Sección 1: Carga de datos

Se cargaron los datos desde el archivo CSV para el sensor 46004e000251353337353037

	date	value	variable	units	range
0	2018-04-25 21:05:41.846	-46.581871	Temperature	Centigrade	[-10, 85]
1	2018-05-02 10:07:03.481	-46.581871	Temperature	Centigrade	[-10, 85]
2	2018-05-02 10:08:39.160	-46.581871	Temperature	Centigrade	[-10, 85]
3	2018-05-02 10:19:05.152	-46.581871	Temperature	Centigrade	[-10, 85]
4	2018-05-02 10:29:05.109	-46.581871	Temperature	Centigrade	[-10, 85]
5	2018-05-02 10:39:03.216	-46.581871	Temperature	Centigrade	[-10, 85]
6	2018-05-02 10:49:04.152	-46.581871	Temperature	Centigrade	[-10, 85]
7	2018-05-02 10:59:05.722	-46.581871	Temperature	Centigrade	[-10, 85]
8	2018-05-02 11:09:05.825	-46.581871	Temperature	Centigrade	[-10, 85]
9	2018-05-02 11:19:05.235	-46.581871	Temperature	Centigrade	[-10, 85]

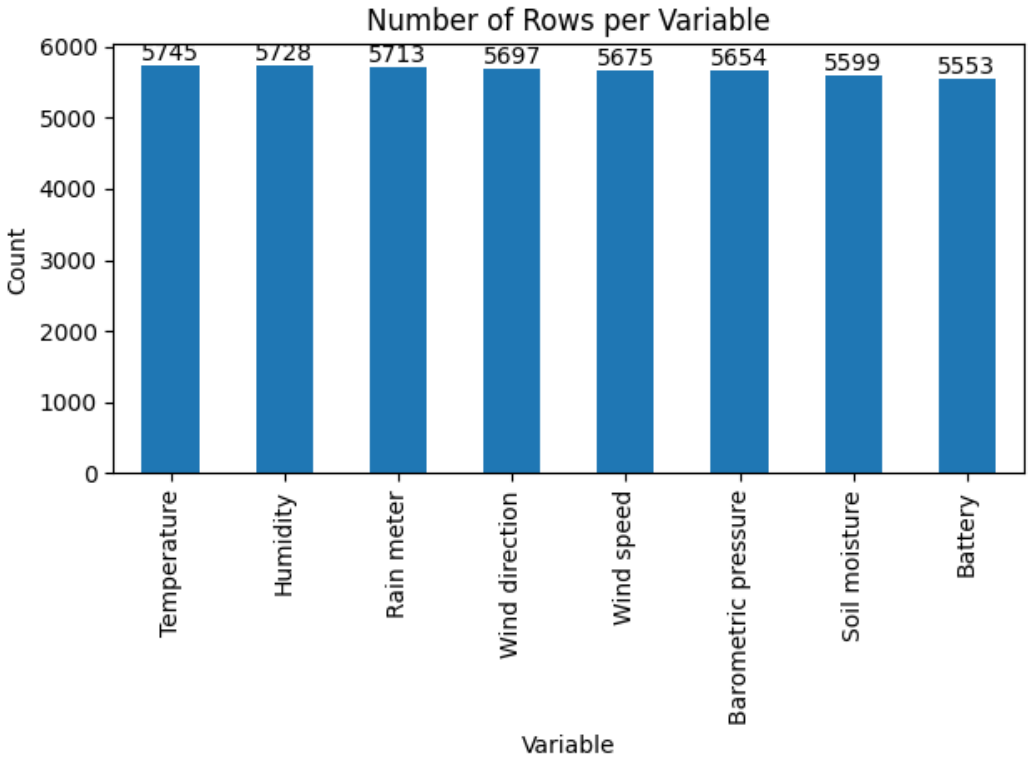
Se conviertien las marcas de tiempo Unix a formato de fecha y hora y se establece como índice. Además se ordena el dataframe por fecha.

date	value	variable	units	range
2018-04-25 19:05:36	-1.000000	Wind direction	Direction (degrees)	[-1, 7]
2018-04-25 21:05:41.846000	-46.581871	Temperature	Centigrade	[-10, 85]
2018-04-25 21:05:42.056000	-5.809265	Humidity	Percentage	[0, 80]
2018-04-25 21:05:42.266000	0.000000	Rain meter	millilitres (mm)	[-1, 7]
2018-04-25 21:05:42.685000	0.000000	Wind speed	km/h	nan
2018-04-25 21:05:42.912000	-9.990000	Barometric pressure	Hectopascal	[500, 1100]
2018-04-25 21:05:43.316000	1304.000000	Soil moisture	Percentage	[0, 85]
2018-04-25 21:05:43.526000	81.726563	Battery	Percentage	[0, 100]
2018-05-02 08:08:30	-1.000000	Wind direction	Direction (degrees)	[-1, 7]
2018-05-02 08:18:30	-1.000000	Wind direction	Direction (degrees)	[-1, 7]

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

date	value	variable	units	range
2018-04-25 21:05:43.316000	31.835938	Soil moisture	Percentage	[0, 85]
2018-05-02 10:07:04.958000	27.050781	Soil moisture	Percentage	[0, 85]
2018-05-02 10:08:40.632000	28.857422	Soil moisture	Percentage	[0, 85]
2018-05-02 10:19:06.626000	28.393555	Soil moisture	Percentage	[0, 85]
2018-05-02 10:29:06.585000	28.613281	Soil moisture	Percentage	[0, 85]
2018-05-02 10:39:04.677000	28.027344	Soil moisture	Percentage	[0, 85]
2018-05-02 10:49:05.622000	28.076172	Soil moisture	Percentage	[0, 85]
2018-05-02 10:59:07.202000	28.710938	Soil moisture	Percentage	[0, 85]
2018-05-02 11:09:07.287000	28.735352	Soil moisture	Percentage	[0, 85]
2018-05-02 11:19:06.709000	27.978516	Soil moisture	Percentage	[0, 85]

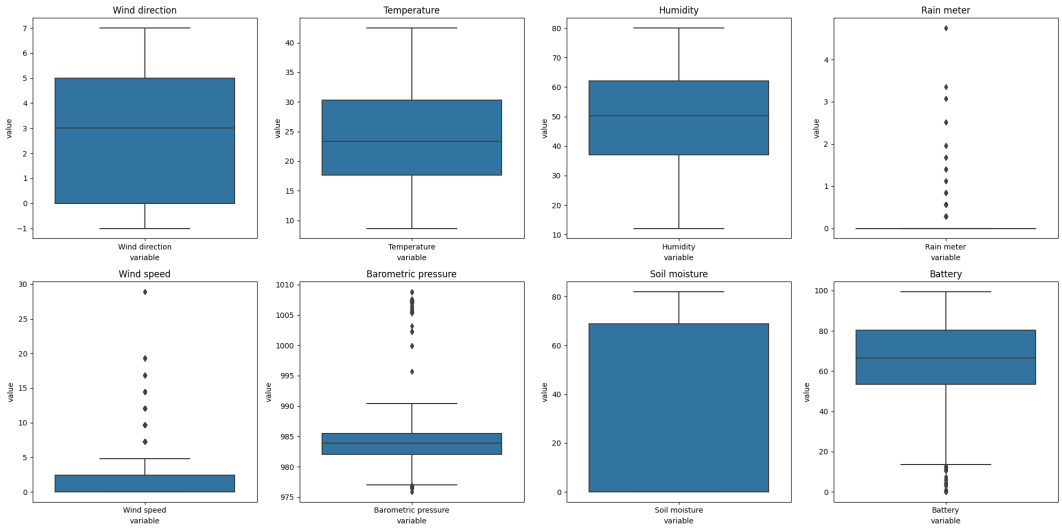
Se realiza un análisis de cada variable y se calcula el porcentaje de valores que están fuera de rango y cuantos de ellos corresponden a valores nulos.



Barometric pressure: 2.02% (null: 0.00%)
Battery: 0.00% (null: 0.00%)
Humidity: 10.13% (null: 0.00%)
Rain meter: 0.02% (null: 0.00%)
Soil moisture: 0.00% (null: 0.00%)
Temperature: 1.53% (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: 2.02% (null: 2.02%)
Battery: 0.00% (null: 0.00%)

Humidity: 10.13% (null: 10.13%)
Rain meter: 0.02% (null: 0.02%)
Soil moisture: 0.00% (null: 0.00%)
Temperature: 1.53% (null: 1.53%)
Wind direction: 0.00% (null: 0.00%)
Wind speed: 100.00% (null: 0.00%)

Variables with more than 5% null values: Humidity

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							
2018-04-25 21:05:00	nan	81.726563	nan	0.000000	31.835938	nan	0.000000
2018-05-02 10:07:00	nan	87.207031	nan	0.000000	27.050781	nan	0.000000
2018-05-02 10:08:00	nan	87.207031	nan	0.000000	28.857422	nan	0.000000
2018-05-02 10:19:00	nan	nan	nan	0.000000	28.393555	nan	0.000000
2018-05-02 10:29:00	nan	87.062500	nan	0.000000	28.613281	nan	0.000000
2018-05-02 10:39:00	nan	86.917969	nan	0.000000	28.027344	nan	0.000000
2018-05-02 10:49:00	nan	86.917969	nan	0.000000	28.076172	nan	0.000000
2018-05-02 10:59:00	nan	86.777344	nan	0.000000	28.710938	nan	0.000000
2018-05-02 11:09:00	nan	86.777344	nan	0.000000	28.735352	nan	0.000000
2018-05-02 11:19:00	nan	86.777344	nan	0.000000	27.978516	nan	0.000000

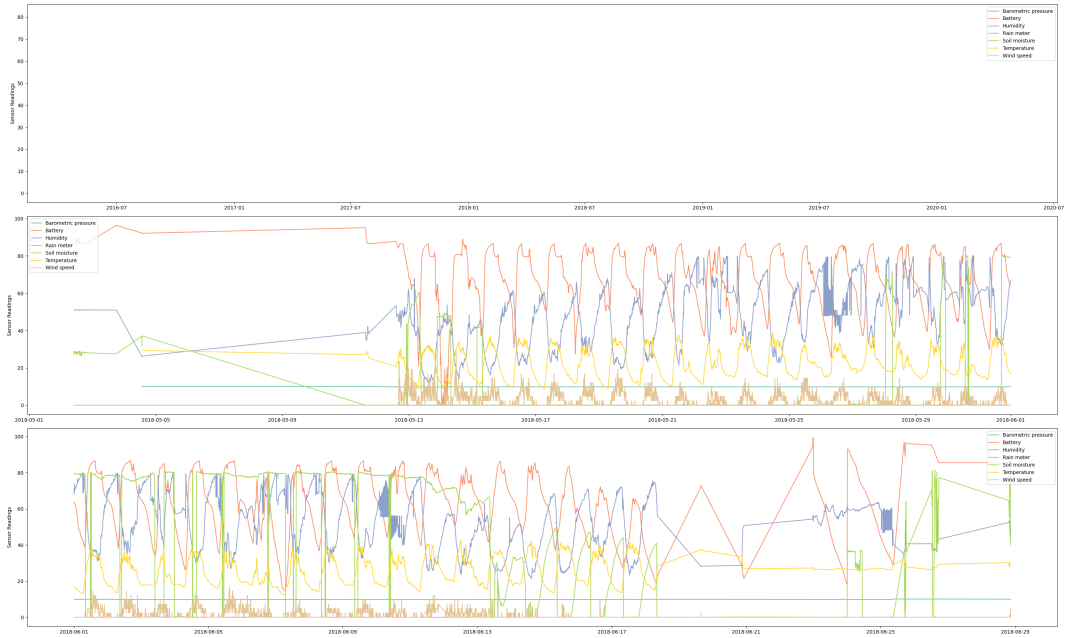
Se imputan los valores faltantes usando el método IterativeImpute. El método IterativeImputer es una técnica de imputación de valores que utiliza un regresor bayesiano como estimador y utilizando el resto de columnas para esa fila.

variable	Barometric pressure	Battery	Humidity	Rain meter	Soil moisture	Temperature	Wind speed
date							
2018-04-25 21:05:00	nan	81.726563	51.414715	0.000000	31.835938	nan	0.000000
2018-05-02 10:07:00	nan	87.207031	50.944202	0.000000	27.050781	nan	0.000000
2018-05-02 10:08:00	nan	87.207031	51.121845	0.000000	28.857422	nan	0.000000
2018-05-02 10:19:00	nan	nan	51.076234	0.000000	28.393555	nan	0.000000
2018-05-02 10:29:00	nan	87.062500	51.097839	0.000000	28.613281	nan	0.000000
2018-05-02 10:39:00	nan	86.917969	51.040225	0.000000	28.027344	nan	0.000000
2018-05-02 10:49:00	nan	86.917969	51.045027	0.000000	28.076172	nan	0.000000
2018-05-02 10:59:00	nan	86.777344	51.107442	0.000000	28.710938	nan	0.000000
2018-05-02 11:09:00	nan	86.777344	51.109842	0.000000	28.735352	nan	0.000000
2018-05-02 11:19:00	nan	86.777344	51.035424	0.000000	27.978516	nan	0.000000

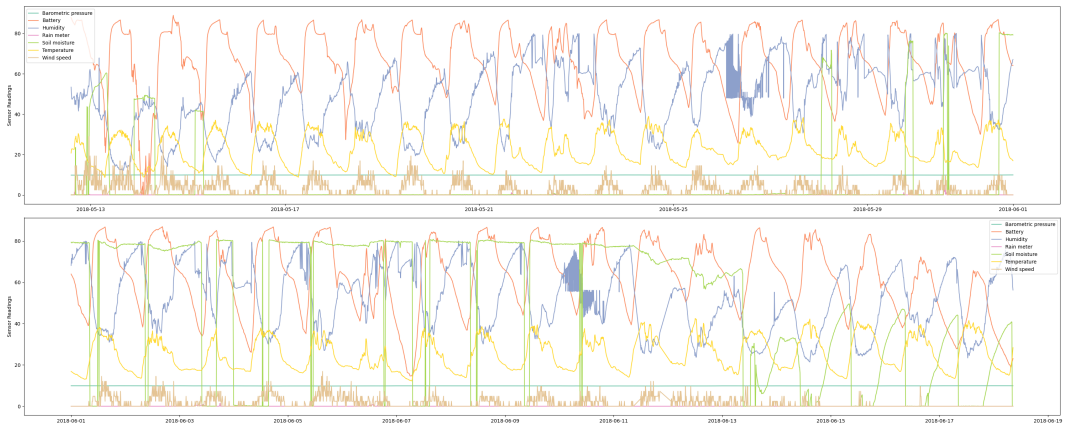
Se realizó una interpolación basada en el tiempo en el dataframe y redondearon todos los valores a 2 decimales.

variable	Barometric pressure	Battery	Humidity	Rain meter	Soil moisture	Temperature	Wind speed
date							
2018-04-25 21:05:00	nan	81.730000	51.410000	0.000000	31.840000	nan	0.000000
2018-05-02 10:07:00	nan	87.210000	50.940000	0.000000	27.050000	nan	0.000000
2018-05-02 10:08:00	nan	87.210000	51.120000	0.000000	28.860000	nan	0.000000
2018-05-02 10:19:00	nan	87.130000	51.080000	0.000000	28.390000	nan	0.000000
2018-05-02 10:29:00	nan	87.060000	51.100000	0.000000	28.610000	nan	0.000000
2018-05-02 10:39:00	nan	86.920000	51.040000	0.000000	28.030000	nan	0.000000
2018-05-02 10:49:00	nan	86.920000	51.050000	0.000000	28.080000	nan	0.000000
2018-05-02 10:59:00	nan	86.780000	51.110000	0.000000	28.710000	nan	0.000000
2018-05-02 11:09:00	nan	86.780000	51.110000	0.000000	28.740000	nan	0.000000
2018-05-02 11:19:00	nan	86.780000	51.040000	0.000000	27.980000	nan	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							
2018-05-12 14:16:00	980.320000	87.780000	53.390000	0.000000	0.020000	20.720000	0.000000
2018-05-12 14:26:00	980.120000	87.350000	49.710000	0.000000	0.070000	22.520000	0.000000
2018-05-12 14:36:00	980.090000	86.920000	47.540000	0.000000	0.050000	22.820000	0.000000
2018-05-12 14:46:00	979.940000	86.490000	47.360000	0.000000	0.050000	22.820000	0.000000
2018-05-12 14:56:00	979.820000	86.200000	47.600000	0.000000	0.070000	22.760000	0.000000
2018-05-12 15:06:00	979.710000	85.770000	47.880000	0.000000	0.020000	22.700000	0.000000
2018-05-12 15:16:00	979.530000	85.480000	48.120000	0.000000	0.070000	22.660000	0.000000
2018-05-12 15:26:00	979.510000	85.200000	48.340000	0.000000	0.020000	22.630000	0.000000
2018-05-12 15:36:00	979.300000	85.050000	48.630000	0.000000	0.120000	22.600000	0.000000
2018-05-12 15:46:00	979.110000	84.770000	48.890000	0.000000	0.000000	22.580000	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								
2018-05-12 14:16:00	980.320000	87.780000	53.390000	0.000000	0.020000	20.720000	0.000000	861
2018-05-12 14:26:00	980.120000	87.350000	49.710000	0.000000	0.070000	22.520000	0.000000	861
2018-05-12 14:36:00	980.090000	86.920000	47.540000	0.000000	0.050000	22.820000	0.000000	861
2018-05-12 14:46:00	979.940000	86.490000	47.360000	0.000000	0.050000	22.820000	0.000000	861
2018-05-12 14:56:00	979.820000	86.200000	47.600000	0.000000	0.070000	22.760000	0.000000	861
2018-05-12 15:06:00	979.710000	85.770000	47.880000	0.000000	0.020000	22.700000	0.000000	861
2018-05-12 15:16:00	979.530000	85.480000	48.120000	0.000000	0.070000	22.660000	0.000000	861
2018-05-12 15:26:00	979.510000	85.200000	48.340000	0.000000	0.020000	22.630000	0.000000	861
2018-05-12 15:36:00	979.300000	85.050000	48.630000	0.000000	0.120000	22.600000	0.000000	861
2018-05-12 15:46:00	979.110000	84.770000	48.890000	0.000000	0.000000	22.580000	0.000000	861