Sección 1: Carga de datos

Se cargaron los datos desde el archivo CSV para el sensor 4e0031000251353337353037

	date	value	variable	units	range
0	2018-05-08 19:46:55.615	-46.581871	Temperature	Centigrade	[-10, 85]
1	2018-05-08 19:47:27.362	-46.581871	Temperature	Centigrade	[-10, 85]
2	2018-05-08 19:48:01.573	-46.581871	Temperature	Centigrade	[-10, 85]
3	2018-05-10 14:10:54.093	-46.581871	Temperature	Centigrade	[-10, 85]
4	2018-05-10 14:18:16.500	28.686863	Temperature	Centigrade	[-10, 85]
5	2018-05-10 14:28:39.737	29.308918	Temperature	Centigrade	[-10, 85]
6	2018-05-10 14:38:39.639	28.290033	Temperature	Centigrade	[-10, 85]
7	2018-05-10 14:40:59.903	28.043358	Temperature	Centigrade	[-10, 85]
8	2018-05-10 14:46:44.209	28.987165	Temperature	Centigrade	[-10, 85]
9	2018-05-10 14:57:08.894	27.710878	Temperature	Centigrade	[-10, 85]

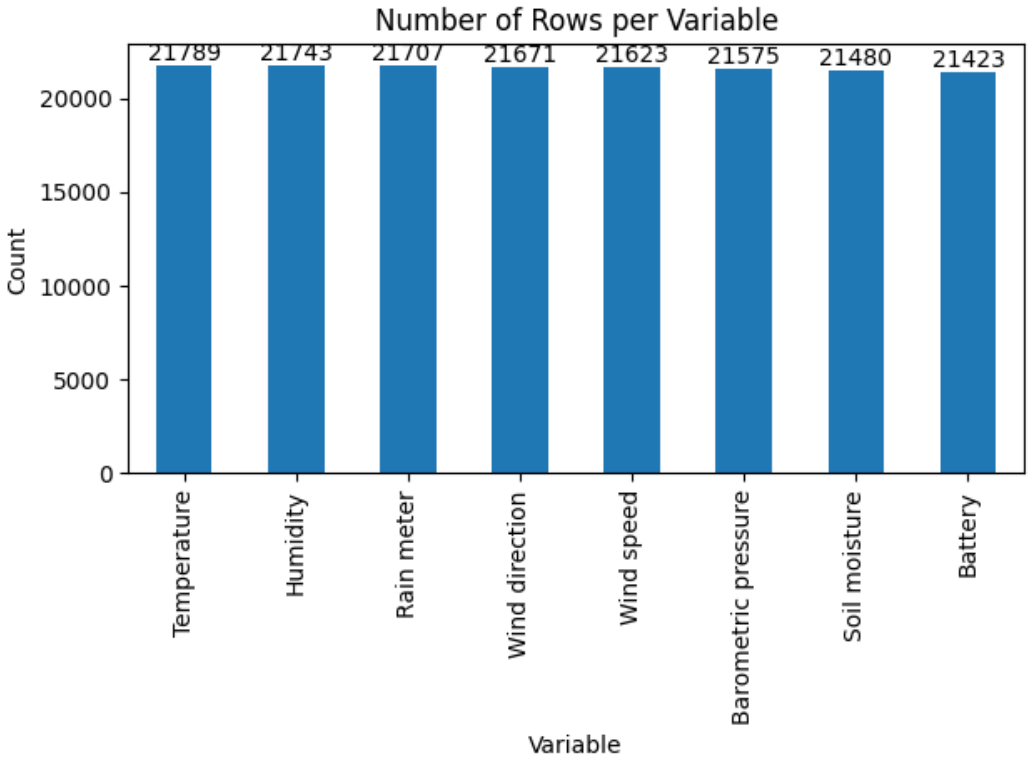
Se conviertien las marcas de tiempo Unix a formato de fecha y hora y se establece como índice. Además se ordena el dataframe por fecha.

	date	value	variable	units	range
2018-05-08	**19:46:55.615000**	-46.581871	Temperature	Centigrade	[-10, 85]
2018-05-08	**19:46:55.809000**	-5.809265	Humidity	Percentage	[0, 80]
2018-05-08	**19:46:56.019000**	0.000000	Rain meter	millilitres (mm)	[-1, 7]
2018-05-08	**19:46:56.439000**	0.000000	Wind speed	km/h	nan
2018-05-08	**19:46:56.649000**	-9.990000	Barometric pressure	Hectopascal	[500, 1100]
2018-05-08	**19:46:57.085000**	903.000000	Soil moisture	Percentage	[0, 85]
2018-05-08	**19:46:57.295000**	87.925781	Battery	Percentage	[0, 100]
2018-05-08	**19:47:27.362000**	-46.581871	Temperature	Centigrade	[-10, 85]
2018-05-08	**19:47:27.554000**	-5.809265	Humidity	Percentage	[0, 80]
2018-05-08	**19:47:27.764000**	0.000000	Rain meter	millilitres (mm)	[-1, 7]

Se convierten los valores de 'Soil moisture' a porcentaje.

	date	value	variable	units	range
2018-05-08	**19:46:57.085000**	22.045898	Soil moisture	Percentage	[0, 85]
2018-05-08	**19:47:28.826000**	19.555664	Soil moisture	Percentage	[0, 85]
2018-05-10	**14:10:55.548000**	30.249023	Soil moisture	Percentage	[0, 85]
2018-05-10	**14:18:17.964000**	79.687500	Soil moisture	Percentage	[0, 85]
2018-05-10	**14:28:41.208000**	0.048828	Soil moisture	Percentage	[0, 85]
2018-05-10	**14:38:41.113000**	85.449219	Soil moisture	Percentage	[0, 85]
2018-05-10	**14:41:01.356000**	0.097656	Soil moisture	Percentage	[0, 85]
2018-05-10	**14:46:45.667000**	74.536133	Soil moisture	Percentage	[0, 85]
2018-05-10	**14:57:10.354000**	0.000000	Soil moisture	Percentage	[0, 85]
2018-05-10	**15:05:28.266000**	86.083984	Soil moisture	Percentage	[0, 85]

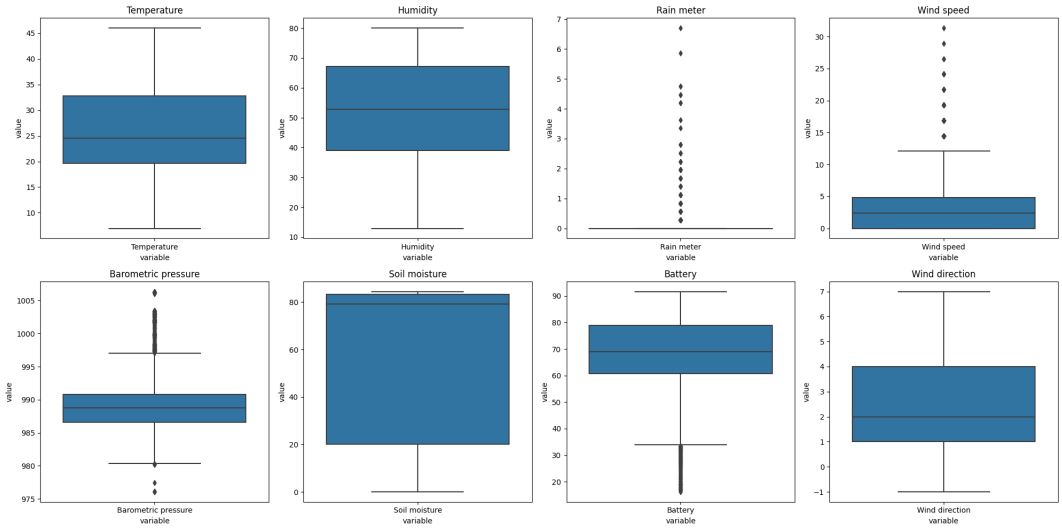
Se realiza un análisis de cada variable y se calcula el porcentaje de valores que están fuera de rango y cuantos de ellos corresponden a valores nulos.



Barometric pressure: 0.03% (null: 0.00%)
Battery: 0.00% (null: 0.00%)
Humidity: 13.12% (null: 0.00%)
Rain meter: 0.01% (null: 0.00%)
Soil moisture: 99.79% (null: 0.00%)
Temperature: 0.02% (null: 0.00%)
Wind direction: 0.00% (null: 0.00%)
Wind speed: 100.00% (null: 0.00%)

Variables with more than 5% null values:

Se ponen a nulo los datos que están fuera de rango y se vuelve a calcular que porcentaje de valores corresponden a valores nulo.



Barometric pressure: 0.03% (null: 0.03%)
Battery: 0.00% (null: 0.00%)

Humidity: 13.12% (null: 13.12%)
Rain meter: 0.01% (null: 0.01%)
Soil moisture: 99.79% (null: 99.79%)
Temperature: 0.02% (null: 0.02%)
Wind direction: 0.00% (null: 0.00%)
Wind speed: 100.00% (null: 0.00%)

Variables with more than 5% null values: Humidity, Soil moisture

Se simplifica el dataframe eliminando columnas innecesarias ('units', 'range', 'within_range') y se pivota uniendo todas las mediciones realizadas en el mismo minuto por todos los sensores. Se elimina también la variable 'Wind direction' dado que tiende a contener muchos valores vacios y el número de mediciones suele ser menor al resto de variables.

variable	Barometric pressure	Battery	Humidity	Rain meter	Soil moisture	Temperature	Wind speed
date							
2018-05-08 19:46:00	nan	87.925781	nan	0.000000	22.045898	nan	0.000000
2018-05-08 19:47:00	nan	87.925781	nan	0.000000	19.555664	nan	0.000000
2018-05-08 19:48:00	nan	nan	nan	0.000000	nan	nan	nan
2018-05-10 14:10:00	nan	87.492188	nan	0.000000	30.249023	nan	0.000000
2018-05-10 14:18:00	nan	91.488281	56.927246	0.000000	79.687500	28.686863	0.000000
2018-05-10 14:28:00	1006.349976	91.648438	36.335510	0.000000	0.048828	29.308918	0.000000
2018-05-10 14:38:00	1006.202515	91.648438	35.236877	0.000000	nan	28.290033	0.000000
2018-05-10 14:40:00	nan	nan	nan	nan	nan	28.043358	nan
2018-05-10 14:41:00	1006.357483	91.648438	36.976379	0.000000	0.097656	nan	0.000000
2018-05-10 14:46:00	1006.325012	90.539063	35.702271	0.000000	74.536133	28.987165	0.000000

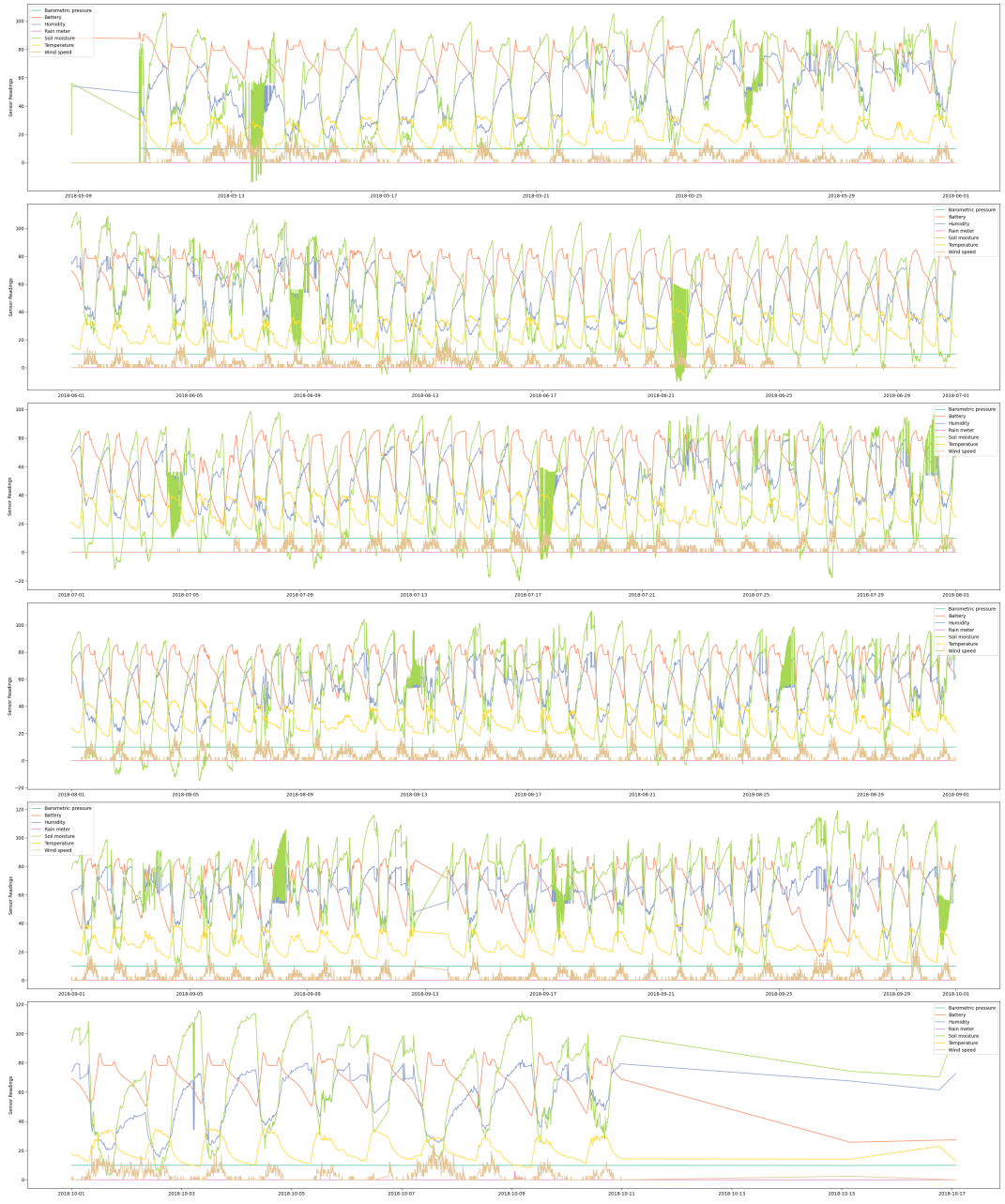
Se imputan los valores faltantes usando el método IterativeImpute. El método IterativeImputer es una técnica de imputación de valores que utiliza un regresor bayesiano como estimador y utilizando el resto de columnas para esa fila.

variable	Barometric pressure	Battery	Humidity	Rain meter	Soil moisture	Temperature	Wind speed
date							
2018-05-08 19:46:00	nan	87.925781	47.677289	0.000000	22.045898	nan	0.000000
2018-05-08 19:47:00	nan	87.925781	47.224471	0.000000	19.555664	nan	0.000000
2018-05-08 19:48:00	nan	nan	53.959871	0.000000	56.009866	nan	nan
2018-05-10 14:10:00	nan	87.492188	49.168924	0.000000	30.249023	nan	0.000000
2018-05-10 14:18:00	nan	91.488281	56.927246	0.000000	79.687500	28.686863	0.000000
2018-05-10 14:28:00	1006.349976	91.648438	36.335510	0.000000	0.048828	29.308918	0.000000
2018-05-10 14:38:00	1006.202515	91.648438	35.236877	0.000000	57.609051	28.290033	0.000000
2018-05-10 14:40:00	nan	nan	51.072574	nan	49.043901	28.043358	nan
2018-05-10 14:41:00	1006.357483	91.648438	36.976379	0.000000	0.097656	nan	0.000000
2018-05-10 14:46:00	1006.325012	90.539063	35.702271	0.000000	74.536133	28.987165	0.000000

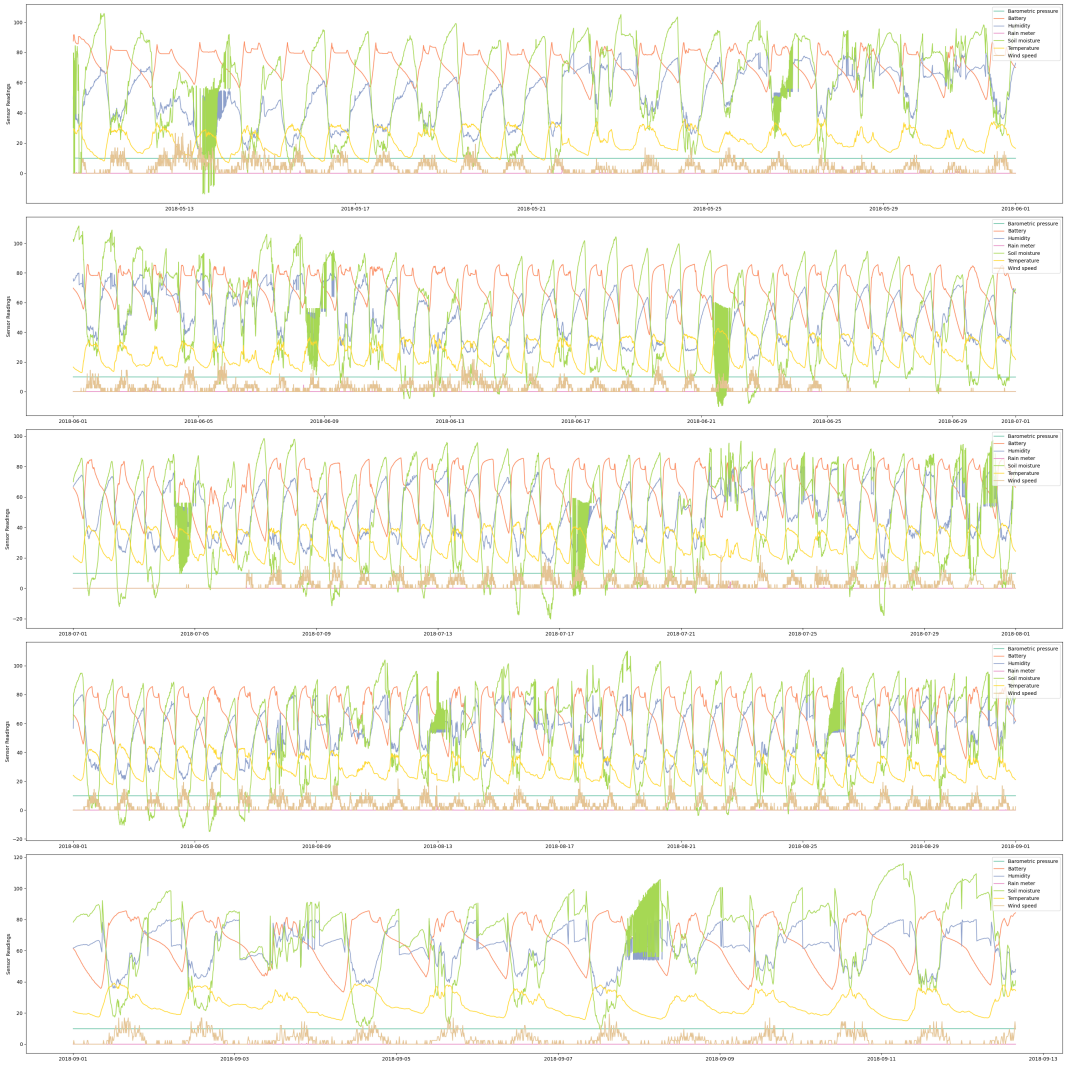
Se realizó una interpolación basada en el tiempo en el dataframe y redondearon todos los valores a 2 decimales.

variable	Barometric pressure	Battery	Humidity	Rain meter	Soil moisture	Temperature	Wind speed
date							
2018-05-08 19:46:00	nan	87.930000	47.680000	0.000000	22.050000	nan	0.000000
2018-05-08 19:47:00	nan	87.930000	47.220000	0.000000	19.560000	nan	0.000000
2018-05-08 19:48:00	nan	87.930000	53.960000	0.000000	56.010000	nan	0.000000
2018-05-10 14:10:00	nan	87.490000	49.170000	0.000000	30.250000	nan	0.000000
2018-05-10 14:18:00	nan	91.490000	56.930000	0.000000	79.690000	28.690000	0.000000
2018-05-10 14:28:00	1006.350000	91.650000	36.340000	0.000000	0.050000	29.310000	0.000000
2018-05-10 14:38:00	1006.200000	91.650000	35.240000	0.000000	57.610000	28.290000	0.000000
2018-05-10 14:40:00	1006.310000	91.650000	51.070000	0.000000	49.040000	28.040000	0.000000
2018-05-10 14:41:00	1006.360000	91.650000	36.980000	0.000000	0.100000	28.200000	0.000000
2018-05-10 14:46:00	1006.330000	90.540000	35.700000	0.000000	74.540000	28.990000	0.000000

Lecturas de los sensores después de toda la limpieza y procesamiento de datos.



Lecturas de los sensores después de eliminar los periodos invalidos.



Se realizó una interpolación basada en el tiempo en el dataframe de nuevo.

variable	Barometric pressure	Battery	Humidity	Rain meter	Soil moisture	Temperature	Wind speed
date							
2018-05-10 14:28:00	1006.350000	91.650000	36.340000	0.000000	0.050000	29.310000	0.000000
2018-05-10 14:38:00	1006.200000	91.650000	35.240000	0.000000	57.610000	28.290000	0.000000
2018-05-10 14:40:00	1006.310000	91.650000	51.070000	0.000000	49.040000	28.040000	0.000000
2018-05-10 14:41:00	1006.360000	91.650000	36.980000	0.000000	0.100000	28.200000	0.000000
2018-05-10 14:46:00	1006.330000	90.540000	35.700000	0.000000	74.540000	28.990000	0.000000
2018-05-10 14:57:00	1006.200000	88.960000	34.120000	0.000000	0.000000	27.710000	0.000000
2018-05-10 15:05:00	1006.260000	87.930000	36.780000	0.280000	76.200000	26.780000	0.000000
2018-05-10 15:06:00	1006.180000	87.930000	37.400000	0.000000	63.010000	26.710000	0.000000
2018-05-10 15:11:00	1006.190000	87.490000	36.700000	0.000000	80.270000	27.220000	0.000000
2018-05-10 15:12:00	1006.150000	87.420000	36.190000	0.000000	78.940000	27.280000	0.000000

Se añade una nueva columna para cada entrada con el numero de minutos total de luz

variable	Barometric pressure	Battery	Humidity	Rain meter	Soil moisture	Temperature	Wind speed	day_length
date								
2018-05-10 14:28:00	1006.350000	91.650000	36.340000	0.000000	0.050000	29.310000	0.000000	857
2018-05-10 14:38:00	1006.200000	91.650000	35.240000	0.000000	57.610000	28.290000	0.000000	857
2018-05-10 14:40:00	1006.310000	91.650000	51.070000	0.000000	49.040000	28.040000	0.000000	857
2018-05-10 14:41:00	1006.360000	91.650000	36.980000	0.000000	0.100000	28.200000	0.000000	857
2018-05-10 14:46:00	1006.330000	90.540000	35.700000	0.000000	74.540000	28.990000	0.000000	857
2018-05-10 14:57:00	1006.200000	88.960000	34.120000	0.000000	0.000000	27.710000	0.000000	857
2018-05-10 15:05:00	1006.260000	87.930000	36.780000	0.280000	76.200000	26.780000	0.000000	857
2018-05-10 15:06:00	1006.180000	87.930000	37.400000	0.000000	63.010000	26.710000	0.000000	857
2018-05-10 15:11:00	1006.190000	87.490000	36.700000	0.000000	80.270000	27.220000	0.000000	857
2018-05-10 15:12:00	1006.150000	87.420000	36.190000	0.000000	78.940000	27.280000	0.000000	857