

## Sección 1: Carga de datos

Se cargaron los datos desde el archivo CSV para el sensor 270043001951343334363036

	date	value	variable	units	range
0	2018-04-05 20:55:42.958	25.941238	Temperature	Centigrade	[-10, 85]
1	2018-04-05 20:56:55.833	22.509207	Temperature	Centigrade	[-10, 85]
2	2018-04-05 20:58:34.591	21.329445	Temperature	Centigrade	[-10, 85]
3	2018-04-05 21:08:04.606	20.578690	Temperature	Centigrade	[-10, 85]
4	2018-04-05 21:15:01.448	22.712984	Temperature	Centigrade	[-10, 85]
5	2018-04-05 21:36:44.882	21.769175	Temperature	Centigrade	[-10, 85]
6	2018-04-05 21:37:17.045	21.715549	Temperature	Centigrade	[-10, 85]
7	2018-04-05 21:37:47.598	21.651199	Temperature	Centigrade	[-10, 85]
8	2018-04-05 21:38:17.547	21.640474	Temperature	Centigrade	[-10, 85]
9	2018-04-05 21:38:47.869	21.597572	Temperature	Centigrade	[-10, 85]

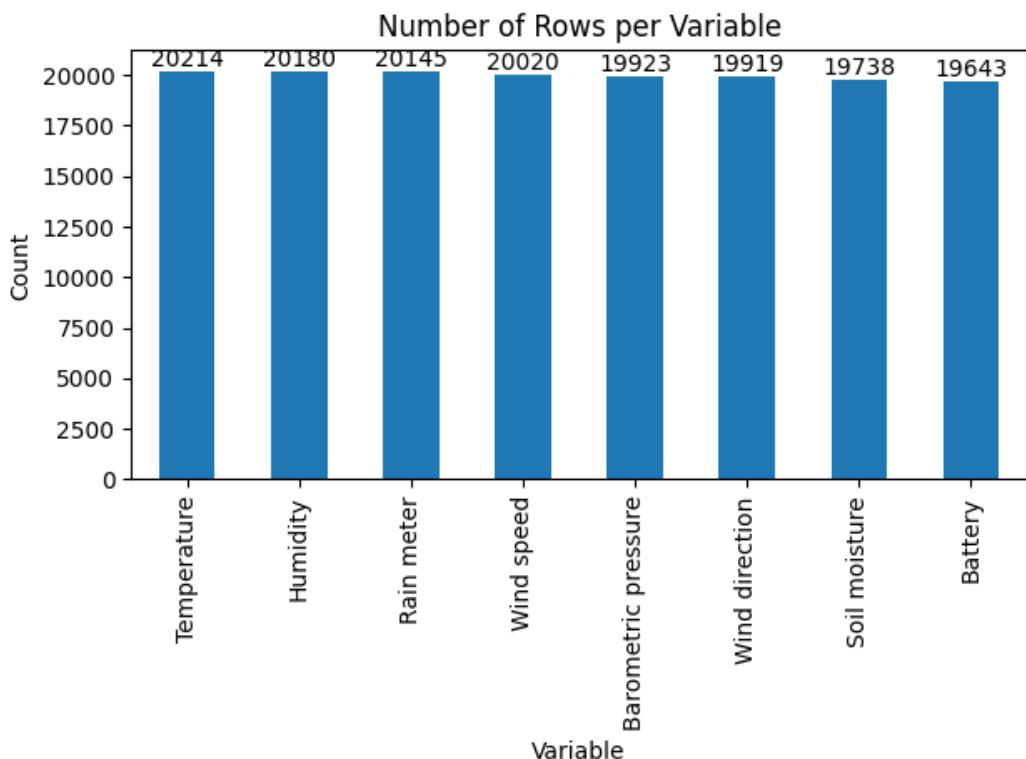
Se convierten las marcas de tiempo Unix a formato de fecha y hora y se establece como índice. Además se ordena el dataframe por fecha.

	value	variable	units	range
date				
2018-04-05 20:55:42.958000	25.941238	Temperature	Centigrade	[-10, 85]
2018-04-05 20:55:43.617000	0.000000	Rain meter	millilitres (mm)	[-1, 7]
2018-04-05 20:55:43.617000	38.952393	Humidity	Percentage	[0, 80]
2018-04-05 20:55:44.476000	0.000000	Barometric pressure	Hectopascal	[500, 1100]
2018-04-05 20:55:44.476000	0.000000	Wind speed	km/h	nan
2018-04-05 20:56:55.833000	22.509207	Temperature	Centigrade	[-10, 85]
2018-04-05 20:56:56.027000	35.084290	Humidity	Percentage	[0, 80]
2018-04-05 20:56:56.255000	0.000000	Rain meter	millilitres (mm)	[-1, 7]
2018-04-05 20:56:56.675000	0.000000	Wind speed	km/h	nan
2018-04-05 20:56:56.887000	0.000000	Barometric pressure	Hectopascal	[500, 1100]

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

	value	variable	units	range
date				
2018-04-05 20:56:57.293000	31.494141	Soil moisture	Percentage	[0, 85]
2018-04-05 20:58:36.067000	34.667969	Soil moisture	Percentage	[0, 85]
2018-04-05 21:08:06.090000	32.373047	Soil moisture	Percentage	[0, 85]
2018-04-05 21:15:02.912000	29.516602	Soil moisture	Percentage	[0, 85]
2018-04-05 21:37:18.530000	32.739258	Soil moisture	Percentage	[0, 85]
2018-04-05 21:37:49.088000	35.400391	Soil moisture	Percentage	[0, 85]
2018-04-05 21:38:19.021000	36.401367	Soil moisture	Percentage	[0, 85]
2018-04-05 21:38:49.339000	34.643555	Soil moisture	Percentage	[0, 85]
2018-04-05 21:39:20.764000	34.863281	Soil moisture	Percentage	[0, 85]
2018-04-05 21:39:51.321000	33.178711	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: 0.08% (null: 0.00%)

Battery: 0.01% (null: 0.00%)

Humidity: 9.49% (null: 0.00%)

Rain meter: 0.00% (null: 0.00%)

Soil moisture: 0.01% (null: 0.00%)

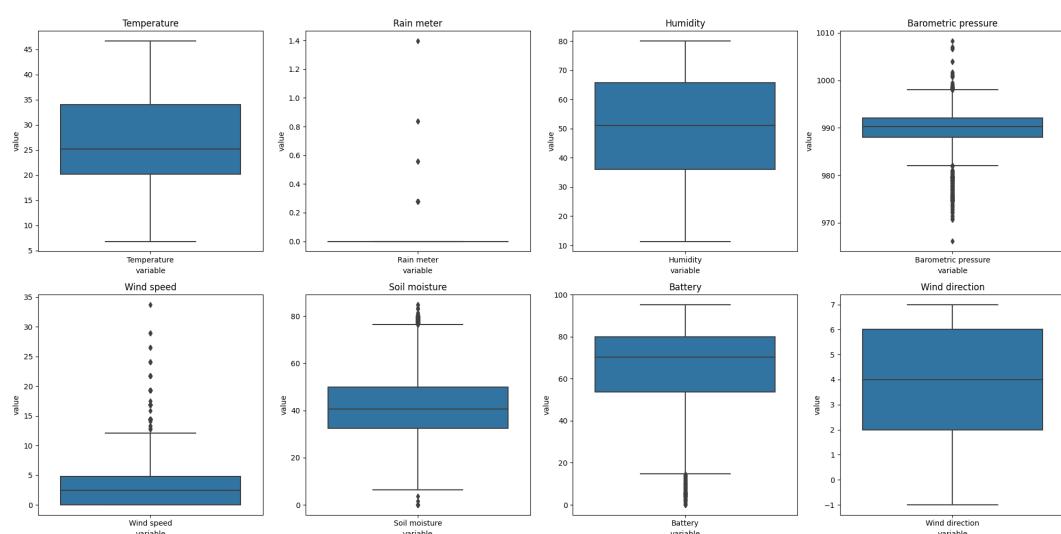
Temperature: 0.01% (null: 0.00%)

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: 0.08% (null: 0.08%)

Battery: 0.01% (null: 0.01%)

Humidity: 9.49% (null: 9.49%)  
 Rain meter: 0.00% (null: 0.00%)  
 Soil moisture: 0.01% (null: 0.01%)  
 Temperature: 0.01% (null: 0.01%)  
 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							
2018-04-05 20:55:00	nan	nan	38.952393	0.000000	nan	25.941238	0.000000
2018-04-05 20:56:00	nan	nan	35.084290	0.000000	31.494141	22.509207	0.000000
2018-04-05 20:58:00	nan	91.488281	37.579102	0.000000	34.667969	21.329445	0.000000
2018-04-05 21:08:00	nan	91.488281	39.272827	0.000000	32.373047	20.578690	0.000000
2018-04-05 21:15:00	nan	91.488281	58.025879	0.000000	29.516602	22.712984	0.000000
2018-04-05 21:36:00	nan	nan	37.319702	0.000000	nan	21.769175	0.000000
2018-04-05 21:37:00	986.728729	92.437500	37.224335	0.000000	34.069824	21.683374	0.000000
2018-04-05 21:38:00	986.751251	92.201172	37.312072	0.000000	35.522461	21.619023	0.000000
2018-04-05 21:39:00	986.753753	91.884766	37.571472	0.000000	34.020996	21.613661	0.000000
2018-04-05 21:40:00	986.746246	91.726563	37.506623	0.000000	34.008789	21.608298	0.000000

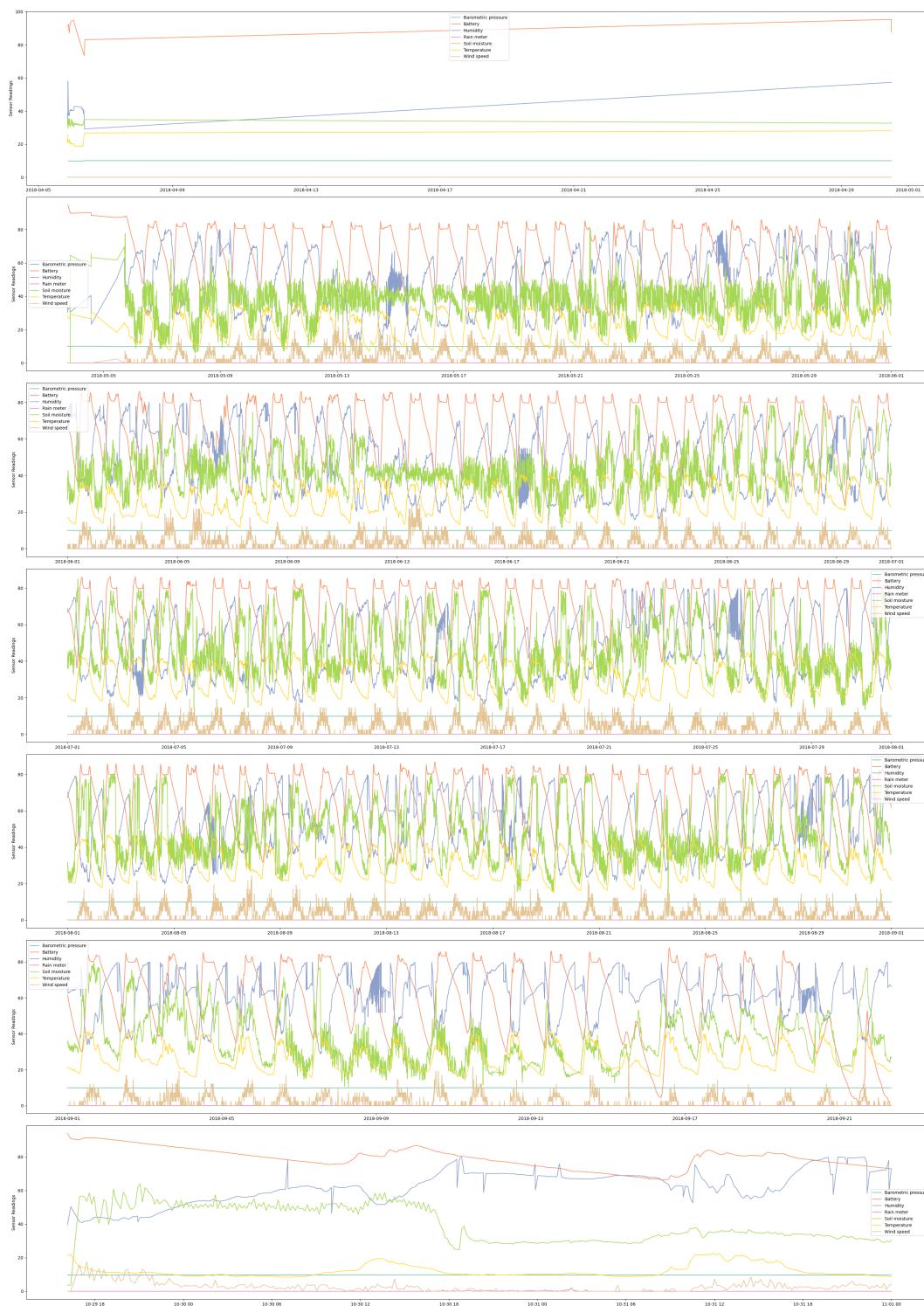
Se imputan los valores faltantes usando el método IterativeImpute. El método IterativeImputer es una técnica de imputación de valores que utiliza un regresor bayesiano como estimador y utilizando el resto de columnas para esa fila.

variable	Barometric pressure	Battery	Humidity	Rain meter	Soil moisture	Temperature	Wind speed
date							
2018-04-05 20:55:00	nan	nan	38.952393	0.000000	nan	25.941238	0.000000
2018-04-05 20:56:00	nan	nan	35.084290	0.000000	31.494141	22.509207	0.000000
2018-04-05 20:58:00	nan	91.488281	37.579102	0.000000	34.667969	21.329445	0.000000
2018-04-05 21:08:00	nan	91.488281	39.272827	0.000000	32.373047	20.578690	0.000000
2018-04-05 21:15:00	nan	91.488281	58.025879	0.000000	29.516602	22.712984	0.000000
2018-04-05 21:36:00	nan	nan	37.319702	0.000000	nan	21.769175	0.000000
2018-04-05 21:37:00	986.728729	92.437500	37.224335	0.000000	34.069824	21.683374	0.000000
2018-04-05 21:38:00	986.751251	92.201172	37.312072	0.000000	35.522461	21.619023	0.000000
2018-04-05 21:39:00	986.753753	91.884766	37.571472	0.000000	34.020996	21.613661	0.000000
2018-04-05 21:40:00	986.746246	91.726563	37.506623	0.000000	34.008789	21.608298	0.000000

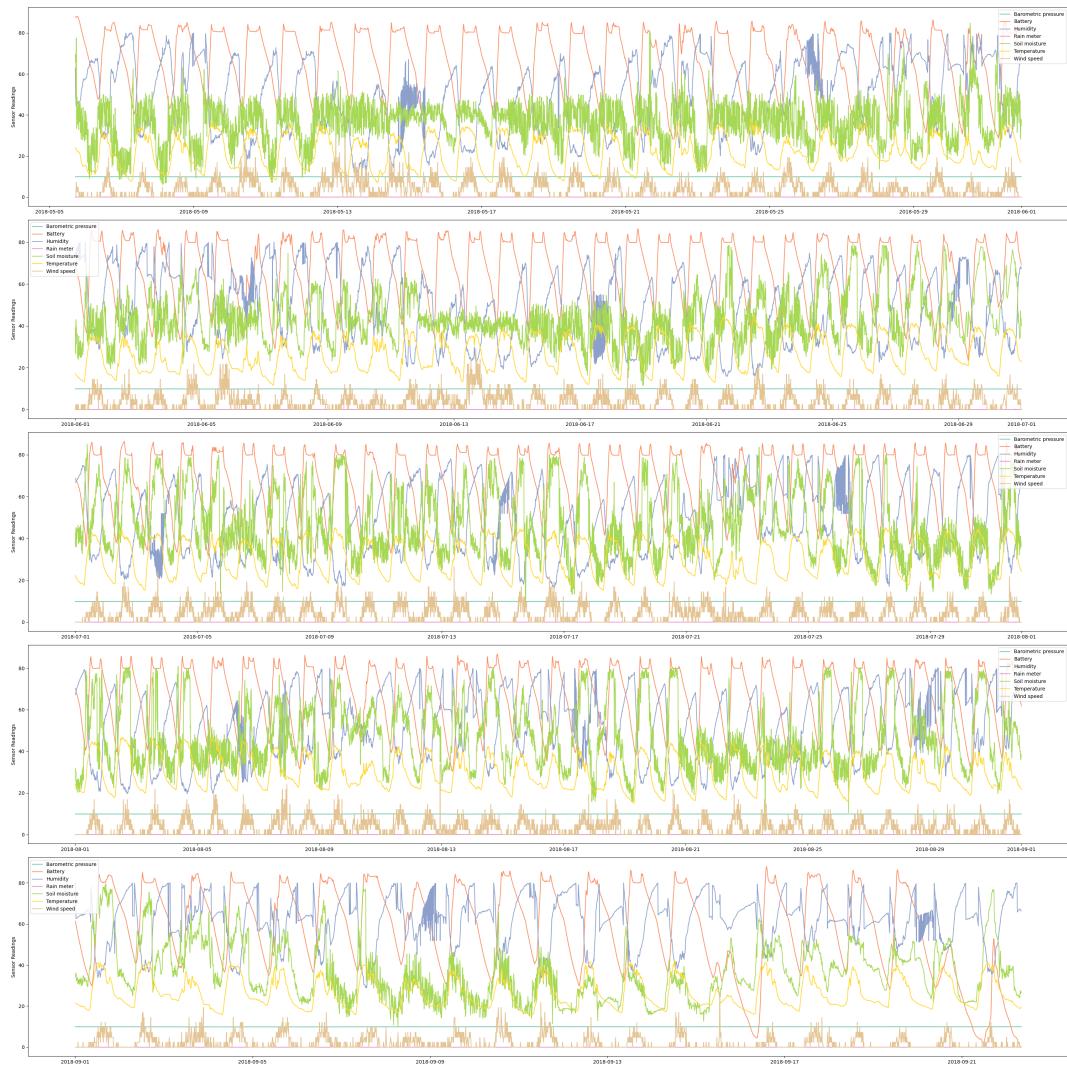
Se realizó una interpolación basada en el tiempo en el dataframe y redondearon todos los valores a 2 decimales.

variable	Barometric pressure	Battery	Humidity	Rain meter	Soil moisture	Temperature	Wind speed
date							
<b>2018-04-05 20:55:00</b>	nan	nan	38.950000 0.000000	nan	25.940000	0.000000	
<b>2018-04-05 20:56:00</b>	nan	nan	35.080000 0.000000	31.490000	22.510000	0.000000	
<b>2018-04-05 20:58:00</b>	nan	91.490000 37.580000 0.000000		34.670000	21.330000	0.000000	
<b>2018-04-05 21:08:00</b>	nan	91.490000 39.270000 0.000000		32.370000	20.580000	0.000000	
<b>2018-04-05 21:15:00</b>	nan	91.490000 58.030000 0.000000		29.520000	22.710000	0.000000	
<b>2018-04-05 21:36:00</b>	nan	92.390000 37.320000 0.000000		33.860000	21.770000	0.000000	
<b>2018-04-05 21:37:00</b>	986.730000	92.440000 37.220000 0.000000		34.070000	21.680000	0.000000	
<b>2018-04-05 21:38:00</b>	986.750000	92.200000 37.310000 0.000000		35.520000	21.620000	0.000000	
<b>2018-04-05 21:39:00</b>	986.750000	91.880000 37.570000 0.000000		34.020000	21.610000	0.000000	
<b>2018-04-05 21:40:00</b>	986.750000	91.730000 37.510000 0.000000		34.010000	21.610000	0.000000	

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							
2018-05-05 17:12:00	985.730000	87.780000	62.900000	0.000000	69.580000	24.000000	0.000000
2018-05-05 17:22:00	985.610000	87.780000	49.740000	0.000000	65.840000	24.270000	0.000000
2018-05-05 17:32:00	985.730000	87.490000	46.900000	0.000000	67.330000	23.770000	0.000000
2018-05-05 17:42:00	990.300000	87.490000	44.520000	0.000000	46.240000	23.360000	4.820000
2018-05-05 17:52:00	990.770000	87.640000	44.370000	0.280000	77.490000	23.160000	7.240000
2018-05-05 18:02:00	990.810000	87.780000	44.610000	0.000000	33.370000	22.950000	4.820000
2018-05-05 18:12:00	990.780000	87.930000	44.970000	0.000000	34.960000	22.670000	2.410000
2018-05-05 18:22:00	990.720000	87.930000	45.220000	0.000000	48.880000	22.660000	4.820000
2018-05-05 18:32:00	990.790000	88.070000	45.800000	0.000000	38.330000	22.320000	2.410000
2018-05-05 18:42:00	990.930000	88.070000	45.730000	0.000000	42.260000	22.190000	4.820000

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								
<b>2018-05-05 17:12:00</b>	985.730000	87.780000	62.900000	0.000000	69.580000	24.000000	0.000000	846
<b>2018-05-05 17:22:00</b>	985.610000	87.780000	49.740000	0.000000	65.840000	24.270000	0.000000	846
<b>2018-05-05 17:32:00</b>	985.730000	87.490000	46.900000	0.000000	67.330000	23.770000	0.000000	846
<b>2018-05-05 17:42:00</b>	990.300000	87.490000	44.520000	0.000000	46.240000	23.360000	4.820000	846
<b>2018-05-05 17:52:00</b>	990.770000	87.640000	44.370000	0.280000	77.490000	23.160000	7.240000	846
<b>2018-05-05 18:02:00</b>	990.810000	87.780000	44.610000	0.000000	33.370000	22.950000	4.820000	846
<b>2018-05-05 18:12:00</b>	990.780000	87.930000	44.970000	0.000000	34.960000	22.670000	2.410000	846
<b>2018-05-05 18:22:00</b>	990.720000	87.930000	45.220000	0.000000	48.880000	22.660000	4.820000	846
<b>2018-05-05 18:32:00</b>	990.790000	88.070000	45.800000	0.000000	38.330000	22.320000	2.410000	846
<b>2018-05-05 18:42:00</b>	990.930000	88.070000	45.730000	0.000000	42.260000	22.190000	4.820000	846