Sección 1: Carga de datos

Se cargaron los datos desde el archivo CSV para el sensor 3800330019513433343630 36

	date	value	variable	units	range
0	2018-04-13 20:36:33.568	21.286545	Temperature	Centigrade	[-10, 85]
1	2018-04-13 20:38:05.143	21.329445	Temperature	Centigrade	[-10, 85]
2	2018-04-13 20:48:29.220	20.878992	Temperature	Centigrade	[-10, 85]
3	2018-04-13 20:58:28.707	19.945908	Temperature	Centigrade	[-10, 85]
4	2018-04-13 21:08:29.311	19.699230	Temperature	Centigrade	[-10, 85]
5	2018-04-13 21:18:31.437	19.023550	Temperature	Centigrade	[-10, 85]
6	2018-04-13 21:28:28.325	18.851948	Temperature	Centigrade	[-10, 85]
7	2018-04-13 21:38:28.651	18.841223	Temperature	Centigrade	[-10, 85]
8	2018-04-13 21:47:49.515	19.484730	Temperature	Centigrade	[-10, 85]
9	2018-04-13 21:58:12.386	19.098625	Temperature	Centigrade	[-10, 85]

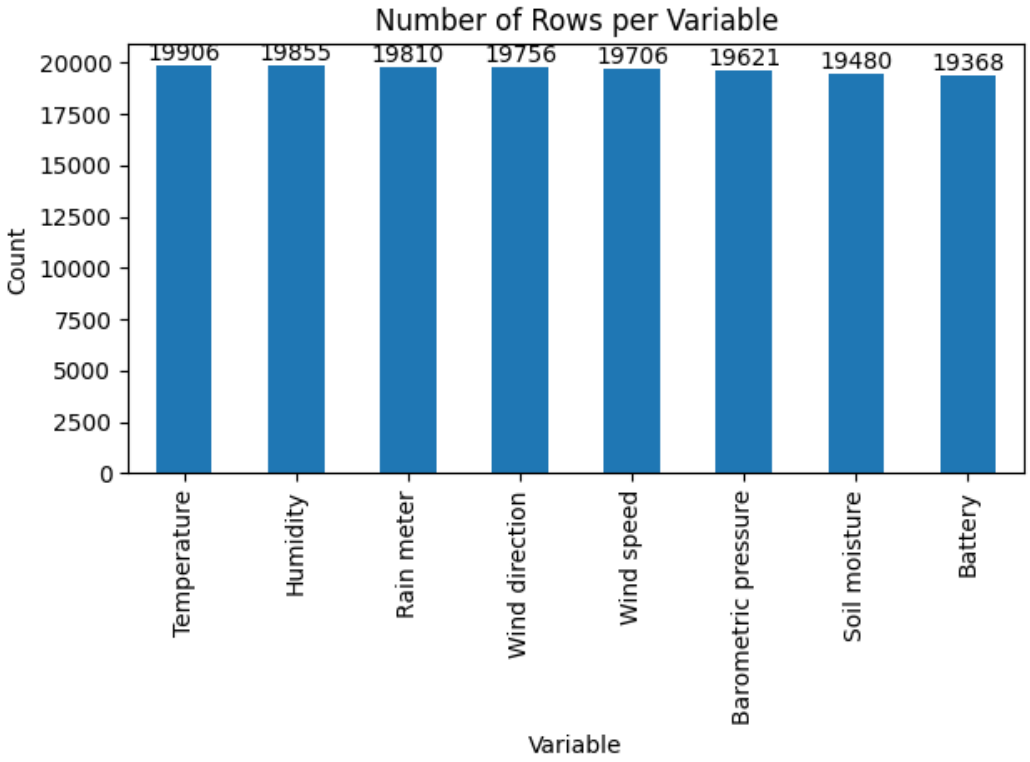
Se conviertien las marcas de tiempo Unix a formato de fecha y hora y se establece como índice. Además se ordena el dataframe por fecha.

	value	variable	units	range
date				
2018-04-13 18:36:06	-1.000000	Wind direction	Direction (degrees)	[-1, 7]
2018-04-13 18:38:02	-1.000000	Wind direction	Direction (degrees)	[-1, 7]
2018-04-13 18:48:02	-1.000000	Wind direction	Direction (degrees)	[-1, 7]
2018-04-13 18:58:02	-1.000000	Wind direction	Direction (degrees)	[-1, 7]
2018-04-13 19:08:02	-1.000000	Wind direction	Direction (degrees)	[-1, 7]
2018-04-13 19:18:02	-1.000000	Wind direction	Direction (degrees)	[-1, 7]
2018-04-13 19:28:02	-1.000000	Wind direction	Direction (degrees)	[-1, 7]
2018-04-13 19:38:02	-1.000000	Wind direction	Direction (degrees)	[-1, 7]
2018-04-13 19:47:47	-1.000000	Wind direction	Direction (degrees)	[-1, 7]
2018-04-13 19:57:46	-1.000000	Wind direction	Direction (degrees)	[-1, 7]

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

	value	variable	units	range
date				
2018-04-13 20:36:35.031000	31.738281	Soil moisture	Percentage	[0, 85]
2018-04-13 20:38:06.623000	32.055664	Soil moisture	Percentage	[0, 85]
2018-04-13 20:48:30.684000	31.665039	Soil moisture	Percentage	[0, 85]
2018-04-13 20:58:30.192000	30.224609	Soil moisture	Percentage	[0, 85]
2018-04-13 21:08:30.777000	30.883789	Soil moisture	Percentage	[0, 85]
2018-04-13 21:18:32.616000	29.589844	Soil moisture	Percentage	[0, 85]
2018-04-13 21:28:29.871000	29.150391	Soil moisture	Percentage	[0, 85]
2018-04-13 21:38:30.183000	30.517578	Soil moisture	Percentage	[0, 85]
2018-04-13 21:47:50.979000	29.931641	Soil moisture	Percentage	[0, 85]
2018-04-13 21:58:13.846000	29.663086	Soil moisture	Percentage	[0, 85]

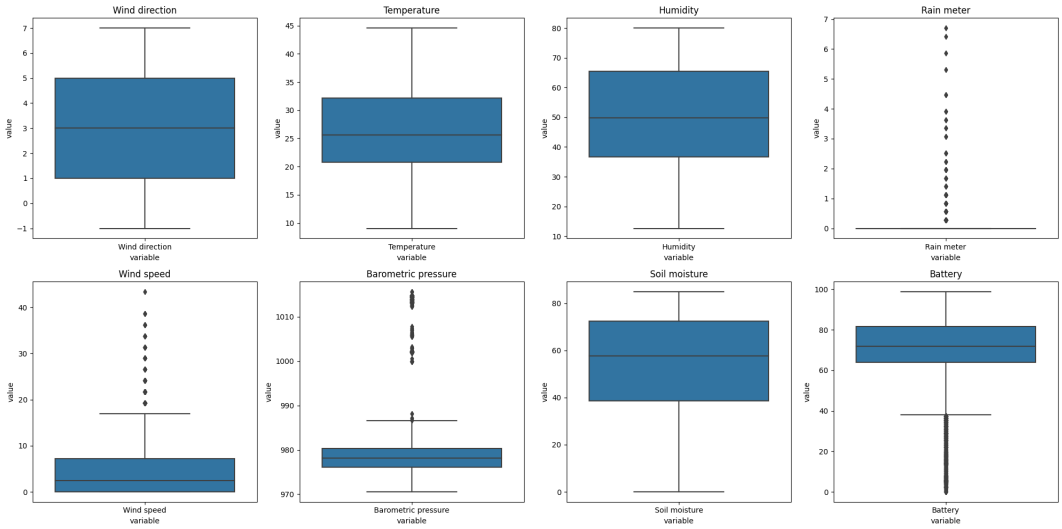
Se realiza un análisis de cada variable y se calcula el porcentaje de valores que están fuera de rango y cuantos de ellos corresponden a valores nulos.



Barometric pressure: 0.31% (null: 0.00%)
Battery: 0.00% (null: 0.00%)
Humidity: 10.32% (null: 0.00%)
Rain meter: 0.02% (null: 0.00%)
Soil moisture: 30.00% (null: 0.00%)
Temperature: 0.10% (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: 0.31% (null: 0.31%)
Battery: 0.00% (null: 0.00%)

Humidity: 10.32% (null: 10.32%)
Rain meter: 0.02% (null: 0.02%)
Soil moisture: 30.00% (null: 30.00%)
Temperature: 0.10% (null: 0.10%)
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							
2018-04-13 20:36:00	977.619995	91.488281	40.318054	0.000000	31.738281	21.286545	0.000000
2018-04-13 20:38:00	nan	77.484375	41.080994	0.000000	32.055664	21.329445	0.000000
2018-04-13 20:48:00	977.812500	81.726563	41.554016	0.000000	31.665039	20.878992	0.000000
2018-04-13 20:58:00	978.000000	82.542969	46.047729	0.000000	30.224609	19.945908	0.000000
2018-04-13 21:08:00	978.292480	81.890625	45.666260	0.000000	30.883789	19.699230	0.000000
2018-04-13 21:18:00	978.679993	80.910156	47.970337	0.000000	29.589844	19.023550	0.000000
2018-04-13 21:28:00	978.825012	79.933594	48.153442	0.000000	29.150391	18.851948	0.000000
2018-04-13 21:38:00	978.902527	79.117188	48.496765	0.000000	30.517578	18.841223	0.000000
2018-04-13 21:47:00	nan	75.855469	50.121826	0.000000	29.931641	19.484730	0.000000
2018-04-13 21:58:00	979.362488	76.019531	49.442810	0.000000	29.663086	19.098625	0.000000

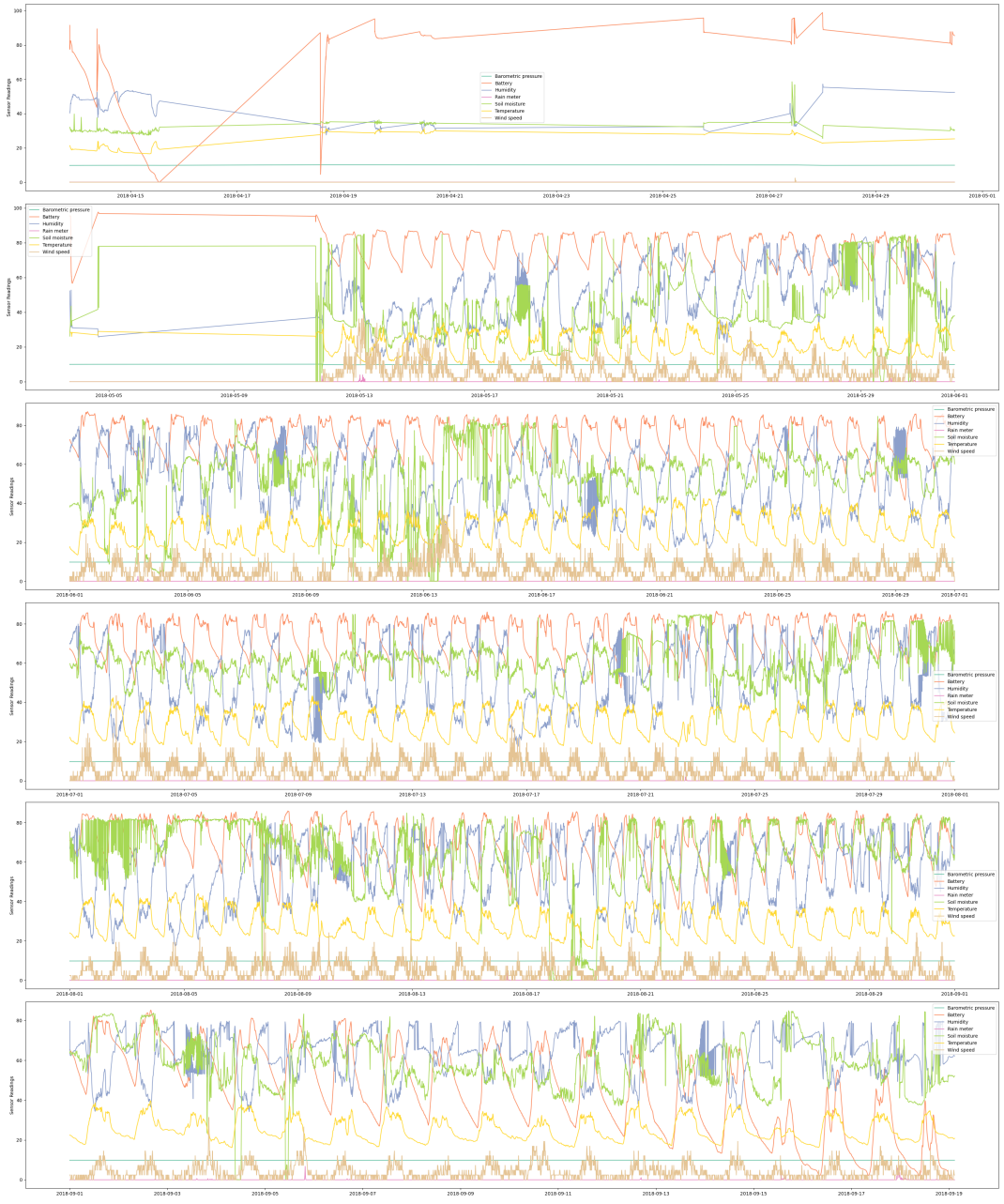
Se imputan los valores faltantes usando el método IterativeImpute. El método IterativeImputer es una técnica de imputación de valores que utiliza un regresor bayesiano como estimador y utilizando el resto de columnas para esa fila.

variable	Barometric pressure	Battery	Humidity	Rain meter	Soil moisture	Temperature	Wind speed
date							
2018-04-13 20:36:00	977.619995	91.488281	40.318054	0.000000	31.738281	21.286545	0.000000
2018-04-13 20:38:00	nan	77.484375	41.080994	0.000000	32.055664	21.329445	0.000000
2018-04-13 20:48:00	977.812500	81.726563	41.554016	0.000000	31.665039	20.878992	0.000000
2018-04-13 20:58:00	978.000000	82.542969	46.047729	0.000000	30.224609	19.945908	0.000000
2018-04-13 21:08:00	978.292480	81.890625	45.666260	0.000000	30.883789	19.699230	0.000000
2018-04-13 21:18:00	978.679993	80.910156	47.970337	0.000000	29.589844	19.023550	0.000000
2018-04-13 21:28:00	978.825012	79.933594	48.153442	0.000000	29.150391	18.851948	0.000000
2018-04-13 21:38:00	978.902527	79.117188	48.496765	0.000000	30.517578	18.841223	0.000000
2018-04-13 21:47:00	nan	75.855469	50.121826	0.000000	29.931641	19.484730	0.000000
2018-04-13 21:58:00	979.362488	76.019531	49.442810	0.000000	29.663086	19.098625	0.000000

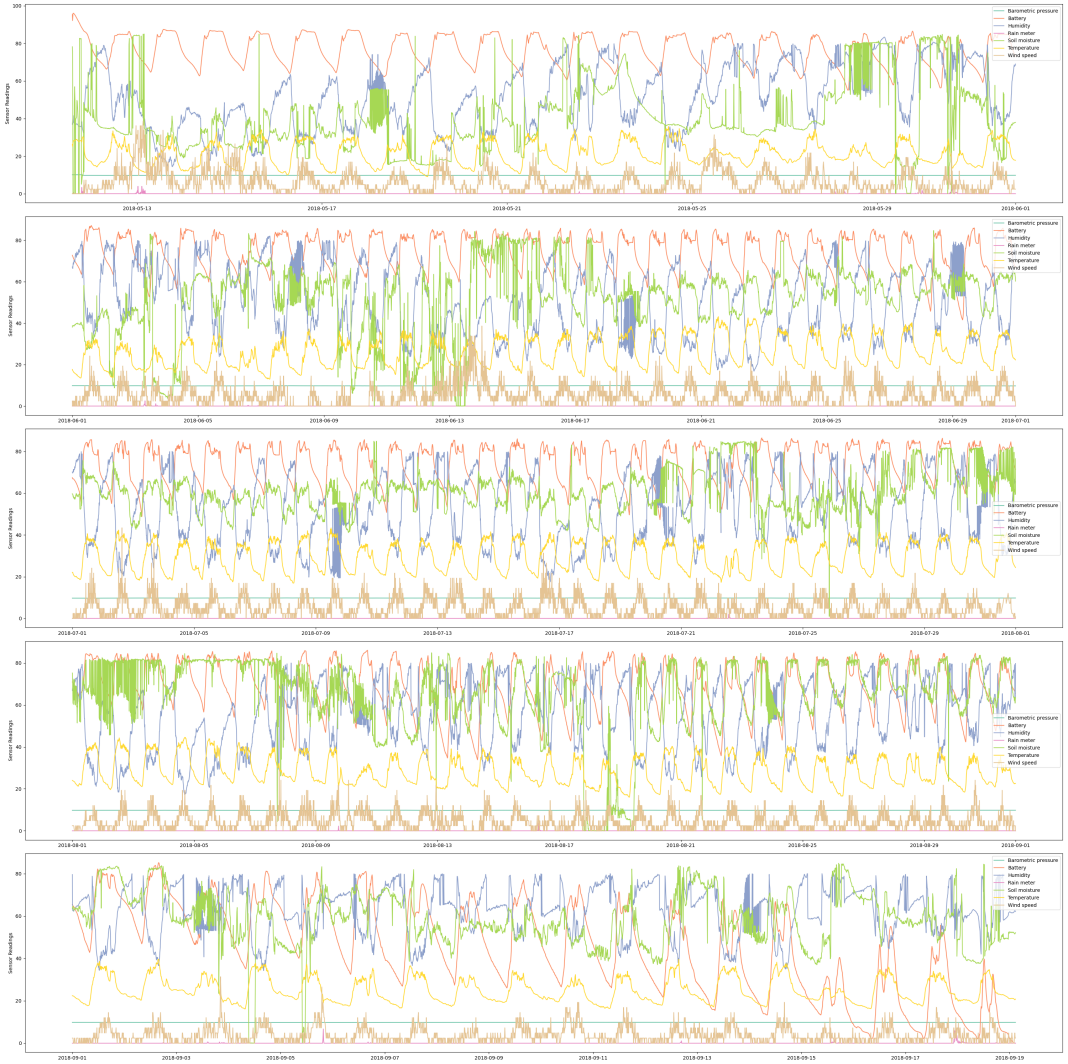
Se realizó una interpolación basada en el tiempo en el dataframe y redondearon todos los valores a 2 decimales.

variable	Barometric pressure	Battery	Humidity	Rain meter	Soil moisture	Temperature	Wind speed
date							
2018-04-13 20:36:00	977.620000	91.490000	40.320000	0.000000	31.740000	21.290000	0.000000
2018-04-13 20:38:00	977.650000	77.480000	41.080000	0.000000	32.060000	21.330000	0.000000
2018-04-13 20:48:00	977.810000	81.730000	41.550000	0.000000	31.670000	20.880000	0.000000
2018-04-13 20:58:00	978.000000	82.540000	46.050000	0.000000	30.220000	19.950000	0.000000
2018-04-13 21:08:00	978.290000	81.890000	45.670000	0.000000	30.880000	19.700000	0.000000
2018-04-13 21:18:00	978.680000	80.910000	47.970000	0.000000	29.590000	19.020000	0.000000
2018-04-13 21:28:00	978.830000	79.930000	48.150000	0.000000	29.150000	18.850000	0.000000
2018-04-13 21:38:00	978.900000	79.120000	48.500000	0.000000	30.520000	18.840000	0.000000
2018-04-13 21:47:00	979.110000	75.860000	50.120000	0.000000	29.930000	19.480000	0.000000
2018-04-13 21:58:00	979.360000	76.020000	49.440000	0.000000	29.660000	19.100000	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
2018-05-11 14:28:00	1007.480000	95.290000	36.950000	0.000000	78.200000	26.220000	0.000000
2018-05-11 14:38:00	1007.480000	92.120000	38.240000	0.000000	75.880000	25.140000	0.000000
2018-05-11 14:49:00	1007.270000	95.920000	39.940000	0.000000	0.020000	26.790000	0.000000
2018-05-11 14:52:00	1007.220000	95.920000	41.590000	0.000000	0.020000	27.400000	0.000000
2018-05-11 15:02:00	1007.030000	96.080000	38.600000	0.560000	0.020000	28.150000	0.000000
2018-05-11 15:12:00	1007.090000	95.920000	37.800000	0.000000	0.020000	28.250000	0.000000
2018-05-11 15:22:00	1006.930000	95.760000	37.530000	0.000000	0.020000	28.260000	0.000000
2018-05-11 15:32:00	1006.800000	95.520000	37.280000	0.000000	48.670000	28.270000	0.000000
2018-05-11 15:42:00	1006.670000	95.290000	50.020000	0.000000	56.930000	28.270000	0.000000
2018-05-11 15:52:00	1006.540000	95.050000	36.880000	0.000000	48.450000	28.290000	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								
2018-05-11 14:28:00	1007.480000	95.290000	36.950000	0.000000	78.200000	26.220000	0.000000	859
2018-05-11 14:38:00	1007.480000	92.120000	38.240000	0.000000	75.880000	25.140000	0.000000	859
2018-05-11 14:49:00	1007.270000	95.920000	39.940000	0.000000	0.020000	26.790000	0.000000	859
2018-05-11 14:52:00	1007.220000	95.920000	41.590000	0.000000	0.020000	27.400000	0.000000	859
2018-05-11 15:02:00	1007.030000	96.080000	38.600000	0.560000	0.020000	28.150000	0.000000	859
2018-05-11 15:12:00	1007.090000	95.920000	37.800000	0.000000	0.020000	28.250000	0.000000	859
2018-05-11 15:22:00	1006.930000	95.760000	37.530000	0.000000	0.020000	28.260000	0.000000	859
2018-05-11 15:32:00	1006.800000	95.520000	37.280000	0.000000	48.670000	28.270000	0.000000	859
2018-05-11 15:42:00	1006.670000	95.290000	50.020000	0.000000	56.930000	28.270000	0.000000	859
2018-05-11 15:52:00	1006.540000	95.050000	36.880000	0.000000	48.450000	28.290000	0.000000	859