.36	26.36	30.28	29.93	24.57	30.35	31.77	31.77	32.63	32.35	27.55	24.72	25.41	31.58	33.19
345	19809	24.51	25.76	26.72	26.72	29.55	27.54	24.04	26.96	31.65	31.65	31.77	30.41	25.90	23.83	26.19	32.02	32.97
346	19810	20.19	20.62	20.42	20.42	20.88	20.71	20.00	20.99	21.11	21.11	20.42	19.66	18.97	19.14	20.52	21.47	21.19
347	19801	17.43	17.51	17.38	17.38	17.15	17.18	17.27	17.66	17.35	17.35	16.27	15.65	15.94	16.55	17.57	17.94	17.46
348	19812	14.84	14.90	14.85	14.85	14.59	14.74	14.98	15.06	14.90	14.90	13.78	13.26	13.71	14.24	15.04	15.18	14.92
349	19801	14.76	14.85	14.68	14.68	14.45	14.49	14.68	14.87	14.56	14.56	13.51	12.93	13.40	14.09	14.98	15.03	14.49
350	19802	14.95	15.13	15.28	15.28	15.68	15.49	14.87	15.65	16.31	16.31	15.01	14.26	14.62	14.49	15.21	16.29	16.48
351	19803	15.32	15.93	15.72	15.72	16.72	16.50	15.08	16.36	16.82	16.82	16.21	15.74	15.13	14.72	15.86	17.14	16.91
352	19804	16.16	17.36	16.82	16.82	18.50	18.10	15.91	17.46	18.30	18.30	16.43	17.88	16.43	15.47	16.39	18.35	18.33
353	19805	18.76	20.45	20.79	20.79	24.17	22.84	18.34	22.23	24.92	24.92	25.53	23.66	20.84	18.37	19.75	24.56	25.33
354	19806	22.16	22.94	24.88	24.88	27.97	25.79	20.76	28.30	31.21	31.21	31.40	29.69	24.29	20.79	25.19	32.50	33.63
355	19807	24.56	25.22	26.68	26.68	29.89	27.94	22.83	27.94	29.08	29.08	31.89	31.89	32.78	32.85	28.56	32.39	33.73
356	19808	25.93	27.12	28.61	28.61	31.12	29.30	25.22	30.30	33.02	33.02	33.65	32.68	27.94	25.24	27.54	33.49	34.88
357	19809	25.49	26.33	27.21	27.21	29.82	28.22	24.68	29.11	31.38	31.38	31.57	30.77	26.91	24.69	27.01	31.93	32.96
358	19810	23.14	23.35	24.55	24.55	25.83	24.46	22.08	26.41	27.88	27.88	26.97	25.85	23.16	21.72	24.79	28.82	29.38
359	19801	18.67	18.84	18.37	18.37	18.28	18.44	18.67	18.51	18.26	18.26	17.42	16.97	17.33	17.87	18.65	18.70	18.42
360	19802	16.92	16.55	16.95	16.95	16.52	16.45	16.42	17.26	17.39	17.39	15.84	15.15	15.75	16.00	17.07	17.87	17.76
361	19803	14.86	14.68	14.68	14.68	14.22	14.24	14.43	14.89	14.49	14.49	13.25	12.63	12.93	13.47	15.19	15.02	14.53
362	19804	14.40	14.46	14.54	14.54	14.42	14.27	14.00	14.95	14.76	14.76	13.74	13.06	12.98	13.35	14.93	15.37	14.96
363	19805	15.15	15.61	15.23	15.23	15.58	15.59	15.00	15.62	15.48	15.48	14.70	14.23	14.26	14.50	15.51	15.90	15.42
364	19806	15.37	16.38	15.93	15.93	17.32	16.06	17.12	15.43	16.52	17.10	17.10	17.02	16.60	15.63	15.07	15.84	17.35
365	19807	18.36	19.28	19.79	19.79	22.06	20.87	17.62	21.31	22.98	22.98	23.04	21.56	19.31	17.61	19.53	23.43	23.68
366	19808	21.86	23.93	24.36	24.36	28.79	26.87	21.20	26.75	30.57	30.57	31.46	30.00	26.00	21.25	23.29	30.01	31.65

Visualización tabla Tª 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	10.46	9.88	12.85
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).

Archivo Editar Buscar Vista Codificación Lenguaje Configuración Herramientas Macro Ejecutar Plugins Ventana ?

TSMM_RCA_HADCM3G0_20CM_196101_200012.csv

250	196110	19.82	19.90	20.90	20.90	21.83	20.63	18.68	22.69	23.68	23.68	22.86	21.57	19.50	18.19	21.27	24.79	25.00	25.00	23.75	2
251	196111	14.00	14.20	14.49	14.49	15.03	14.58	12.97	15.24	15.74	15.74	14.74	14.14	13.35	12.33	14.36	16.20	16.16	16.16	14.67	1
252	196112	11.23	11.78	11.69	11.89	13.11	12.56	10.00	12.76	13.95	13.95	13.04	12.57	11.48	9.25	11.73	14.03	14.50	14.50	13.25	1
253	196201	10.43	10.51	11.21	11.21	11.88	11.24	9.27	12.41	13.63	13.63	12.35	11.55	10.57	8.63	11.10	14.07	14.66	14.66	13.02	1
254	196202	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1
255	196203	14.11	14.69	15.22	15.22	16.39	15.72	13.66	16.60	17.72	17.72	16.66	15.46	14.31	12.96	15.38	18.35	18.52	18.52	16.78	1
256	196204	19.29	20.50	20.93	20.93	23.48	22.15	18.45	22.48	24.06	24.06	24.20	23.28	20.22	18.28	20.81	24.61	24.75	24.75	23.84	2
257	196205	21.39	22.63	22.72	22.72	25.27	24.23	21.34	24.39	25.89	25.89	26.13	25.44	22.58	21.05	22.56	26.55	26.78	26.78	26.19	2
258	196206	26.32	27.73	27.82	27.82	30.43	29.31	26.48	29.54	30.98	30.98	31.79	31.18	27.31	25.87	28.15	31.49	31.79	31.79	31.60	3
259	196207	30.98	32.75	32.29	32.29	35.76	34.76	31.35	33.51	35.93	35.93	37.62	37.31	33.10	31.28	33.39	35.59	36.82	36.82	37.71	3
260	196208	27.94	29.27	29.05	29.05	31.20	30.56	28.41	30.33	31.63	31.63	32.38	31.85	29.00	28.03	28.97	32.02	32.31	32.31	32.35	3
261	196209	23.70	24.02	25.20	25.20	26.52	25.35	22.85	27.29	28.01	28.01	28.04	26.57	23.88	22.13	26.26	29.38	29.49	29.49	28.84	2
262	196210	19.72	20.40	20.66	20.66	22.18	21.16	18.49	22.05	23.07	23.07	22.52	21.57	19.57	17.89	20.63	23.65	23.96	23.96	22.83	2
263	196211	14.62	14.93	15.06	15.06	16.01	15.51	13.46	15.88	16.78	16.78	15.68	15.07	14.31	12.81	14.95	17.04	17.29	17.29	15.79	1
264	196212	12.15	12.26	12.70	12.70	13.04	12.27	10.68	13.60	14.06	14.06	12.63	11.66	10.84	9.93	12.62	14.56	14.54	14.54	12.87	1
265	196301	10.37	10.37	10.83	10.83	10.90	10.40	9.28	11.56	11.78	11.78	10.45	9.62	9.14	8.56	10.94	12.28	12.12	12.12	10.59	1
266	196302	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1
267	196303	15.30	16.11	16.77	16.77	19.52	17.33	14.64	18.02	20.25	20.25	19.08	17.75	16.20	14.10	15.99	19.94	20.72	20.72	19.25	1
268	196304	15.47	16.53	16.14	16.14	17.64	17.32	15.14	17.14	17.97	17.97	17.65	17.00	15.79	14.59	15.89	18.57	18.43	18.43	17.32	1
269	196305	18.56	19.83	19.54	19.54	21.78	21.36	18.64	20.79	22.04	22.04	22.19	20.92	19.42	18.22	19.25	22.39	22.50	22.50	21.06	1
270	196306	23.11	24.23	24.17	24.17	26.03	25.53	23.25	25.76	26.68	26.68	27.04	26.12	24.07	22.97	24.62	27.51	27.57	27.57	26.87	2
271	196307	25.81	27.47	26.95	26.95	29.78	29.36	26.48	28.40	30.06	30.06	31.22	30.95	27.97	26.45	28.84	30.44	31.00	31.00	31.00	2
272	196308	28.47	30.47	30.03	30.03	33.92	32.72	28.90	31.56	34.37	34.37	35.89	35.81	31.22	29.09	29.25	34.00	35.34	35.34	35.95	3
273	196309	24.50	25.41	25.75	25.75	27.81	26.64	23.81	27.47	28.84	28.84	28.78	27.81	24.73	23.42	25.86	29.67	30.00	30.00	29.00	2
274	196310	17.52	17.98	18.30	18.30	19.50	18.78	16.77	19.50	20.50	20.50	19.44	18.70	17.17	16.19	18.30	21.05	21.19	21.19	19.73	1
275	196311	16.12	16.60	16.57	16.57	17.48	17.04	15.03	17.53	18.20	18.20	17.12	16.51	15.65	14.45	16.45	18.64	18.59	18.59	17.20	1
276	196312	14.83	15.47	15.18	15.18	16.10	15.66	13.69	15.83	16.46	16.46	15.57	14.99	14.22	13.04	14.85	16.58	16.50	16.50	15.30	1
277	196401	12.51	12.88	12.95	12.95	13.51	12.98	11.40	13.62	14.10	14.10	13.07	12.40	11.52	10.60	12.82	14.41	14.24	14.24	13.01	1
278	196402	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1
279	196403	12.97	13.34	13.25	13.25	13.80	13.47	12.13	13.86	14.12	14.12	13.30	12.67	12.03	11.33	13.08	14.55	14.23	14.23	13.14	1
280	196404	14.74	15.58	15.19	15.19	16.72	16.30	14.24	15.69	16.42	16.42	16.33	15.96	14.72	13.71	14.58	16.47	16.32	16.32	15.74	1
281	196405	17.15	18.07	17.54	17.54	19.23	18.99	17.35	17.91	18.76	18.76	19.13	18.79	17.71	17.02	16.88	18.71	18.63	18.63	18.30	1
282	196406	24.27	25.29	25.40	25.40	27.69	26.82	24.87	26.40	28.18	28.18	28.69	27.12	25.60	24.62	24.99	28.29	28.69	28.69	27.74	2
283	196407	28.64	29.80	29.19	29.19	32.72	31.48	28.59	31.83	33.72	33.72	34.52	33.19	29.59	27.86	30.16	34.23	34.83	34.83	34.38	3
284	196408	27.80	29.75	29.11	29.11	33.13	31.92	28.28	30.10	32.85	32.85	34.42	34.27	30.30	28.14	27.61	31.97	33.22	33.22	33.83	3
285	196409	23.80	24.70	24.89	24.89	26.83	25.75	23.67	26.47	27.67	27.67	27.94	27.06	24.22	22.97	25.03	28.42	28.69	28.69	28.08	2
286	196410	19.83	20.12	20.44	20.44	21.12	20.61	19.09	21.47	22.05	22.05	21.34	20.41	19.39	18.61	20.53	22.70	22.69	22.69	21.55	1
287	196411	14.48	14.92	14.78	14.78	15.59	15.26	13.40	15.48	16.14	16.14	15.14	14.64	13.82	12.70	14.48	16.40	16.42	16.42	15.04	1
288	196412	12.84	12.84	13.49	13.49	13.77	13.16	11.71	14.68	15.37	15.37	13.87	12.91	12.21	11.10	13.57	16.09	16.18	16.18	14.45	1
289	196501	12.58	12.78	13.13	13.13	13.74	13.14	11.30	14.24	15.11	15.11	13.61	12.77	11.91	10.66	13.02	15.70	15.89	15.89	14.02	1
290	196502	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1
291	196503	16.27	17.10	17.01	17.01	18.20	17.63	15.56	17.91	18.77	18.77	17.93	17.20	16.01	14.87	16.56	19.20	18.93	18.93	17.81	1

Normal text file length: 6246240 lines: 481 Ln: 1 Col: 1 Pos: 1 Windows (CR LF) UTF-8 INS

Tiene sentido que falten valores en un dataset porque los datos en crudo se han obtenido de estaciones ubicadas, tanto en lugares institucionales (aeropuertos, colegios, etc....) como particulares y las lecturas no siempre se han podido hacer en tiempo y forma.

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).

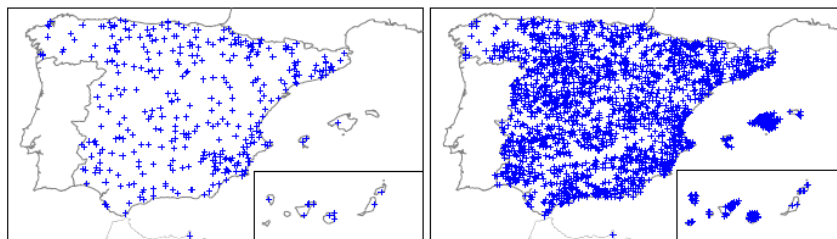


Figura 6. Distribución de las localizaciones de las 374 estaciones con datos de temperatura y las 2.321 estaciones con datos de precipitación para los productos de Escenarios PNACC Datos mensuales

Mapa de estaciones climatológicas ubicadas en España

La primera conclusión de este estudio se refiere a la detección de errores provocados por desplazamientos de ubicación, cambios en la instrumentación o el entorno, calibraciones, etc... lo que se refleja en las series como cambios artificiales añadidos con formato diferente.

No obstante, la serie que se ha utilizado para estudiar la influencia de la temperatura en el cambio climático tiene la misma tendencia que la señal climática real y los datos climáticos son homogéneos, con largo recorrido temporal y probada calidad de modo que aceptan como válidos para esta investigación de carácter general, sin especificar zonas concretas de la geografía española.

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

Plots y visualización de datos

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

Plots y visualización de los datos

```
: 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

Visualización básica: Scatterplot (nube de dispersión/puntos)

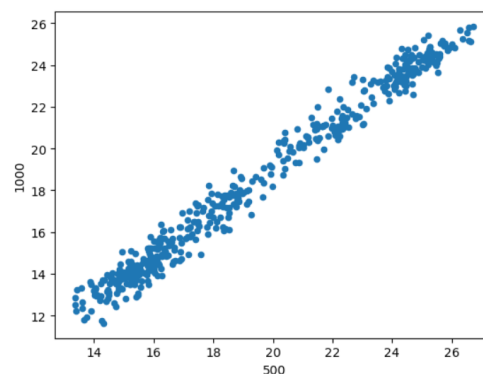
Representación de TODOS los puntos que recoge el dataset, en la estación 500 y la estación 1000 a lo largo del tiempo desde enero de 1961 hasta diciembre del año 2000.

Parece que hay una tendencia ascendente y relativamente similar entre ambas

```
data1_1961=data1
```

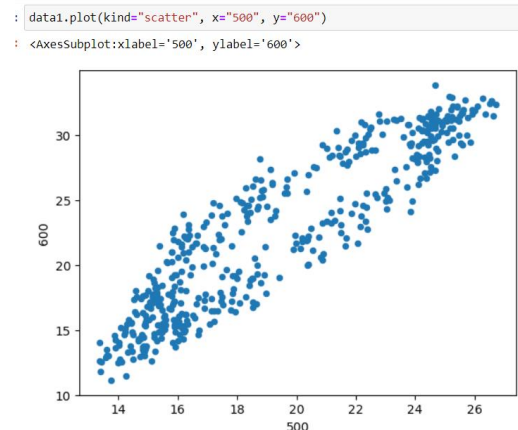
```
data1.plot(kind="scatter", x="500", y="1000")
```

```
<AxesSubplot:xlabel='500', ylabel='1000'>
```



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 en 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.

2

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

```
[196101.0,
 16.08,
 16.33,
 16.08,
 16.08,
 16.22,
 16.12,
 16.01,
 16.36,
 16.5,
 16.5,
 15.39,
 14.73,
 15.17,
 15.48,
 16.23,
 16.75,
 16.56,
 16.56,
```

```
1 | a = set(desired_columns)
2 | b = set(all_columns_list)
3 | sublist = b-a
4 | sublist = list(sublist)
```

Concatenación de datos (función merge)

Tipos de joins

Inner join

Left join