Sección 1: Carga de datos

Se cargaron los datos desde el archivo CSV para el sensor 4e0022000251353337353037

|	date	value	variable	units	range
0	2018-04-27 09:47:00.681	27.968281	Temperature	Centigrade	[-10, 85]
1	2018-04-27 09:47:45.005	27.979006	Temperature	Centigrade	[-10, 85]
2	2018-04-27 09:49:32.616	28.290033	Temperature	Centigrade	[-10, 85]
3	2018-04-27 09:50:34.470	28.139883	Temperature	Centigrade	[-10, 85]
4	2018-04-27 10:02:56.611	28.515261	Temperature	Centigrade	[-10, 85]
5	2018-04-27 10:08:57.609	28.236408	Temperature	Centigrade	[-10, 85]
6	2018-04-27 17:59:43.383	-46.581871	Temperature	Centigrade	[-10, 85]
7	2018-04-27 18:07:09.005	28.054083	Temperature	Centigrade	[-10, 85]
8	2018-04-27 18:17:35.163	28.096983	Temperature	Centigrade	[-10, 85]
9	2018-04-30 13:06:33.917	-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
2018-04-27 07:46:57	-1.000000	Wind direction	Direction (degrees)	[-1, 7]
2018-04-27 07:47:42	-1.000000	Wind direction	Direction (degrees)	[-1, 7]
2018-04-27 07:49:28	-1.000000	Wind direction	Direction (degrees)	[-1, 7]
2018-04-27 07:50:28	-1.000000	Wind direction	Direction (degrees)	[-1, 7]
2018-04-27 08:02:53	-1.000000	Wind direction	Direction (degrees)	[-1, 7]
2018-04-27 08:08:54	-1.000000	Wind direction	Direction (degrees)	[-1, 7]
2018-04-27 09:47:00.681000	27.968281	Temperature	Centigrade	[-10, 85]
2018-04-27 09:47:00.891000	39.478821	Humidity	Percentage	[0, 80]
2018-04-27 09:47:01.103000	0.000000	Rain meter	millilitres (mm)	[-1, 7]
2018-04-27 09:47:01.507000	0.000000	Wind speed	km/h	nan

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

date	value	variable	units	range
2018-04-27 09:47:02.157000	31.982422	Soil moisture	Percentage	[0, 85]
2018-04-27 09:49:34.087000	32.153320	Soil moisture	Percentage	[0, 85]
2018-04-27 09:50:35.943000	31.884766	Soil moisture	Percentage	[0, 85]
2018-04-27 10:02:58.068000	32.373047	Soil moisture	Percentage	[0, 85]
2018-04-27 10:08:59.085000	31.909180	Soil moisture	Percentage	[0, 85]
2018-04-27 17:59:44.864000	24.975586	Soil moisture	Percentage	[0, 85]
2018-04-27 18:07:10.465000	7.495117	Soil moisture	Percentage	[0, 85]
2018-04-27 18:17:36.638000	0.000000	Soil moisture	Percentage	[0, 85]
2018-04-30 13:06:35.371000	28.540039	Soil moisture	Percentage	[0, 85]
2018-04-30 13:07:34.201000	30.102539	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.

Number of Rows per Variable
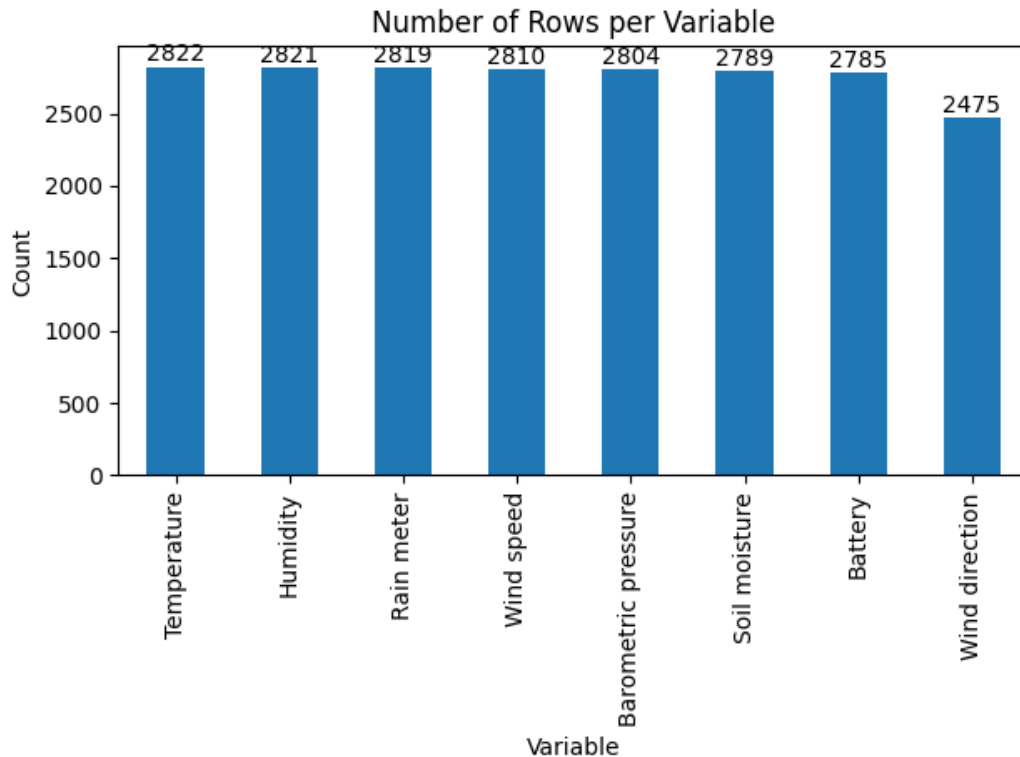
Barometric pressure: 1.89% (null: 0.00%)
Battery: 0.00% (null: 0.00%)
Humidity: 32.68% (null: 0.00%)
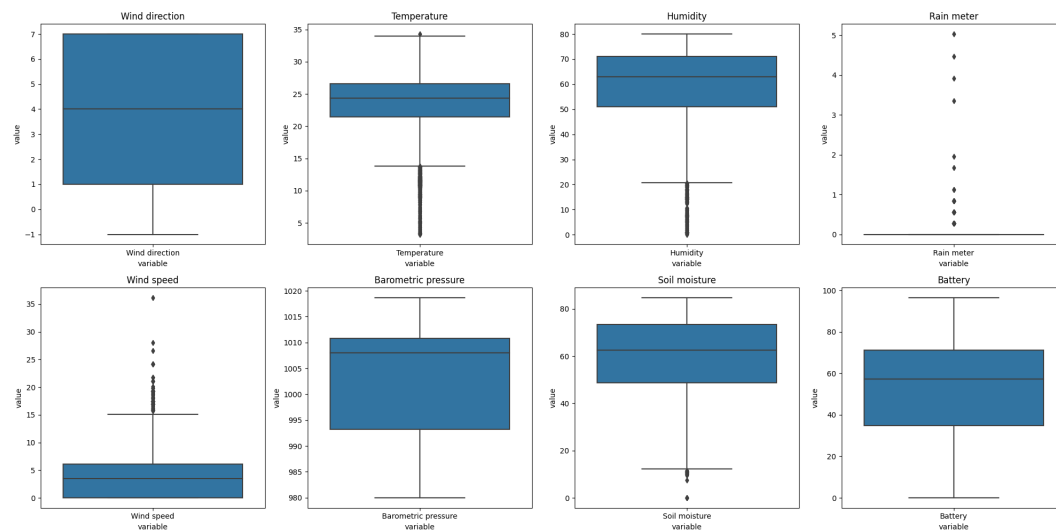Rain meter: 0.04% (null: 0.00%)
Soil moisture: 0.93% (null: 0.00%)
Temperature: 1.13% (null: 0.00%)
Wind direction: 0.00% (null: 0.00%)
Wind speed: 100.00% (null: 0.00%)

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: 1.89% (null: 1.89%)
Battery: 0.00% (null: 0.00%)

Humidity: 32.68% (null: 32.68%)
Rain meter: 0.04% (null: 0.04%)
Soil moisture: 0.93% (null: 0.93%)
Temperature: 1.13% (null: 1.13%)
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.

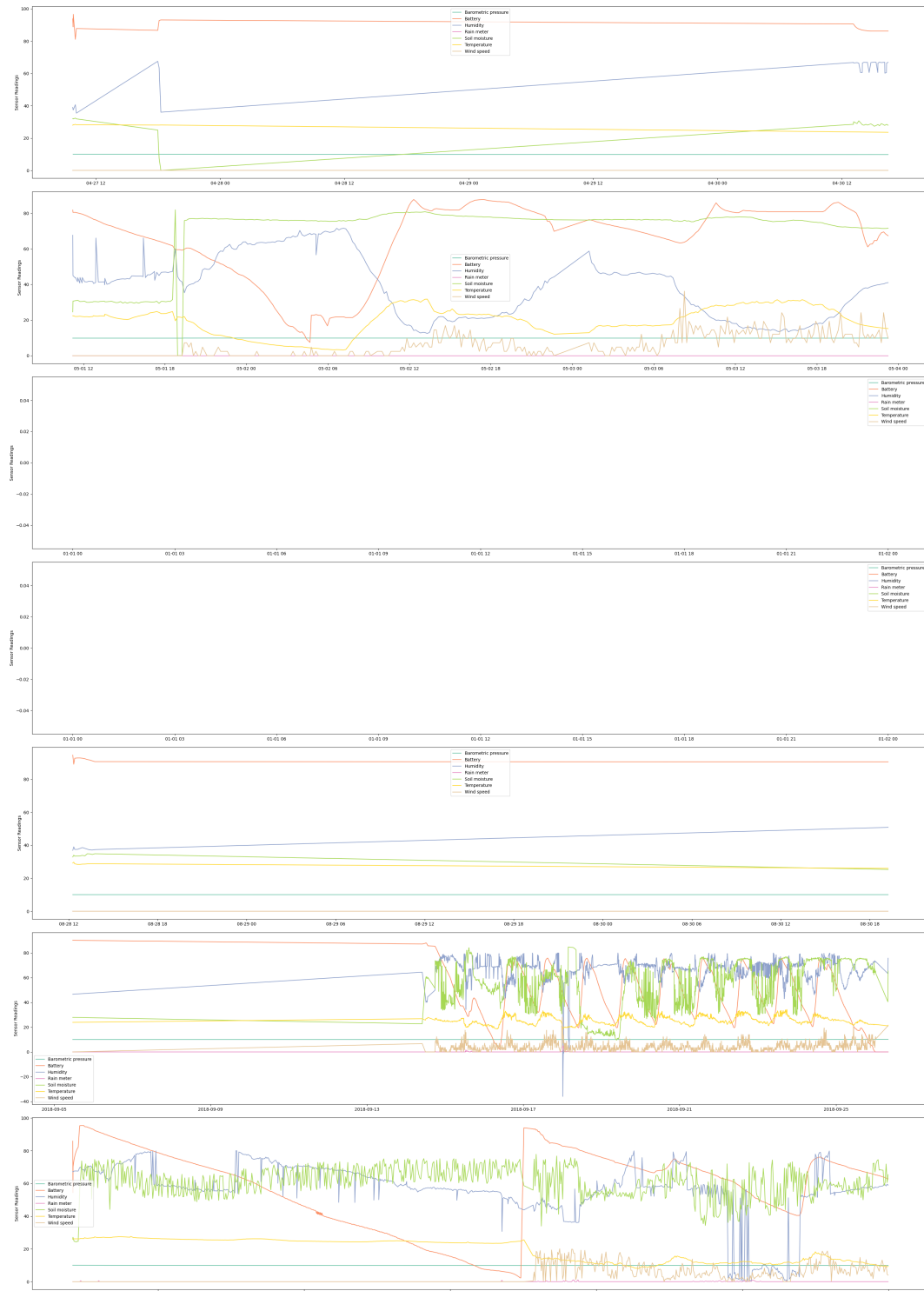| variable date	Barometric pressure	Battery	Humidity	Rain meter	Soil moisture	Temperature	Wind speed
2018-04-27 09:47:00	nan	93.703125	39.047760	0.000000	31.982422	27.973644	0.000000
2018-04-27 09:49:00	1002.565002	88.800781	37.487549	0.000000	32.153320	28.290033	0.000000
2018-04-27 09:50:00	nan	96.550781	37.457031	0.000000	31.884766	28.139883	0.000000
2018-04-27 10:02:00	nan	81.074219	40.653748	0.000000	32.373047	28.515261	0.000000
2018-04-27 10:08:00	nan	87.781250	35.435242	0.000000	31.909180	28.236408	0.000000
2018-04-27 17:59:00	nan	86.632813	nan	0.000000	24.975586	nan	0.000000
2018-04-27 18:07:00	nan	92.753906	63.450378	0.000000	7.495117	28.054083	0.000000
2018-04-27 18:17:00	1001.299988	93.070313	36.015076	0.000000	0.000000	28.096983	0.000000
2018-04-30 13:06:00	nan	90.539063	nan	0.000000	28.540039	nan	0.000000
2018-04-30 13:07:00	nan	90.382813	nan	0.000000	30.102539	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.

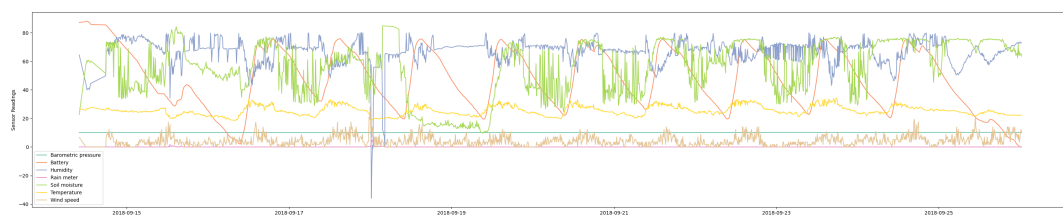| variable date	Barometric pressure	Battery	Humidity	Rain meter	Soil moisture	Temperature	Wind speed
2018-04-27 09:47:00	nan	93.703125	39.047760	0.000000	31.982422	27.973644	0.000000
2018-04-27 09:49:00	1002.565002	88.800781	37.487549	0.000000	32.153320	28.290033	0.000000
2018-04-27 09:50:00	nan	96.550781	37.457031	0.000000	31.884766	28.139883	0.000000
2018-04-27 10:02:00	nan	81.074219	40.653748	0.000000	32.373047	28.515261	0.000000
2018-04-27 10:08:00	nan	87.781250	35.435242	0.000000	31.909180	28.236408	0.000000
2018-04-27 17:59:00	nan	86.632813	67.522998	0.000000	24.975586	nan	0.000000
2018-04-27 18:07:00	nan	92.753906	63.450378	0.000000	7.495117	28.054083	0.000000
2018-04-27 18:17:00	1001.299988	93.070313	36.015076	0.000000	0.000000	28.096983	0.000000
2018-04-30 13:06:00	nan	90.539063	66.779735	0.000000	28.540039	nan	0.000000
2018-04-30 13:07:00	nan	90.382813	66.453922	0.000000	30.102539	nan	0.000000

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

variable date	Barometric pressure	Battery	Humidity	Rain meter	Soil moisture	Temperature	Wind speed
2018-04-27 09:47:00	nan	93.700000	39.050000	0.000000	31.980000	27.970000	0.000000
2018-04-27 09:49:00	1002.570000	88.800000	37.490000	0.000000	32.150000	28.290000	0.000000
2018-04-27 09:50:00	1002.560000	96.550000	37.460000	0.000000	31.880000	28.140000	0.000000
2018-04-27 10:02:00	1002.530000	81.070000	40.650000	0.000000	32.370000	28.520000	0.000000
2018-04-27 10:08:00	1002.520000	87.780000	35.440000	0.000000	31.910000	28.240000	0.000000
2018-04-27 17:59:00	1001.340000	86.630000	67.520000	0.000000	24.980000	28.060000	0.000000
2018-04-27 18:07:00	1001.320000	92.750000	63.450000	0.000000	7.500000	28.050000	0.000000
2018-04-27 18:17:00	1001.300000	93.070000	36.020000	0.000000	0.000000	28.100000	0.000000
2018-04-30 13:06:00	986.760000	90.540000	66.780000	0.000000	28.540000	23.820000	0.000000
2018-04-30 13:07:00	986.750000	90.380000	66.450000	0.000000	30.100000	23.820000	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 date	Barometric pressure	Battery	Humidity	Rain meter	Soil moisture	Temperature	Wind speed
2018-09-14 09:59:00	1009.960000	87.210000	64.470000	0.000000	22.750000	26.850000	6.720000
2018-09-14 11:58:00	1009.970000	87.780000	44.120000	0.000000	58.030000	25.690000	0.000000
2018-09-14 12:08:00	1009.970000	88.070000	41.440000	0.000000	59.200000	26.600000	0.000000
2018-09-14 12:18:00	1009.950000	88.070000	39.990000	0.000000	60.620000	26.870000	0.000000
2018-09-14 12:28:00	1009.920000	87.930000	40.700000	0.000000	59.520000	26.920000	0.000000
2018-09-14 12:38:00	1009.860000	87.780000	41.240000	0.000000	60.720000	27.080000	0.000000
2018-09-14 12:48:00	1009.830000	87.350000	40.610000	0.000000	60.740000	27.040000	0.000000
2018-09-14 12:58:00	1009.850000	87.060000	40.140000	0.000000	59.620000	26.990000	0.000000
2018-09-14 13:08:00	1009.930000	86.630000	41.620000	0.000000	59.470000	27.220000	0.000000
2018-09-14 13:18:00	1010.020000	86.200000	44.450000	0.000000	59.280000	27.640000	0.000000

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

variable date	Barometric pressure	Battery	Humidity	Rain meter	Soil moisture	Temperature	Wind speed	day_length
2018-09-14 09:59:00	1009.960000	87.210000	64.470000	0.000000	22.750000	26.850000	6.720000	753
2018-09-14 11:58:00	1009.970000	87.780000	44.120000	0.000000	58.030000	25.690000	0.000000	753
2018-09-14 12:08:00	1009.970000	88.070000	41.440000	0.000000	59.200000	26.600000	0.000000	753
2018-09-14 12:18:00	1009.950000	88.070000	39.990000	0.000000	60.620000	26.870000	0.000000	753
2018-09-14 12:28:00	1009.920000	87.930000	40.700000	0.000000	59.520000	26.920000	0.000000	753
2018-09-14 12:38:00	1009.860000	87.780000	41.240000	0.000000	60.720000	27.080000	0.000000	753
2018-09-14 12:48:00	1009.830000	87.350000	40.610000	0.000000	60.740000	27.040000	0.000000	753
2018-09-14 12:58:00	1009.850000	87.060000	40.140000	0.000000	59.620000	26.990000	0.000000	753
2018-09-14 13:08:00	1009.930000	86.630000	41.620000	0.000000	59.470000	27.220000	0.000000	753
2018-09-14 13:18:00	1010.020000	86.200000	44.450000	0.000000	59.280000	27.640000	0.000000	753