

Sección 1: Carga de datos

Se cargaron los datos desde el archivo CSV para el sensor 200034001951343334363036

	date	value	variable	units	range
0	2018-04-01 00:04:04.340	14.325957	Temperature	Centigrade	[-10, 85]
1	2018-04-01 00:14:03.711	14.325957	Temperature	Centigrade	[-10, 85]
2	2018-04-01 00:24:04.259	14.304507	Temperature	Centigrade	[-10, 85]
3	2018-04-01 00:34:04.284	14.079280	Temperature	Centigrade	[-10, 85]
4	2018-04-01 00:44:04.253	14.047105	Temperature	Centigrade	[-10, 85]
5	2018-04-01 00:54:04.954	14.240156	Temperature	Centigrade	[-10, 85]
6	2018-04-01 01:04:03.666	14.411757	Temperature	Centigrade	[-10, 85]
7	2018-04-01 01:14:03.357	14.293781	Temperature	Centigrade	[-10, 85]
8	2018-04-01 01:24:02.288	14.218706	Temperature	Centigrade	[-10, 85]
9	2018-04-01 01:34:04.481	14.250881	Temperature	Centigrade	[-10, 85]

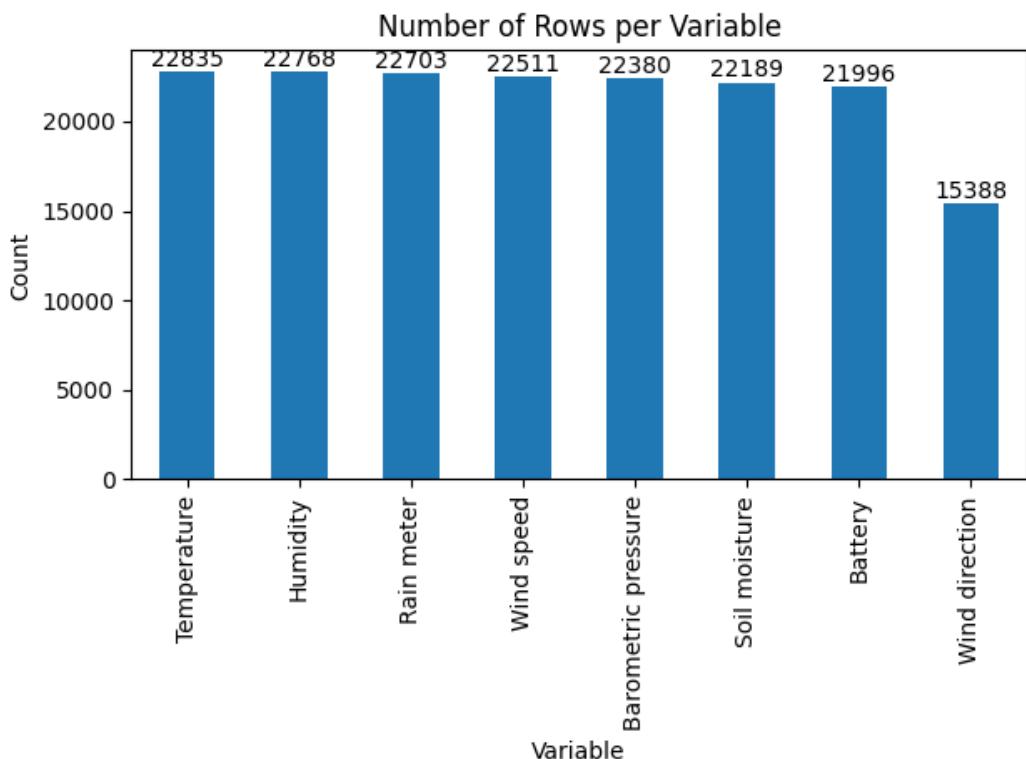
Se convierten las marcas de tiempo Unix a formato de fecha y hora y se establece como índice. Además se ordena el dataframe por fecha.

	value	variable	units	range
date				
2018-03-31 22:04:47	0.000000	Wind direction	Direction (degrees)	[-1, 7]
2018-03-31 22:14:47	-1.000000	Wind direction	Direction (degrees)	[-1, 7]
2018-03-31 22:24:47	-1.000000	Wind direction	Direction (degrees)	[-1, 7]
2018-03-31 22:34:47	4.000000	Wind direction	Direction (degrees)	[-1, 7]
2018-03-31 22:44:47	0.000000	Wind direction	Direction (degrees)	[-1, 7]
2018-03-31 22:54:47	7.000000	Wind direction	Direction (degrees)	[-1, 7]
2018-03-31 23:04:47	7.000000	Wind direction	Direction (degrees)	[-1, 7]
2018-03-31 23:14:47	7.000000	Wind direction	Direction (degrees)	[-1, 7]
2018-03-31 23:24:47	7.000000	Wind direction	Direction (degrees)	[-1, 7]
2018-03-31 23:34:47	7.000000	Wind direction	Direction (degrees)	[-1, 7]

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

	value	variable	units	range
date				
2018-04-01 00:04:05.805000	24.169922	Soil moisture	Percentage	[0, 85]
2018-04-01 00:14:05.181000	23.437500	Soil moisture	Percentage	[0, 85]
2018-04-01 00:24:05.729000	23.120117	Soil moisture	Percentage	[0, 85]
2018-04-01 00:34:05.770000	24.584961	Soil moisture	Percentage	[0, 85]
2018-04-01 00:44:05.723000	23.583984	Soil moisture	Percentage	[0, 85]
2018-04-01 00:54:06.419000	23.217773	Soil moisture	Percentage	[0, 85]
2018-04-01 01:04:05.135000	23.071289	Soil moisture	Percentage	[0, 85]
2018-04-01 01:14:04.828000	23.364258	Soil moisture	Percentage	[0, 85]
2018-04-01 01:24:03.742000	23.486328	Soil moisture	Percentage	[0, 85]
2018-04-01 01:34:05.951000	23.583984	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: 0.06% (null: 0.00%)

Battery: 0.00% (null: 0.00%)

Humidity: 15.51% (null: 0.00%)

Rain meter: 0.00% (null: 0.00%)

Soil moisture: 0.00% (null: 0.00%)

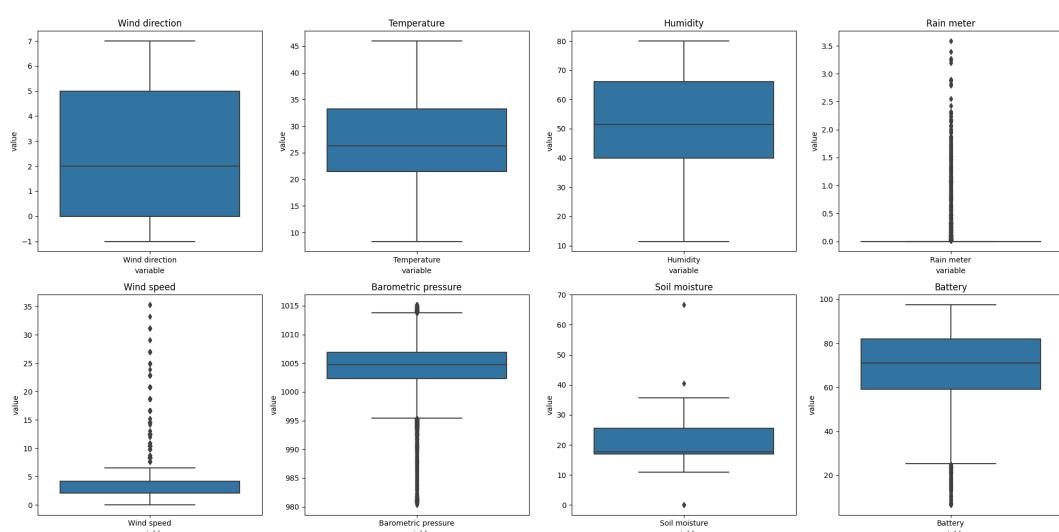
Temperature: 0.00% (null: 0.00%)

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: 0.06% (null: 0.06%)

Battery: 0.00% (null: 0.00%)

Humidity: 15.51% (null: 15.51%)
 Rain meter: 0.00% (null: 0.00%)
 Soil moisture: 0.00% (null: 0.00%)
 Temperature: 0.00% (null: 0.00%)
 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 vacíos 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							
2018-04-01 00:04:00	1006.227478	78.789063	23.777527	0.000000	24.169922	14.325957	4.155989
2018-04-01 00:14:00	1006.289978	78.300781	23.869080	0.000000	23.437500	14.325957	2.077994
2018-04-01 00:24:00	1006.539978	77.812500	23.899597	0.000000	23.120117	14.304507	2.077994
2018-04-01 00:34:00	1006.710022	77.484375	24.158997	0.000000	24.584961	14.079280	4.155989
2018-04-01 00:44:00	1006.807495	76.996094	24.326843	0.000000	23.583984	14.047105	6.233984
2018-04-01 00:54:00	1006.747498	76.671875	24.372620	0.000000	23.217773	14.240156	4.155989
2018-04-01 01:04:00	1006.724976	76.343750	24.349731	0.000000	23.071289	14.411757	6.233984
2018-04-01 01:14:00	1006.837524	76.019531	24.364990	0.000000	23.364258	14.293781	4.155989
2018-04-01 01:24:00	1006.809998	75.691406	24.868530	0.000000	23.486328	14.218706	2.077994
2018-04-01 01:34:00	1006.882507	75.367188	25.318665	0.000000	23.583984	14.250881	4.155989

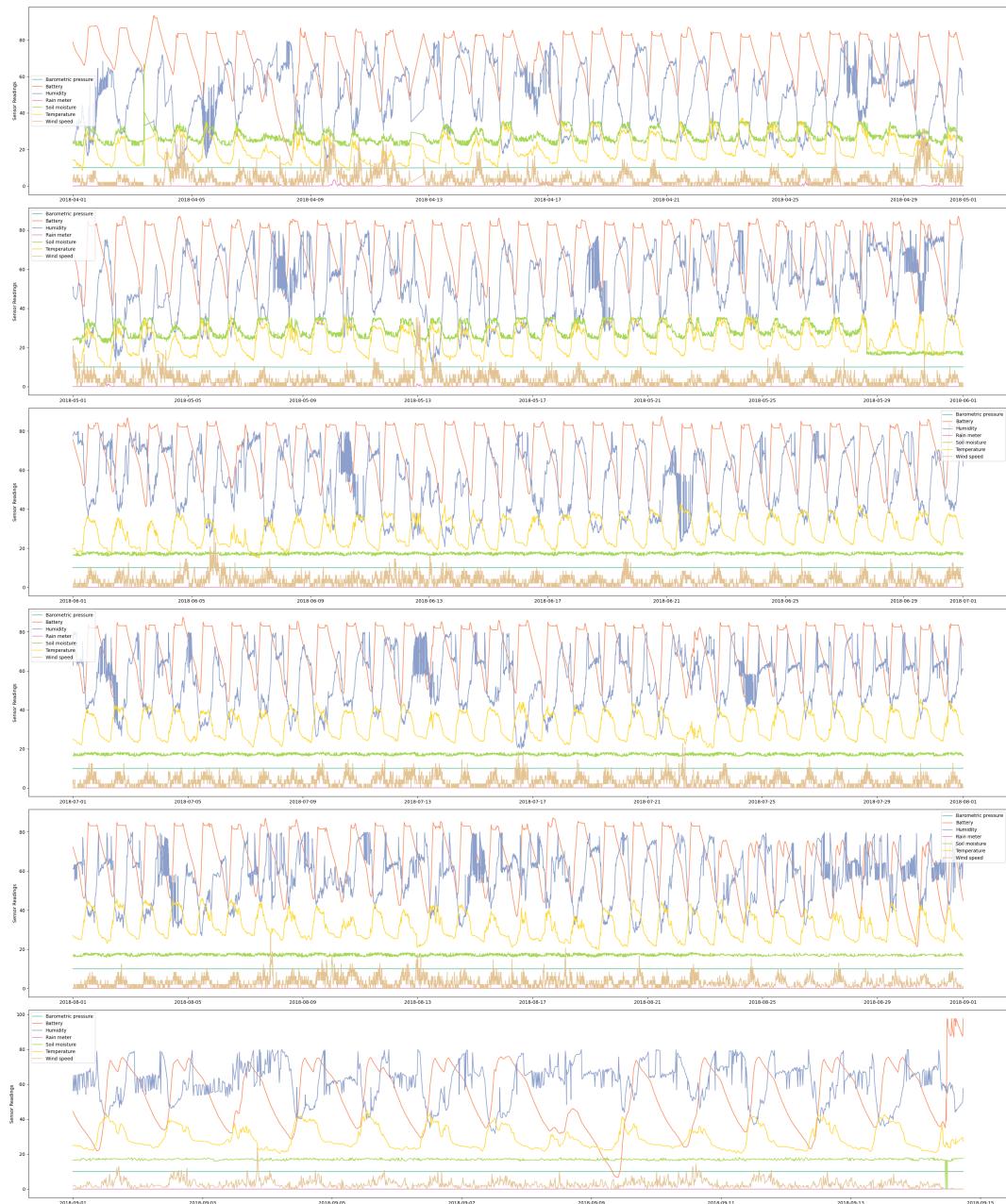
Se imputan los valores faltantes usando el método IterativeImpute. El método IterativeImputer es una técnica de imputación de valores que utiliza un regresor bayesiano como estimador y utilizando el resto de columnas para esa fila.

variable	Barometric pressure	Battery	Humidity	Rain meter	Soil moisture	Temperature	Wind speed
date							
2018-04-01 00:04:00	1006.227478	78.789063	23.777527	0.000000	24.169922	14.325957	4.155989
2018-04-01 00:14:00	1006.289978	78.300781	23.869080	0.000000	23.437500	14.325957	2.077994
2018-04-01 00:24:00	1006.539978	77.812500	23.899597	0.000000	23.120117	14.304507	2.077994
2018-04-01 00:34:00	1006.710022	77.484375	24.158997	0.000000	24.584961	14.079280	4.155989
2018-04-01 00:44:00	1006.807495	76.996094	24.326843	0.000000	23.583984	14.047105	6.233984
2018-04-01 00:54:00	1006.747498	76.671875	24.372620	0.000000	23.217773	14.240156	4.155989
2018-04-01 01:04:00	1006.724976	76.343750	24.349731	0.000000	23.071289	14.411757	6.233984
2018-04-01 01:14:00	1006.837524	76.019531	24.364990	0.000000	23.364258	14.293781	4.155989
2018-04-01 01:24:00	1006.809998	75.691406	24.868530	0.000000	23.486328	14.218706	2.077994
2018-04-01 01:34:00	1006.882507	75.367188	25.318665	0.000000	23.583984	14.250881	4.155989

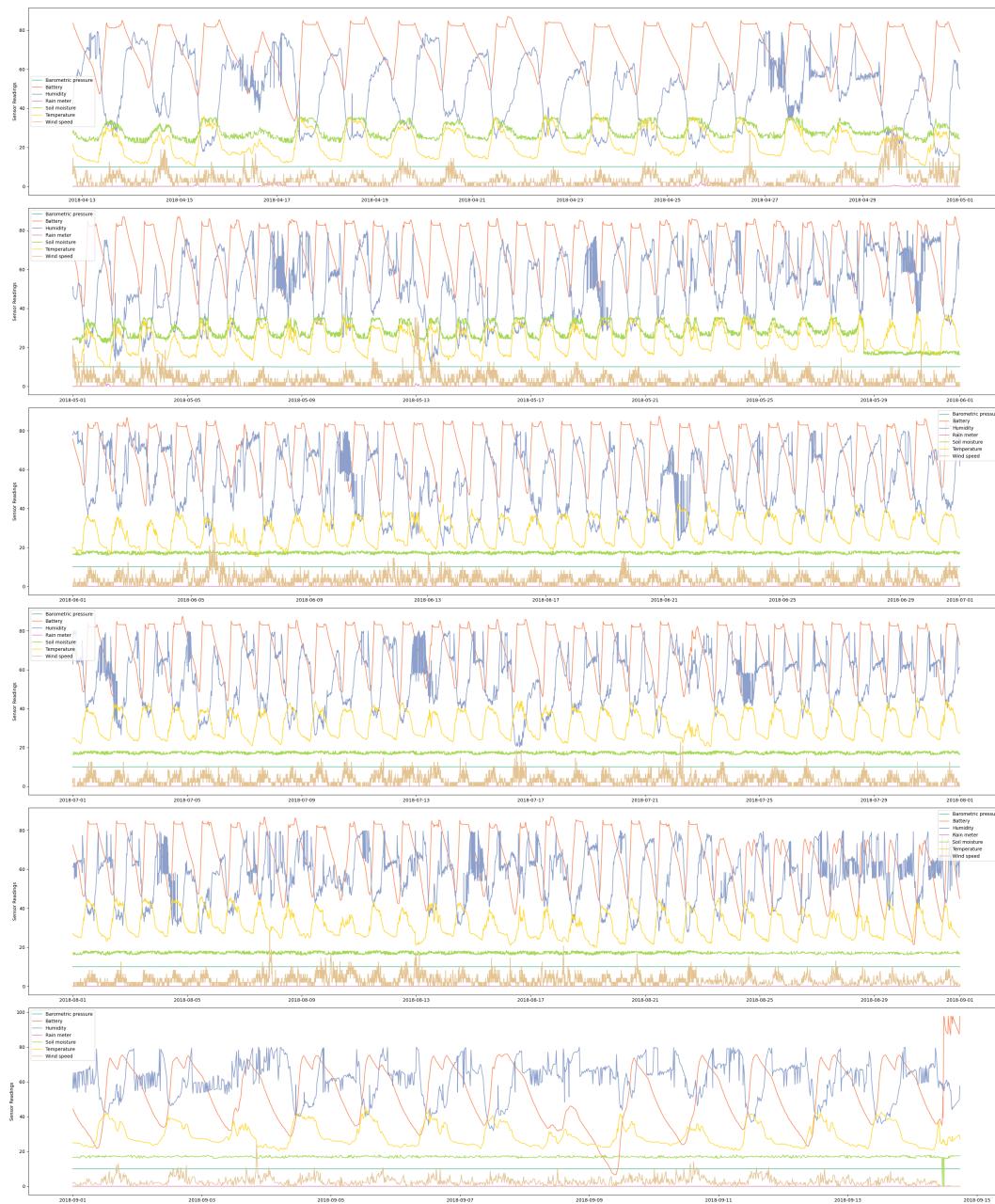
Se realizó una interpolación basada en el tiempo en el dataframe y redondearon todos los valores a 2 decimales.

variable	Barometric pressure	Battery	Humidity	Rain meter	Soil moisture	Temperature	Wind speed
date							
2018-04-01 00:04:00	1006.230000	78.790000	23.780000	0.000000	24.170000	14.330000	4.160000
2018-04-01 00:14:00	1006.290000	78.300000	23.870000	0.000000	23.440000	14.330000	2.080000
2018-04-01 00:24:00	1006.540000	77.810000	23.900000	0.000000	23.120000	14.300000	2.080000
2018-04-01 00:34:00	1006.710000	77.480000	24.160000	0.000000	24.580000	14.080000	4.160000
2018-04-01 00:44:00	1006.810000	77.000000	24.330000	0.000000	23.580000	14.050000	6.230000
2018-04-01 00:54:00	1006.750000	76.670000	24.370000	0.000000	23.220000	14.240000	4.160000
2018-04-01 01:04:00	1006.720000	76.340000	24.350000	0.000000	23.070000	14.410000	6.230000
2018-04-01 01:14:00	1006.840000	76.020000	24.360000	0.000000	23.360000	14.290000	4.160000
2018-04-01 01:24:00	1006.810000	75.690000	24.870000	0.000000	23.490000	14.220000	2.080000
2018-04-01 01:34:00	1006.880000	75.370000	25.320000	0.000000	23.580000	14.250000	4.160000

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



Lecturas de los sensores después de eliminar los períodos 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							
2018-04-12 19:27:00	990.350000	83.520000	42.260000	0.000000	28.050000	21.600000	6.230000
2018-04-12 19:37:00	990.450000	83.030000	43.930000	0.000000	27.250000	20.590000	14.550000
2018-04-12 19:47:00	990.440000	83.030000	46.390000	0.000000	28.710000	20.290000	6.230000
2018-04-12 19:57:00	990.330000	82.380000	49.450000	0.000000	28.150000	19.470000	8.310000
2018-04-12 20:07:00	990.280000	81.890000	51.660000	0.000000	27.880000	18.800000	8.310000
2018-04-12 20:17:00	990.430000	81.240000	52.930000	0.000000	26.510000	18.210000	10.390000
2018-04-12 20:27:00	990.500000	80.750000	52.510000	0.000000	26.510000	17.930000	8.310000
2018-04-12 20:37:00	990.770000	80.260000	53.750000	0.000000	26.070000	17.590000	4.160000
2018-04-12 20:47:00	991.060000	79.930000	54.830000	0.000000	27.000000	17.260000	8.310000
2018-04-12 20:57:00	991.230000	79.440000	55.540000	0.000000	27.690000	17.060000	4.160000

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								
2018-04-12 19:27:00	990.350000	83.520000	42.260000	0.000000	28.050000	21.600000	6.230000	791
2018-04-12 19:37:00	990.450000	83.030000	43.930000	0.000000	27.250000	20.590000	14.550000	791
2018-04-12 19:47:00	990.440000	83.030000	46.390000	0.000000	28.710000	20.290000	6.230000	791
2018-04-12 19:57:00	990.330000	82.380000	49.450000	0.000000	28.150000	19.470000	8.310000	791
2018-04-12 20:07:00	990.280000	81.890000	51.660000	0.000000	27.880000	18.800000	8.310000	791
2018-04-12 20:17:00	990.430000	81.240000	52.930000	0.000000	26.510000	18.210000	10.390000	791
2018-04-12 20:27:00	990.500000	80.750000	52.510000	0.000000	26.510000	17.930000	8.310000	791
2018-04-12 20:37:00	990.770000	80.260000	53.750000	0.000000	26.070000	17.590000	4.160000	791
2018-04-12 20:47:00	991.060000	79.930000	54.830000	0.000000	27.000000	17.260000	8.310000	791
2018-04-12 20:57:00	991.230000	79.440000	55.540000	0.000000	27.690000	17.060000	4.160000	791