Sección 1: Carga de datos

Se cargaron los datos desde el archivo CSV para el sensor 46005a000351353337353037

	date	value	variable	units	range
0	2018-08-23 13:03:08.888	31.378862	Temperature	Centigrade	[-10, 85]
1	2018-08-23 13:07:10.293	31.078560	Temperature	Centigrade	[-10, 85]
2	2018-08-23 13:17:35.764	30.102575	Temperature	Centigrade	[-10, 85]
3	2018-08-23 13:37:38.648	29.512695	Temperature	Centigrade	[-10, 85]
4	2018-08-23 15:05:15.335	28.654688	Temperature	Centigrade	[-10, 85]
5	2018-08-23 15:15:38.389	30.939133	Temperature	Centigrade	[-10, 85]
6	2018-08-23 15:25:38.101	30.585205	Temperature	Centigrade	[-10, 85]
7	2018-08-23 15:35:39.231	28.761938	Temperature	Centigrade	[-10, 85]
8	2018-08-23 15:45:38.869	29.941700	Temperature	Centigrade	[-10, 85]
9	2018-08-23 15:55:38.493	31.164360	Temperature	Centigrade	[-10, 85]

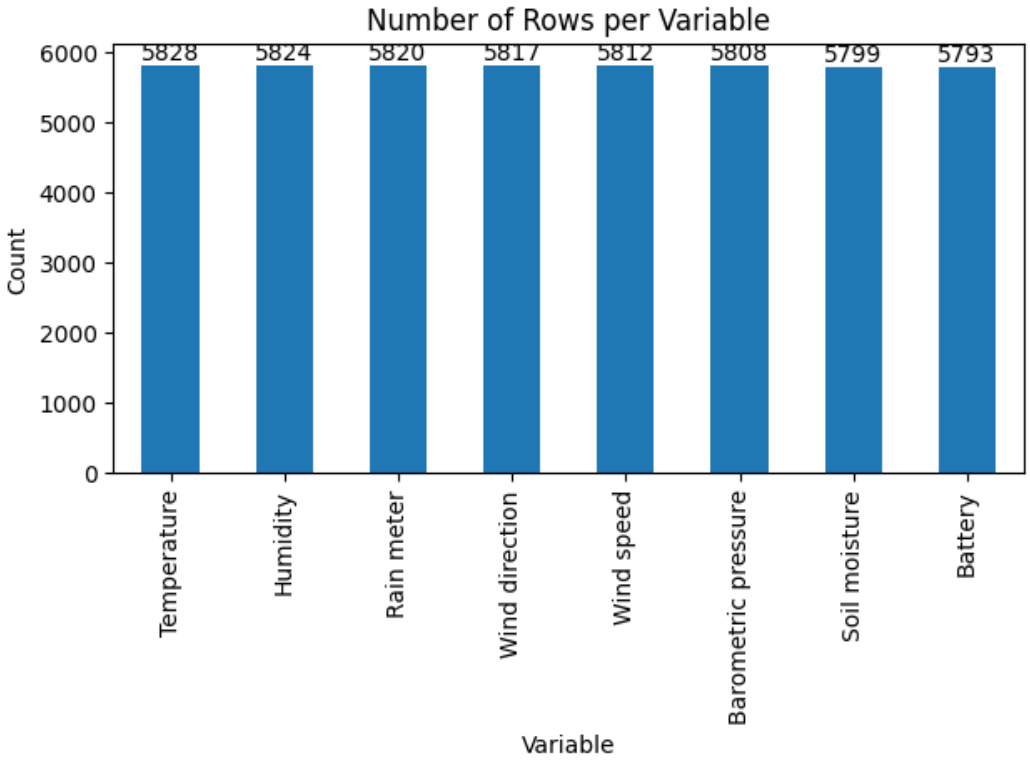
Se conviertien las marcas de tiempo Unix a formato de fecha y hora y se establece como índice. Además se ordena el dataframe por fecha.

date	value	variable	units	range
2018-08-23 11:03:05	-1.000000	Wind direction	Direction (degrees)	[-1, 7]
2018-08-23 11:07:07	-1.000000	Wind direction	Direction (degrees)	[-1, 7]
2018-08-23 11:17:07	-1.000000	Wind direction	Direction (degrees)	[-1, 7]
2018-08-23 11:37:07	-1.000000	Wind direction	Direction (degrees)	[-1, 7]
2018-08-23 13:03:08.888000	31.378862	Temperature	Centigrade	[-10, 85]
2018-08-23 13:03:09.096000	54.516357	Humidity	Percentage	[0, 80]
2018-08-23 13:03:09.308000	0.000000	Rain meter	millilitres (mm)	[-1, 7]
2018-08-23 13:05:12	0.000000	Wind direction	Direction (degrees)	[-1, 7]
2018-08-23 13:07:10.293000	31.078560	Temperature	Centigrade	[-10, 85]
2018-08-23 13:07:10.503000	54.302734	Humidity	Percentage	[0, 80]

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

date	value	variable	units	range
2018-08-23 13:07:11.783000	94.799805	Soil moisture	Percentage	[0, 85]
2018-08-23 13:17:37.242000	7.421875	Soil moisture	Percentage	[0, 85]
2018-08-23 13:37:40.126000	76.098633	Soil moisture	Percentage	[0, 85]
2018-08-23 15:05:16.805000	78.588867	Soil moisture	Percentage	[0, 85]
2018-08-23 15:15:39.859000	80.200195	Soil moisture	Percentage	[0, 85]
2018-08-23 15:25:39.571000	80.590820	Soil moisture	Percentage	[0, 85]
2018-08-23 15:35:40.761000	80.273438	Soil moisture	Percentage	[0, 85]
2018-08-23 15:45:40.334000	80.517578	Soil moisture	Percentage	[0, 85]
2018-08-23 15:55:39.974000	82.080078	Soil moisture	Percentage	[0, 85]
2018-08-23 16:05:39.978000	82.031250	Soil moisture	Percentage	[0, 85]

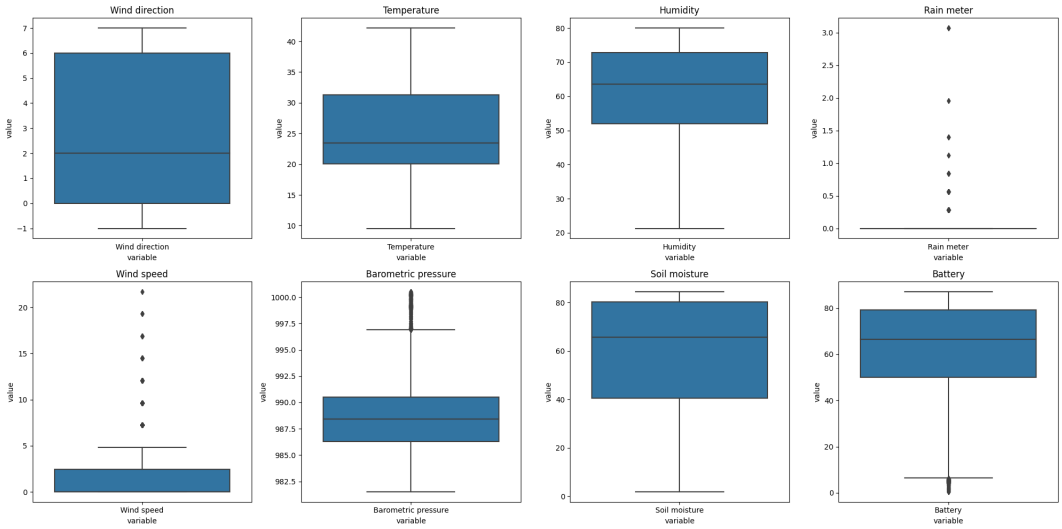
Se realiza un análisis de cada variable y se calcula el porcentaje de valores que están fuera de rango y cuantos de ellos corresponden a valores nulos.



Barometric pressure: 0.02% (null: 0.00%)
Battery: 0.00% (null: 0.00%)
Humidity: 40.92% (null: 0.00%)
Rain meter: 0.00% (null: 0.00%)
Soil moisture: 0.31% (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: 0.02% (null: 0.02%)
Battery: 0.00% (null: 0.00%)

Humidity: 40.92% (null: 40.92%)
Rain meter: 0.00% (null: 0.00%)
Soil moisture: 0.31% (null: 0.31%)
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 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							
2018-08-23 13:03:00	nan	nan	54.516357	0.000000	nan	31.378862	nan
2018-08-23 13:07:00	984.222473	86.203125	54.302734	0.000000	nan	31.078560	0.000000
2018-08-23 13:17:00	984.257507	85.769531	57.072205	0.000000	7.421875	30.102575	0.000000
2018-08-23 13:37:00	984.377502	82.703125	58.575195	0.000000	76.098633	29.512695	0.000000
2018-08-23 15:05:00	nan	83.031250	69.553894	0.000000	78.588867	28.654688	2.412060
2018-08-23 15:15:00	983.794983	82.703125	62.313599	3.073400	80.200195	30.939133	0.000000
2018-08-23 15:25:00	983.719971	82.542969	60.963196	0.000000	80.590820	30.585205	0.000000
2018-08-23 15:35:00	983.729980	82.050781	57.766479	0.000000	80.273438	28.761938	0.000000
2018-08-23 15:45:00	986.539978	82.378906	58.743042	1.955800	80.517578	29.941700	0.000000
2018-08-23 15:55:00	985.567505	83.355469	56.789917	0.000000	82.080078	31.164360	4.824121

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							
2018-08-23 13:03:00	nan	nan	54.516357	0.000000	nan	31.378862	nan
2018-08-23 13:07:00	984.222473	86.203125	54.302734	0.000000	nan	31.078560	0.000000
2018-08-23 13:17:00	984.257507	85.769531	57.072205	0.000000	7.421875	30.102575	0.000000
2018-08-23 13:37:00	984.377502	82.703125	58.575195	0.000000	76.098633	29.512695	0.000000
2018-08-23 15:05:00	nan	83.031250	69.553894	0.000000	78.588867	28.654688	2.412060
2018-08-23 15:15:00	983.794983	82.703125	62.313599	3.073400	80.200195	30.939133	0.000000
2018-08-23 15:25:00	983.719971	82.542969	60.963196	0.000000	80.590820	30.585205	0.000000
2018-08-23 15:35:00	983.729980	82.050781	57.766479	0.000000	80.273438	28.761938	0.000000
2018-08-23 15:45:00	986.539978	82.378906	58.743042	1.955800	80.517578	29.941700	0.000000
2018-08-23 15:55:00	985.567505	83.355469	56.789917	0.000000	82.080078	31.164360	4.824121

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							
2018-08-23 13:03:00	nan	nan	54.520000	0.000000	nan	31.380000	nan
2018-08-23 13:07:00	984.220000	86.200000	54.300000	0.000000	nan	31.080000	0.000000
2018-08-23 13:17:00	984.260000	85.770000	57.070000	0.000000	7.420000	30.100000	0.000000
2018-08-23 13:37:00	984.380000	82.700000	58.580000	0.000000	76.100000	29.510000	0.000000
2018-08-23 15:05:00	983.850000	83.030000	69.550000	0.000000	78.590000	28.650000	2.410000
2018-08-23 15:15:00	983.790000	82.700000	62.310000	3.070000	80.200000	30.940000	0.000000
2018-08-23 15:25:00	983.720000	82.540000	60.960000	0.000000	80.590000	30.590000	0.000000
2018-08-23 15:35:00	983.730000	82.050000	57.770000	0.000000	80.270000	28.760000	0.000000
2018-08-23 15:45:00	986.540000	82.380000	58.740000	1.960000	80.520000	29.940000	0.000000
2018-08-23 15:55:00	985.570000	83.360000	56.790000	0.000000	82.080000	31.160000	4.820000

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							
2018-08-23 13:17:00	984.260000	85.770000	57.070000	0.000000	7.420000	30.100000	0.000000
2018-08-23 13:37:00	984.380000	82.700000	58.580000	0.000000	76.100000	29.510000	0.000000
2018-08-23 15:05:00	983.850000	83.030000	69.550000	0.000000	78.590000	28.650000	2.410000
2018-08-23 15:15:00	983.790000	82.700000	62.310000	3.070000	80.200000	30.940000	0.000000
2018-08-23 15:25:00	983.720000	82.540000	60.960000	0.000000	80.590000	30.590000	0.000000
2018-08-23 15:35:00	983.730000	82.050000	57.770000	0.000000	80.270000	28.760000	0.000000
2018-08-23 15:45:00	986.540000	82.380000	58.740000	1.960000	80.520000	29.940000	0.000000
2018-08-23 15:55:00	985.570000	83.360000	56.790000	0.000000	82.080000	31.160000	4.820000
2018-08-23 16:05:00	985.820000	84.340000	55.310000	0.000000	82.030000	32.190000	2.410000
2018-08-23 16:15:00	985.710000	84.340000	54.160000	0.000000	82.130000	33.620000	2.410000

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								
2018-08-23 13:17:00	984.260000	85.770000	57.070000	0.000000	7.420000	30.100000	0.000000	809
2018-08-23 13:37:00	984.380000	82.700000	58.580000	0.000000	76.100000	29.510000	0.000000	809
2018-08-23 15:05:00	983.850000	83.030000	69.550000	0.000000	78.590000	28.650000	2.410000	809
2018-08-23 15:15:00	983.790000	82.700000	62.310000	3.070000	80.200000	30.940000	0.000000	809
2018-08-23 15:25:00	983.720000	82.540000	60.960000	0.000000	80.590000	30.590000	0.000000	809
2018-08-23 15:35:00	983.730000	82.050000	57.770000	0.000000	80.270000	28.760000	0.000000	809
2018-08-23 15:45:00	986.540000	82.380000	58.740000	1.960000	80.520000	29.940000	0.000000	809
2018-08-23 15:55:00	985.570000	83.360000	56.790000	0.000000	82.080000	31.160000	4.820000	809
2018-08-23 16:05:00	985.820000	84.340000	55.310000	0.000000	82.030000	32.190000	2.410000	809
2018-08-23 16:15:00	985.710000	84.340000	54.160000	0.000000	82.130000	33.620000	2.410000	809