

ANEXO: ANÓMALOS II

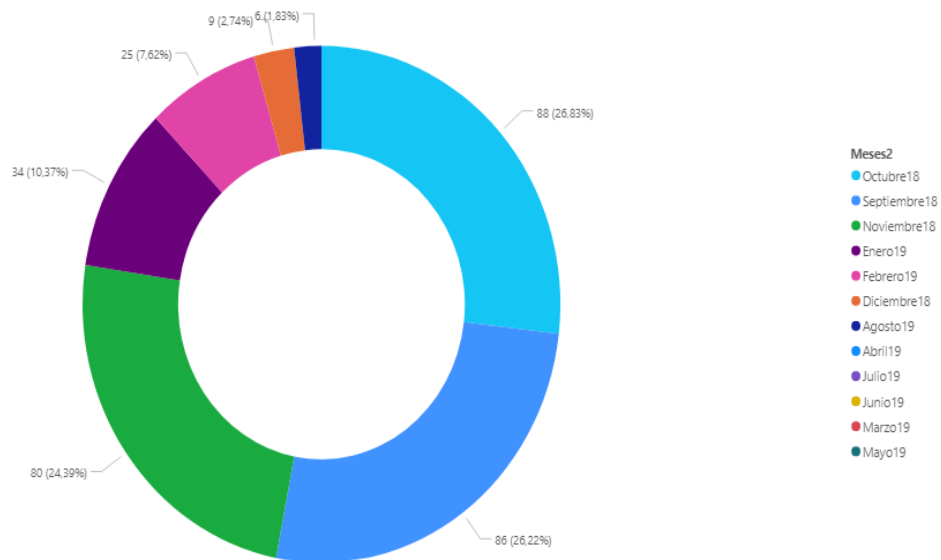
En este documento se presentarán de manera gráfica las variables de la base de datos que recogía la información de la cantidad de cada tipo de dato anómalo. Esta era la base de datos.

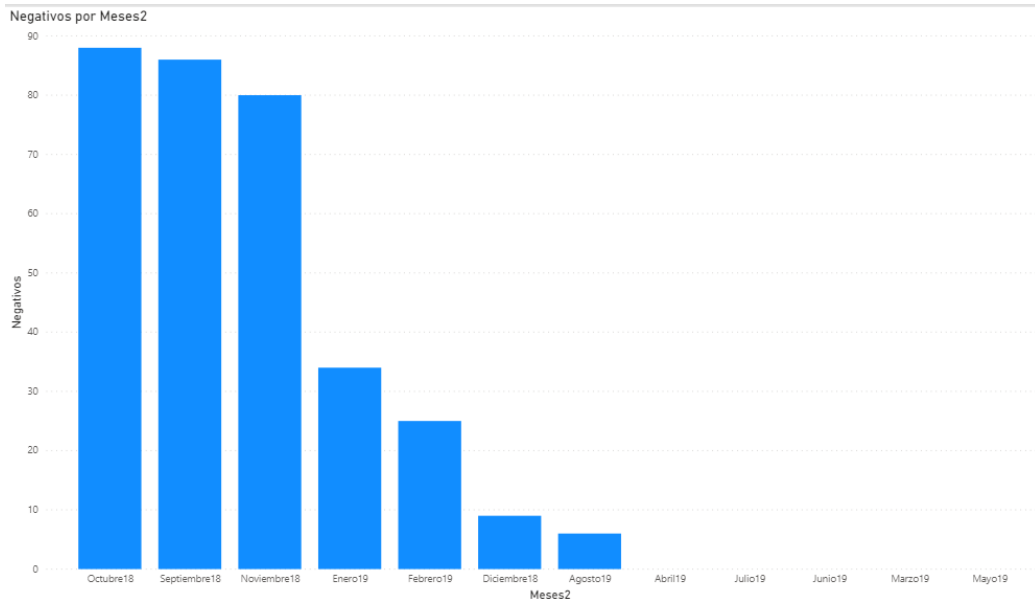
| Meses | Negativos | Nulos | NaN | Mes Siguient | NegCada15 | NulCada15 | Atípicos | Clima | Edificios Antr | Edificios Des | Total Filas Ar | Total Filas Después |
|------------|-----------|-------|-------|--------------|-----------|-----------|----------|--------|----------------|---------------|----------------|---------------------|
| Septiembre | 86 | 18 | 8970 | 141 | 354 | 83730 | 242 | 163722 | 143 | 85 | 406201 | 211446 |
| Octubre | 88 | 0 | 6702 | 138 | 737 | 102847 | 278 | 169699 | 142 | 86 | 421999 | 220006 |
| Noviembre | 80 | 8 | 9481 | 137 | 376 | 93363 | 199 | 164066 | 142 | 84 | 407870 | 213700 |
| Diciembre | 9 | 4 | 4782 | 141 | 298 | 106505 | 201 | 169605 | 143 | 85 | 423807 | 221949 |
| Enero | 34 | 163 | 31183 | 140 | 300 | 83137 | 150 | 169604 | 143 | 85 | 424345 | 214019 |
| Febrero | 25 | 67 | 11493 | 143 | 157 | 59301 | 140 | 114969 | 143 | 85 | 287835 | 150417 |
| Marzo | 0 | 0 | 1556 | 140 | 100 | 110408 | 5 | 167699 | 143 | 86 | 420789 | 222671 |
| Abril | 0 | 0 | 1485 | 142 | 97 | 1099979 | 1 | 163588 | 143 | 85 | 408976 | 215108 |
| Mayo | 0 | 0 | 1584 | 140 | 47 | 104378 | 0 | 163254 | 144 | 86 | 408928 | 217288 |
| Junio | 0 | 0 | 960 | 141 | 489 | 93227 | 1 | 159571 | 144 | 86 | 401754 | 212933 |
| Julio | 0 | 0 | 0 | 138 | 864 | 85779 | 0 | 163265 | 142 | 86 | 410426 | 218789 |
| Agosto | 6 | 11 | 0 | 139 | 283 | 136722 | 4 | 160050 | 143 | 86 | 396310 | 191216 |

Encontramos 6 tipos que son anómalos (negativos, nulos, Nan, Mes siguiente, Negativos cada 15 minutos, atípicos), pero eliminamos 2 más (clima y nulos cada 15 min) por decisiones tomadas en común que ya hemos explicado. A continuación, se muestran los gráficos de cada tipo en función de los meses.

Negativos

Negativos por Meses2

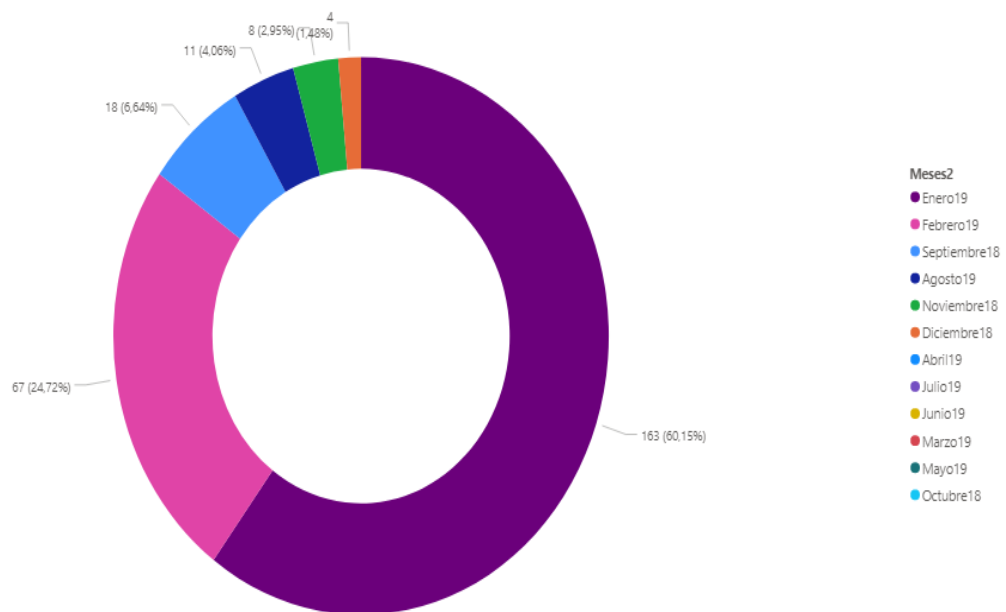




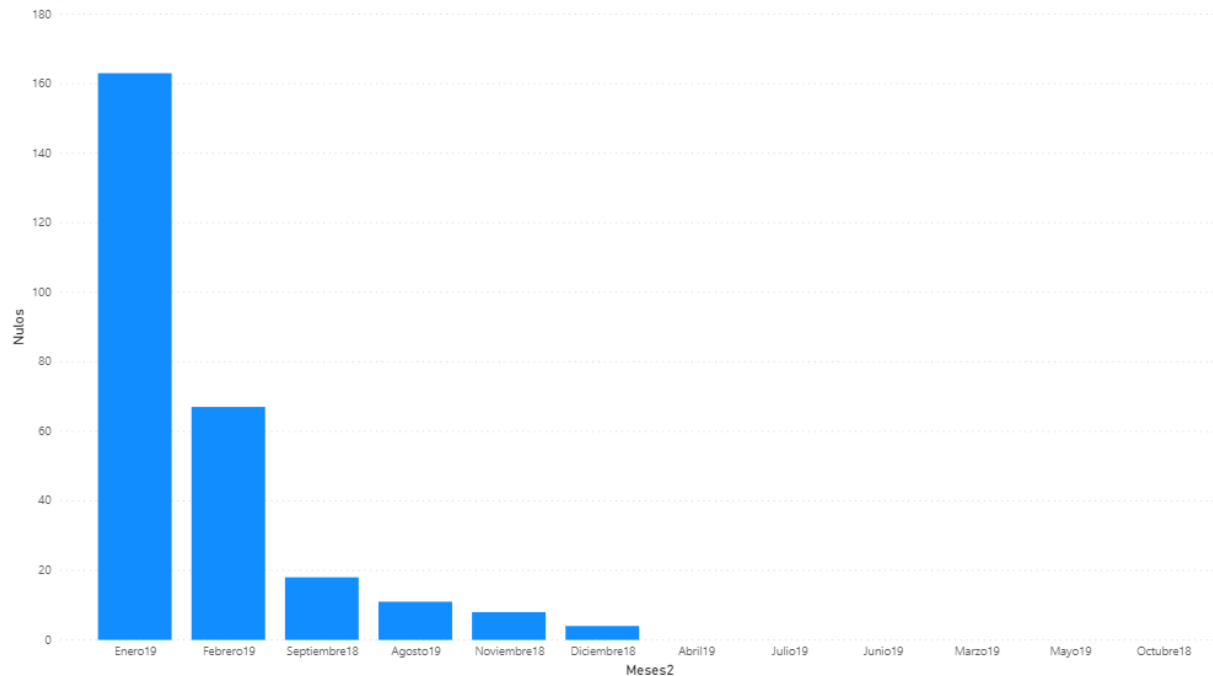
Mediante el Gráfico de Rosquilla y el gráfico de barras podemos observar que en los meses (octubre, septiembre, noviembre, enero, febrero y diciembre) existe una gran cantidad de datos negativos, lo que nos permite deducir que durante este período de tiempo hubo muchos fallos en los sensores. Destacar por último que a partir del mes de agosto este porcentaje sufrió una gran disminución.

Nulos

Nulos por Meses2



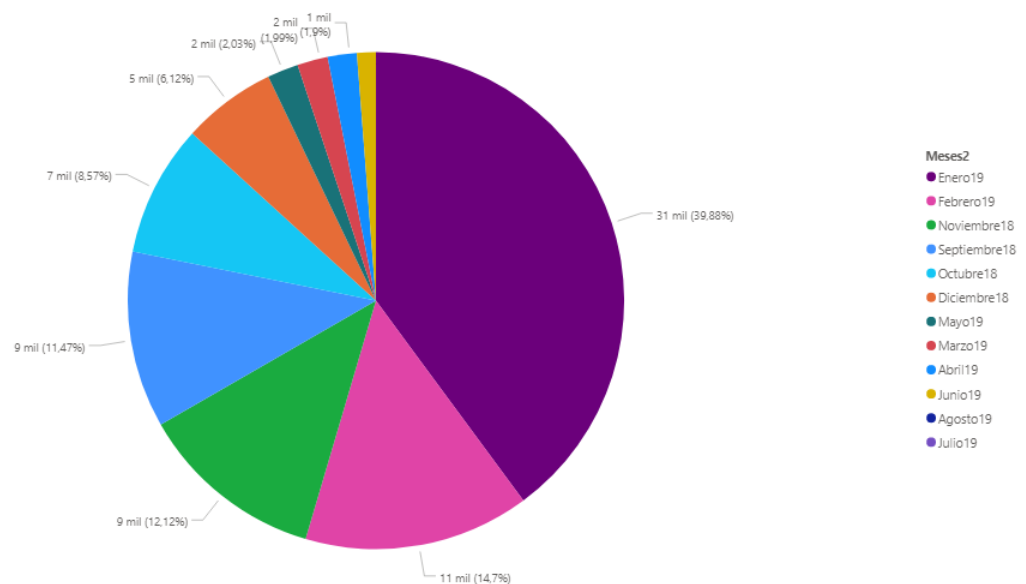
Nulos por Meses2



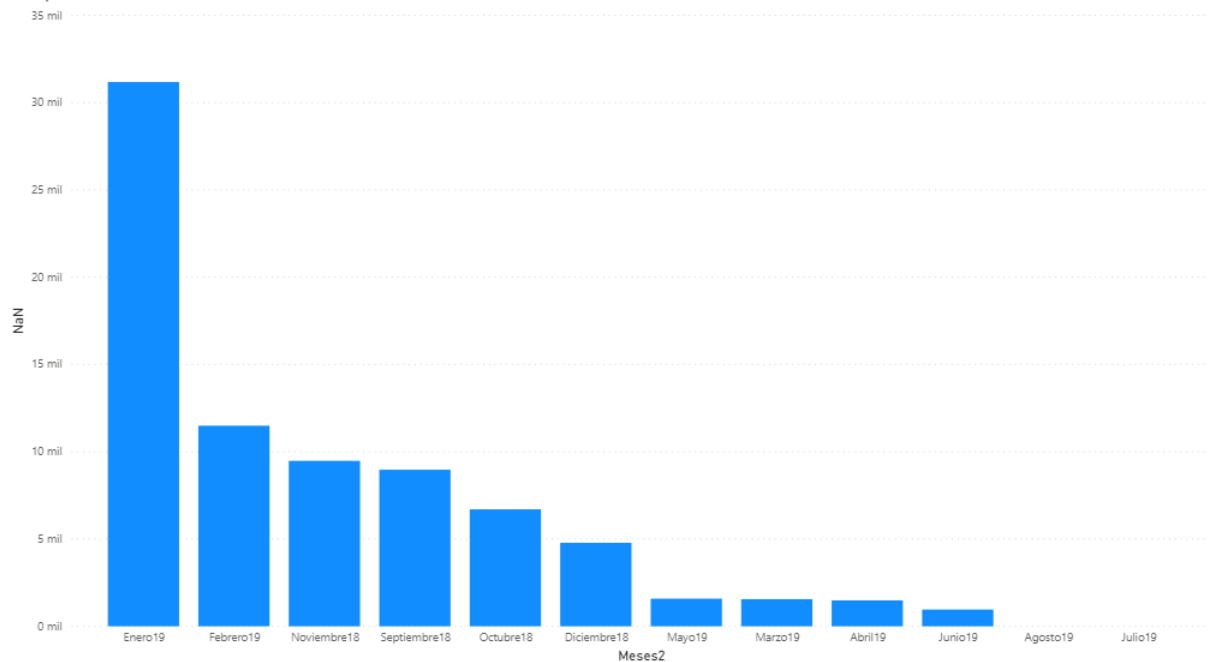
Observando las gráficas de datos nulos, también relacionada con fallos de los sensores, observamos que durante esos 5 meses encontramos un gran presencia de datos nulos(de manera decreciente). Estos datos nulos aparecían tras un cierto periodo de tiempo dependiendo del mes e interrumpen una información de datos acumlados recogidos cada 15 min. Es decir, en la casilla se presenta un 0.

Nan

NaN por Meses2



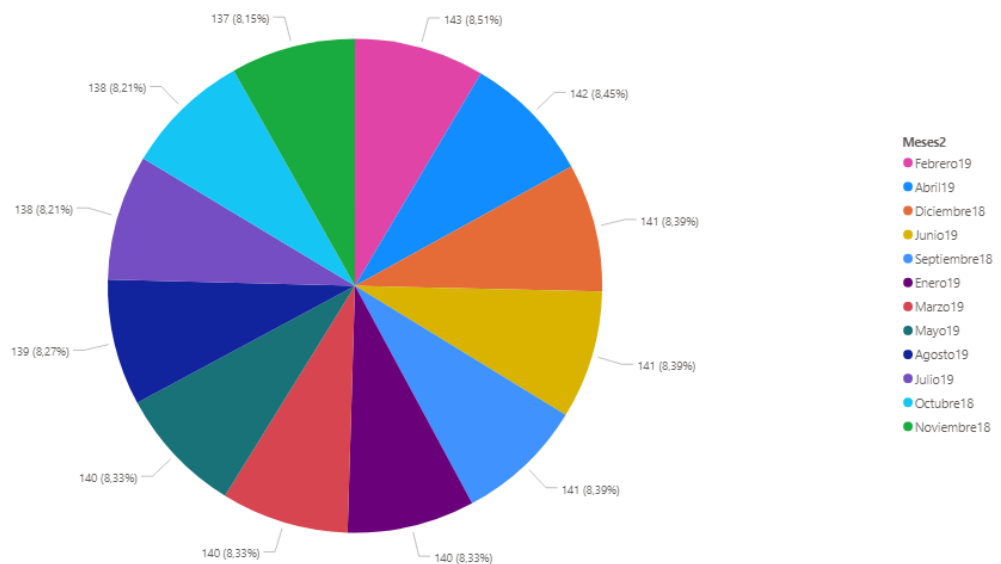
NaN por Meses2



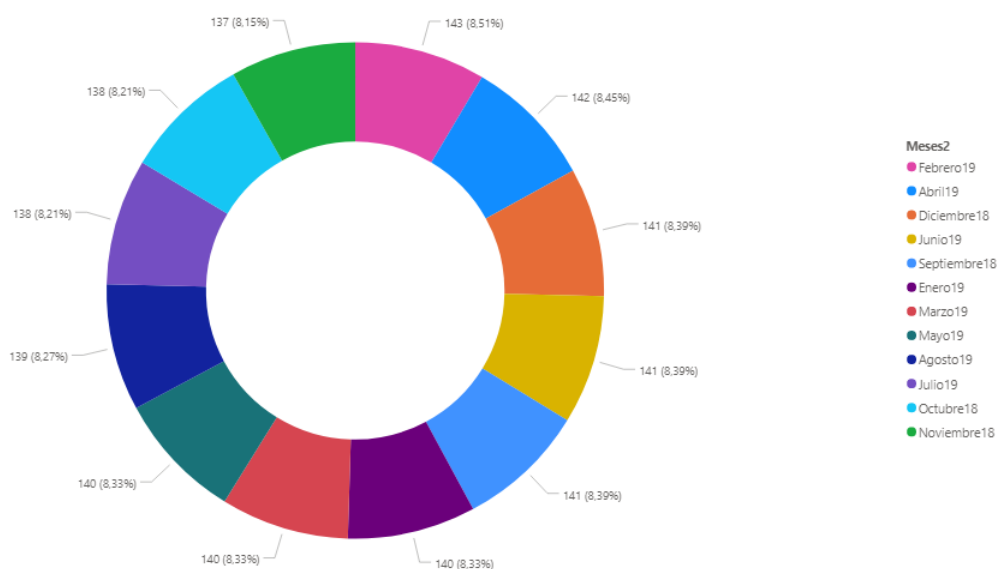
En esta ocasión mediante un gráfico de tarta y diagrama de barras analizamos los datos con casilla vacía(Nan) de forma decreciente. Observamos que la mayoría de los meses tienen una falta importante de datos, también debido a los fallos de los sensores.

Mes siguiente

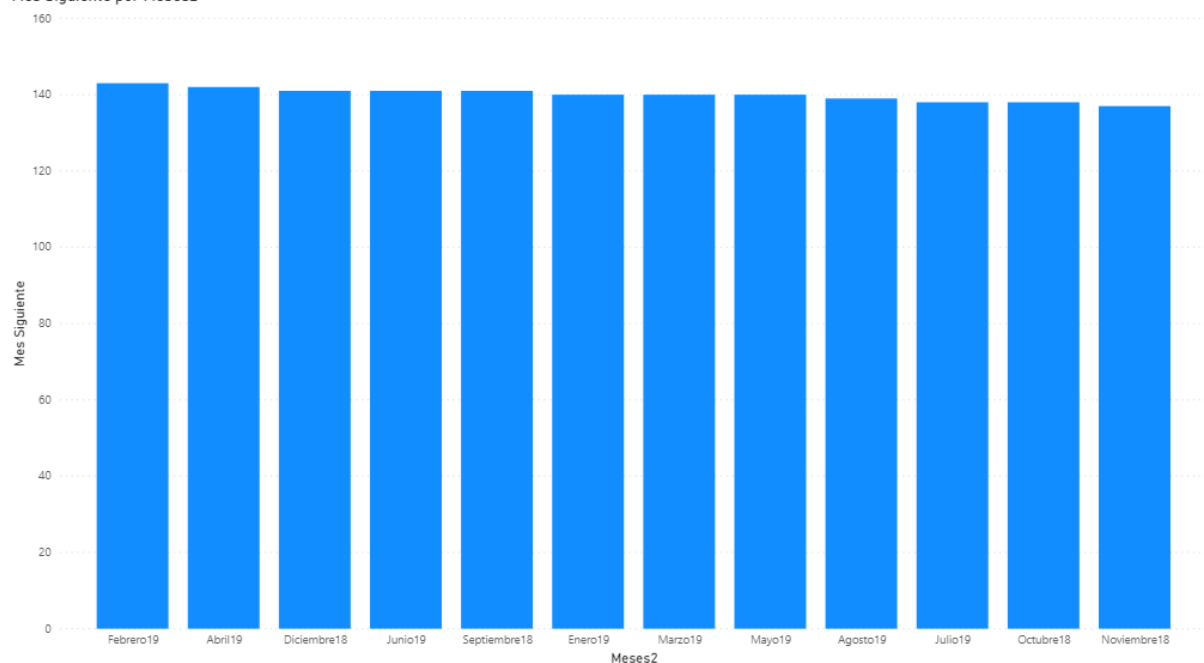
Mes Siguiente por Meses2



Mes Siguiente por Meses2



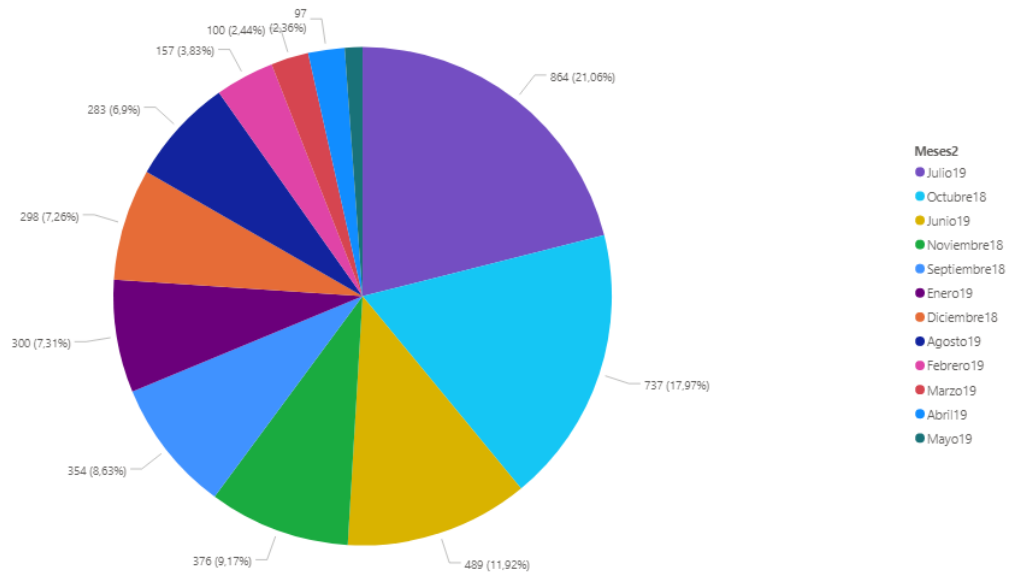
Mes Siguiente por Meses2



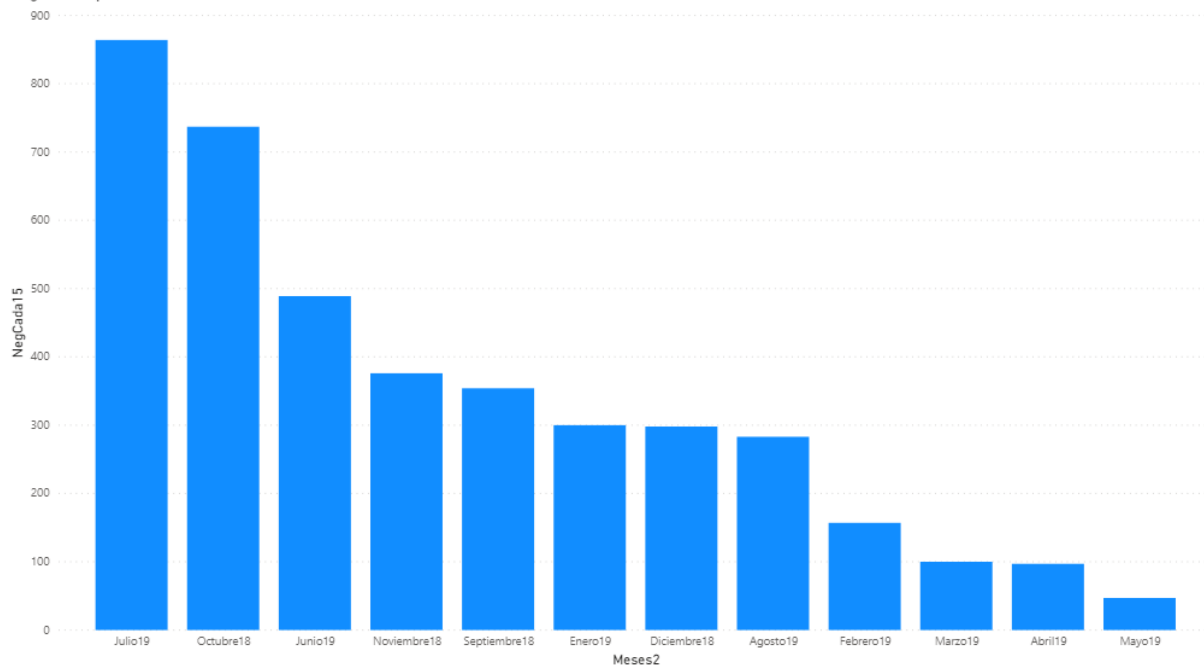
Remarcar que en todos los meses de la base datos hemos encontrado muchos datos repetidos de diferentes meses que aparecen en otros meses.

Negativos Cada15 min

NegCada15 por Meses2



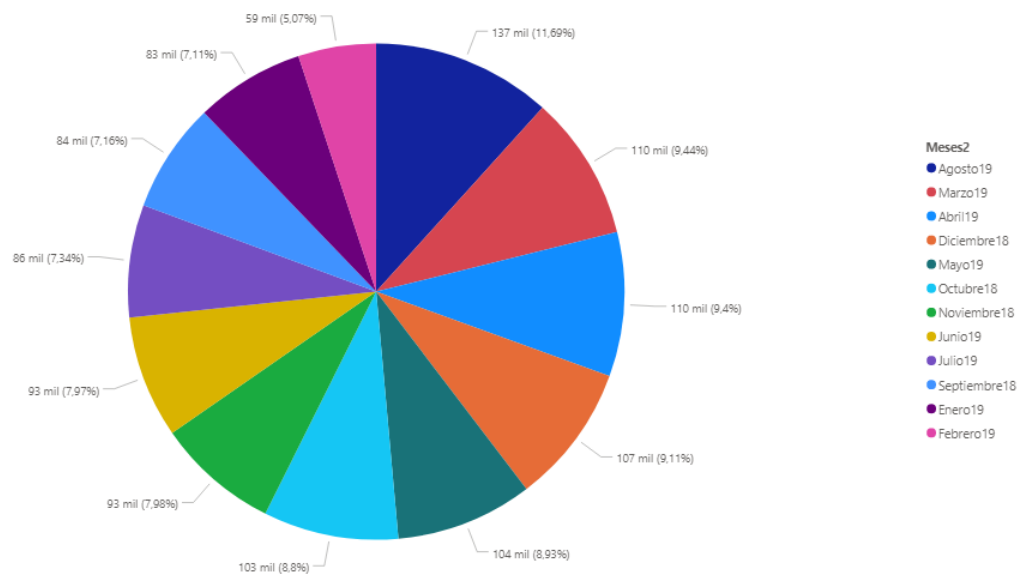
NegCada15 por Meses2



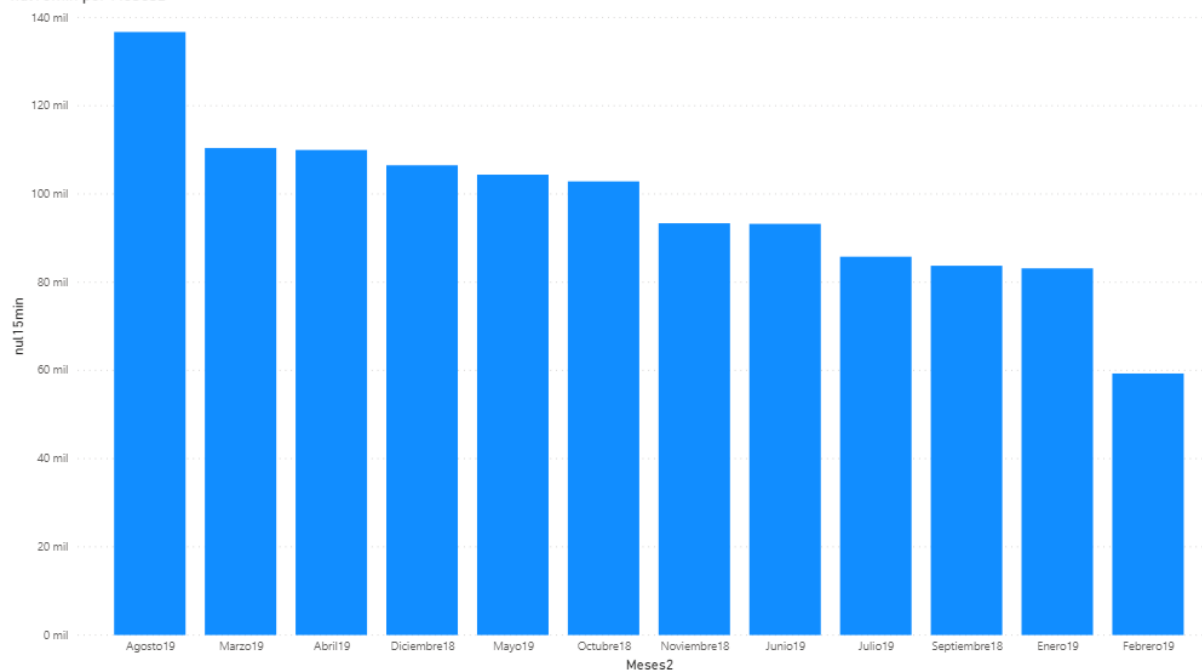
Resaltar que respecto a la nueva columna que creamos(Neg/15min) en la mayoría de los meses hay una gran cantidad de datos negativos, en algunos meses más y otros menos, pero en general un alto porcentaje.

NulCada15

nul15min por Meses2



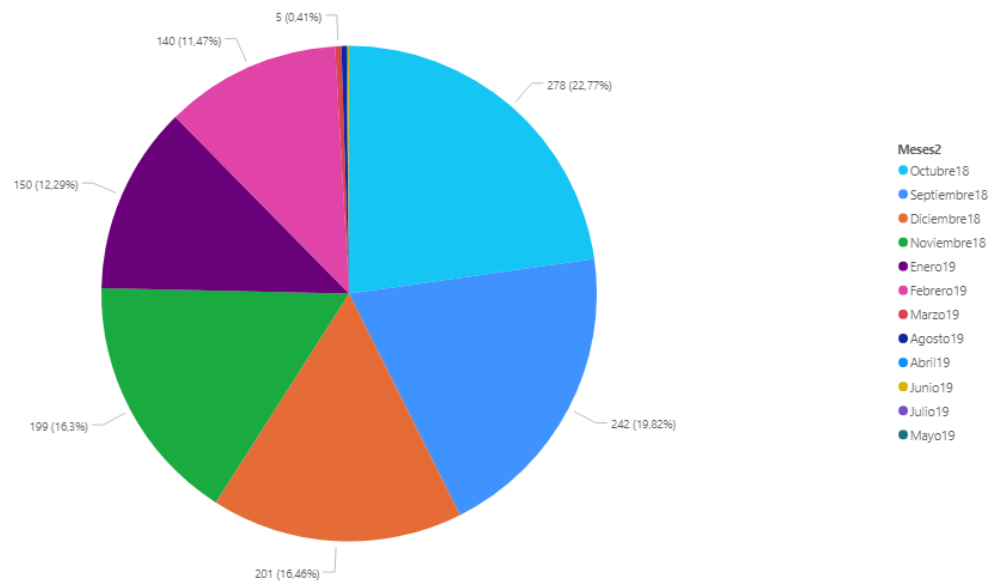
nul15min por Meses2



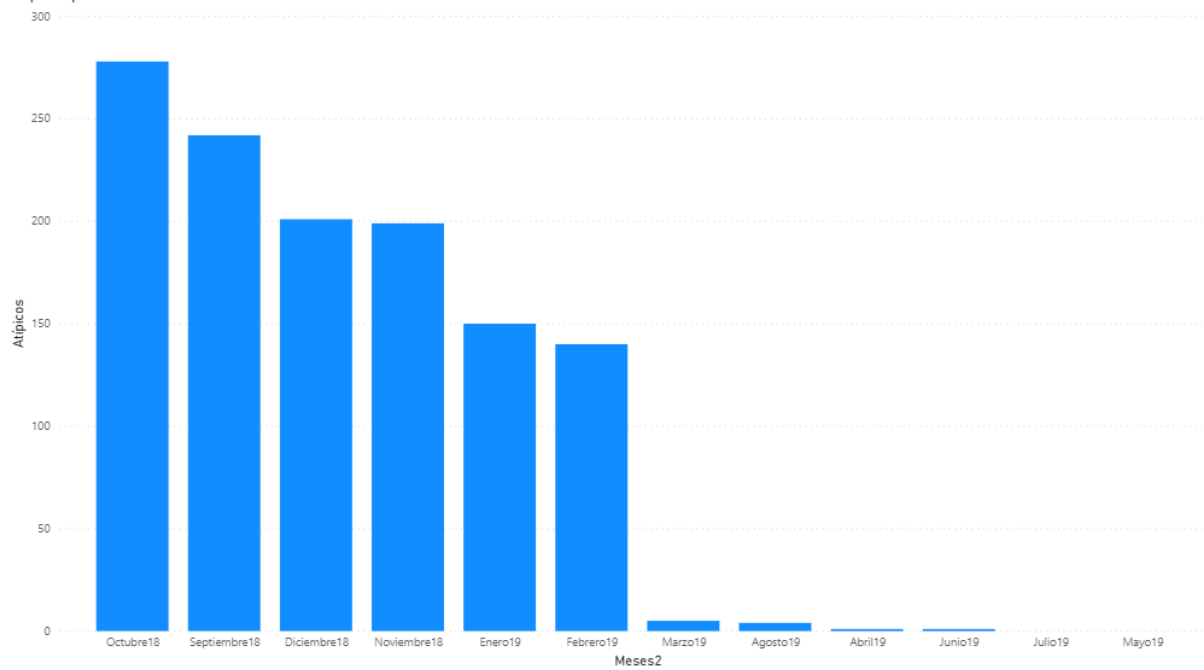
Observamos que los nulos que se crean en la columna nueva están repartidos de manera muy similar en los distintos meses. Esto no nos ayudan a decidir si cada edificio de la Upv consumen más o menos. Por último recalcar que no siempre son fallos de sensores, también pueden ser que no haya habido ningún consumo durante ese periodo.

Atípicos

Atípicos por Meses2



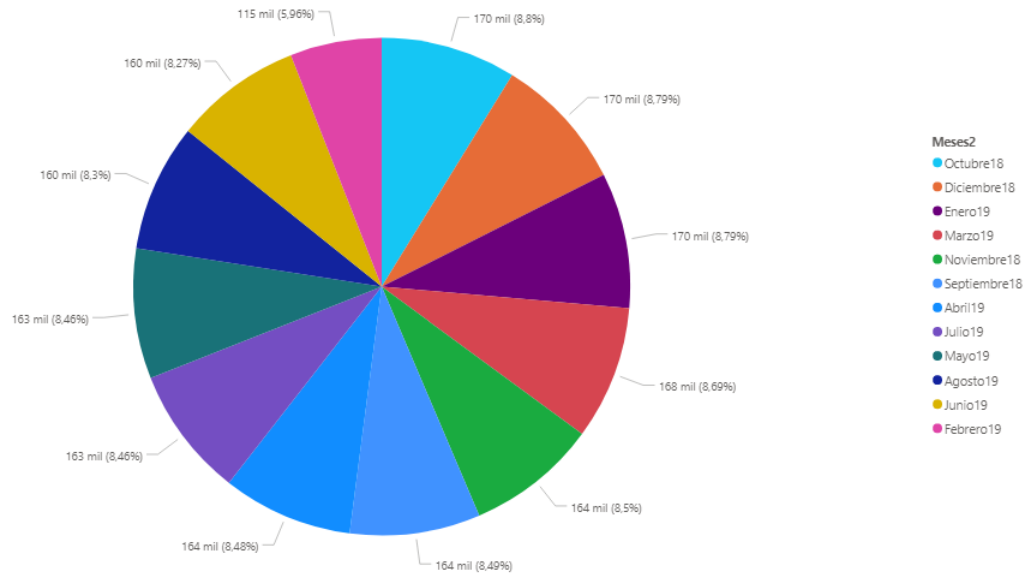
Atípicos por Meses2



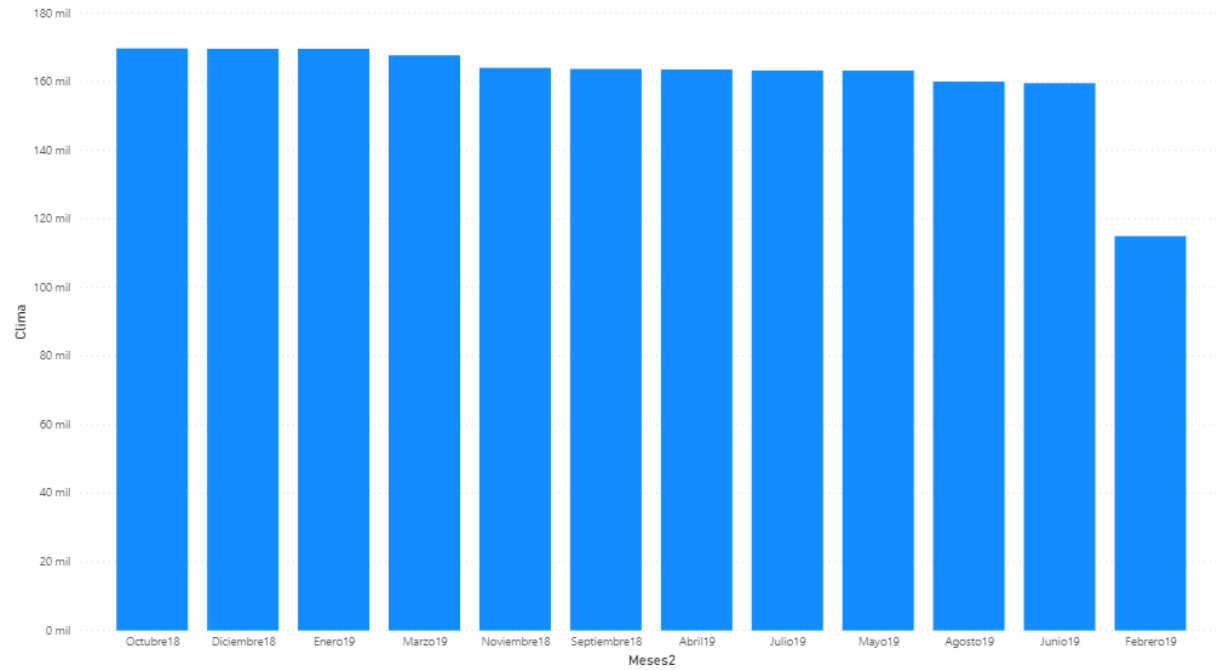
En estos datos, también vemos la evolución comentada. Es decir, los primeros 6 meses tienen muchos más valores que en los 6 últimos.

Clima

Clima por Meses2



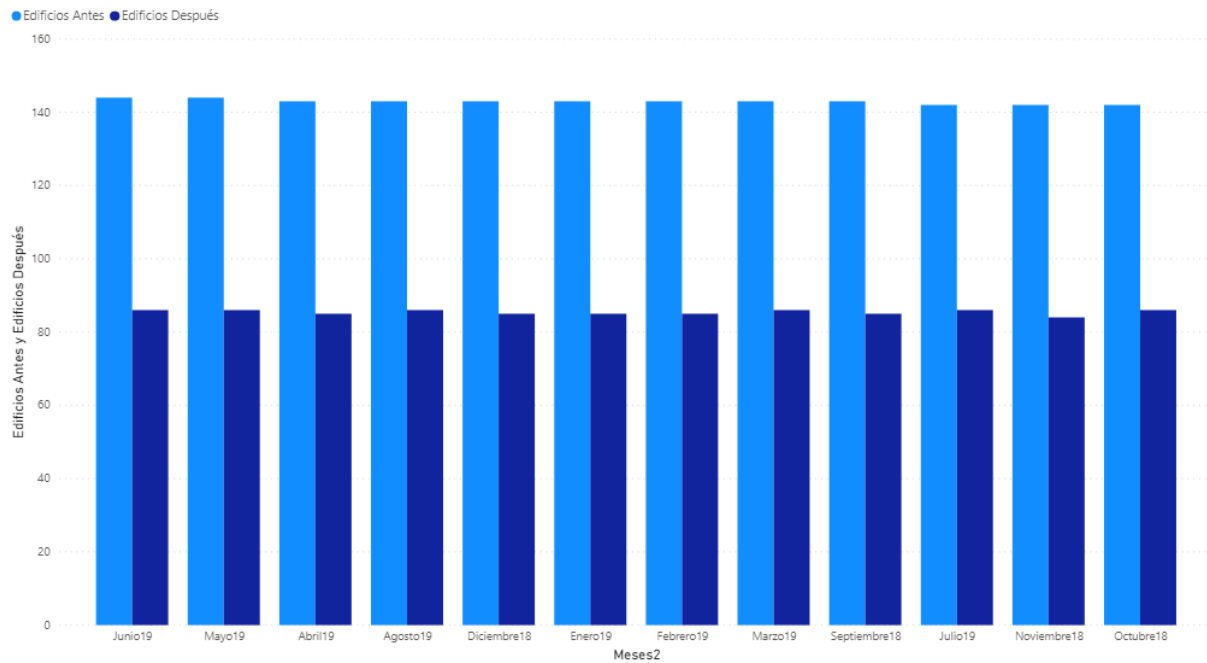
Clima por Meses2



Los datos en relación al clima se presentan en la misma frecuencia en todos los meses. Febrero tiene menos puesto que tiene menos días el mes.

Edificios antes y después

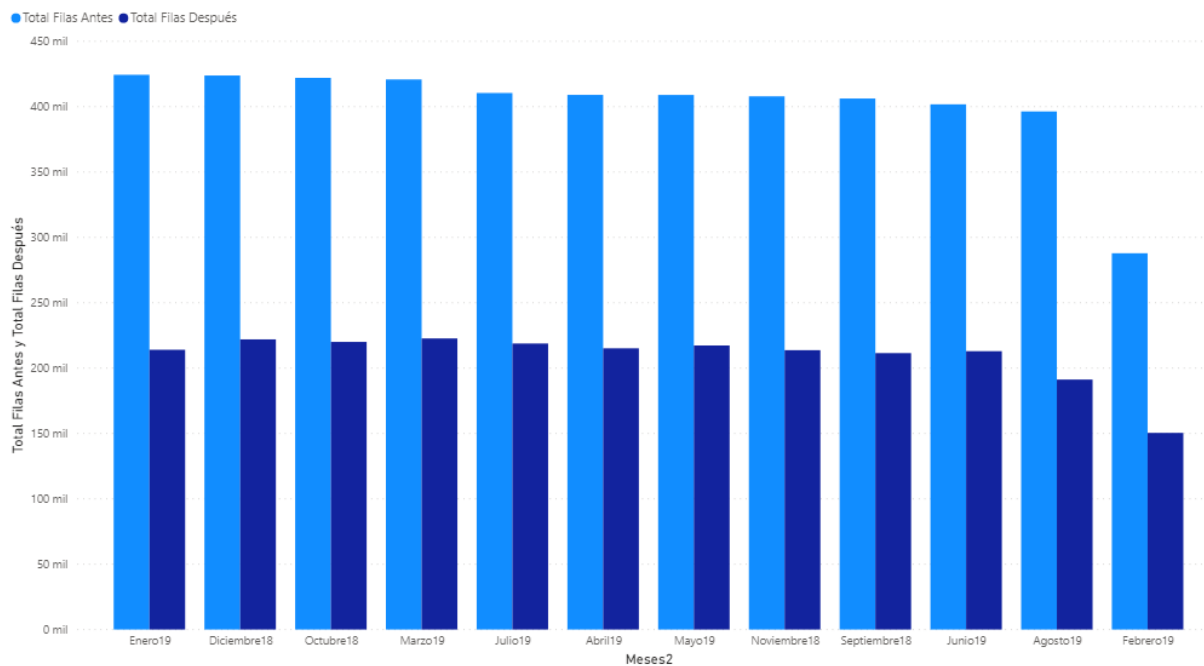
Edificios Antes y Edificios Después por Meses2



Se reduce de la misma forma en todos los meses. La explicación de la disminución es que existen edificios dedicados a recoger información de clima solamente.

Filas antes y después

Total Filas Antes y Total Filas Después por Meses2

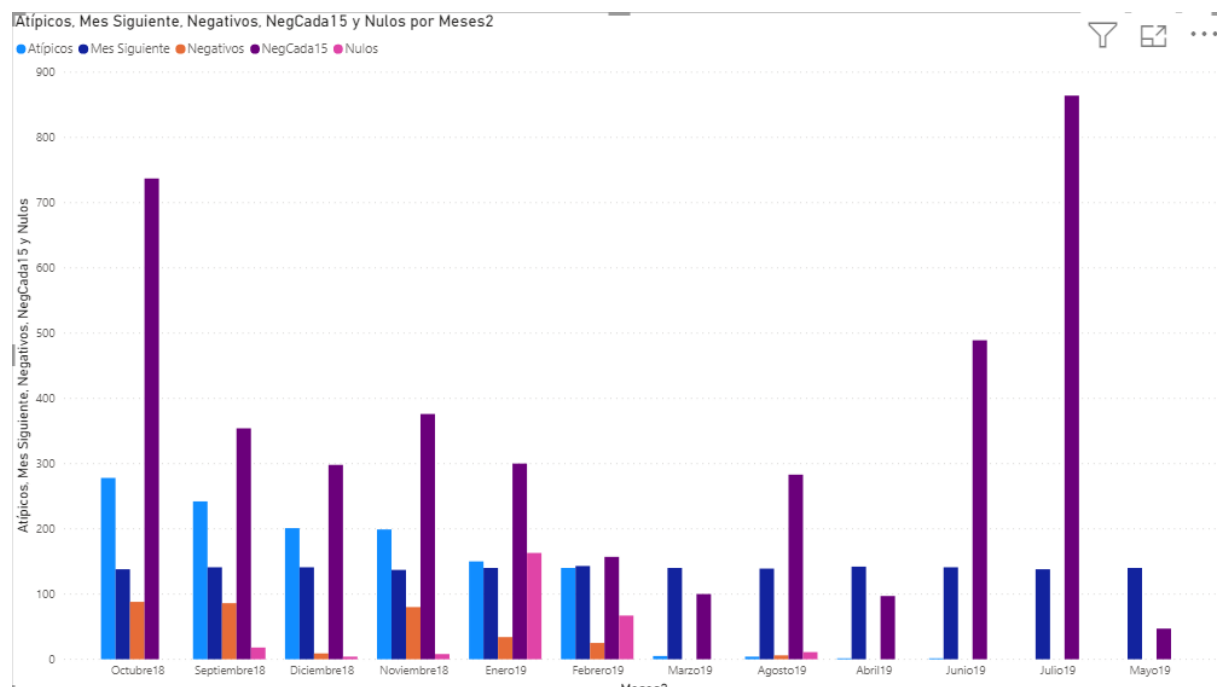


En todos los meses se reduce en la misma proporción. Esta disminución se debe a la limpieza de los datos.

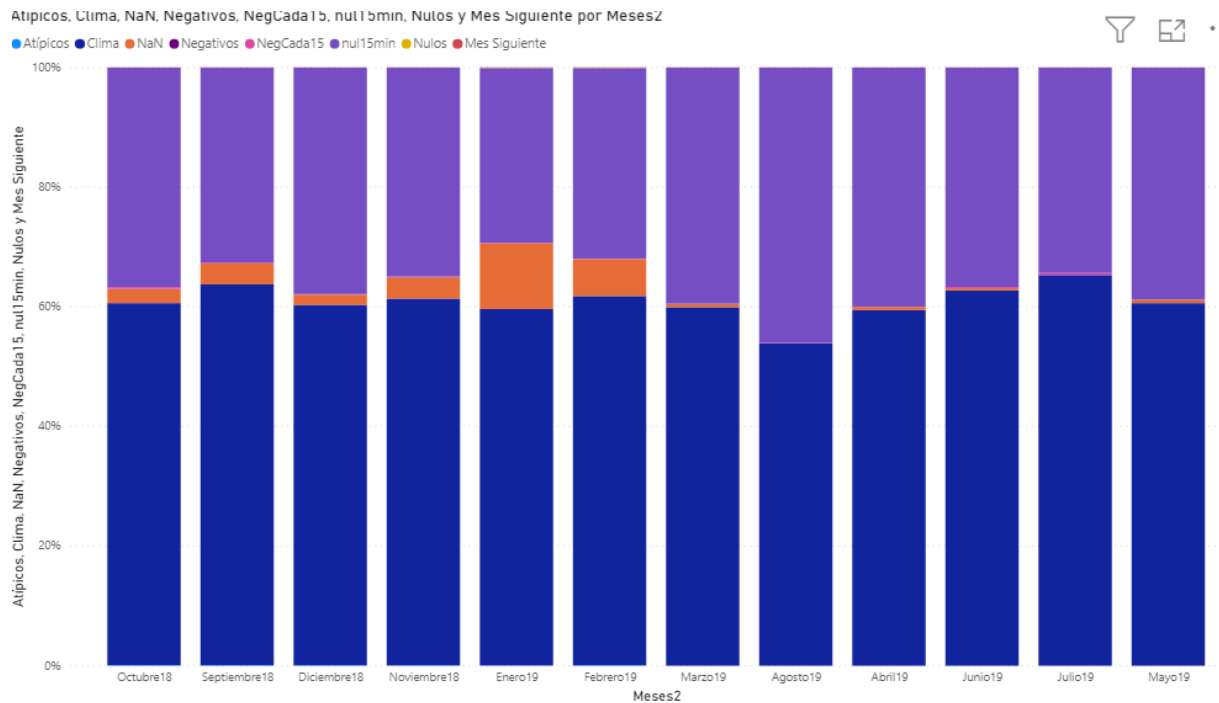
Con el objetivo de ver la calidad de los sensores haremos un análisis de los 6 tipos de fallos que produce.



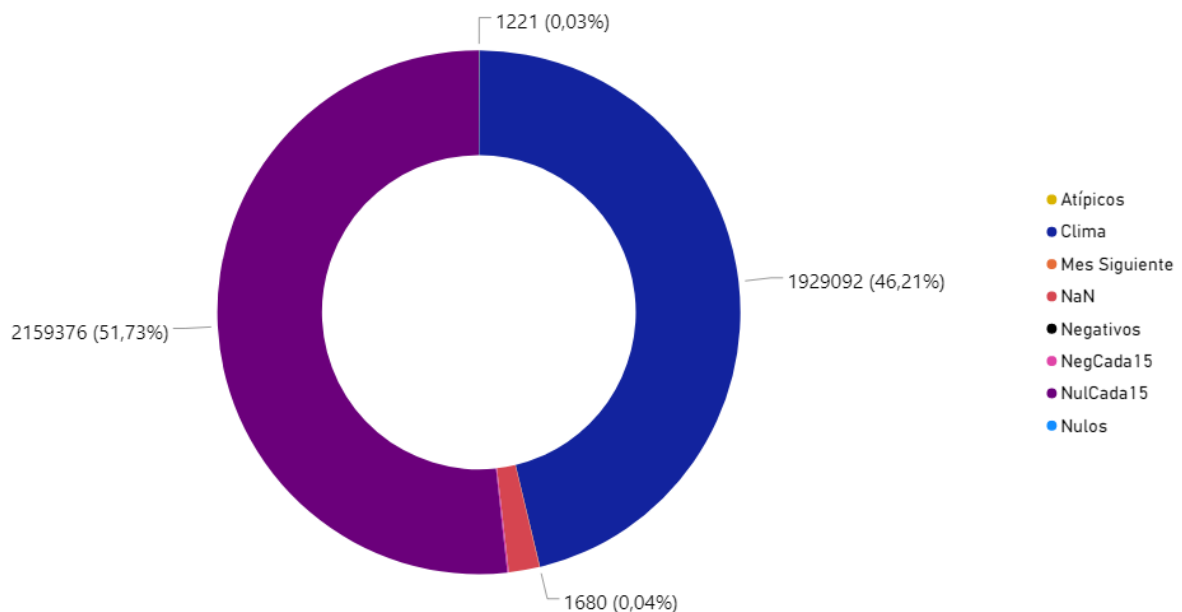
Para poder ver con más detalle los 5 tipos que no podemos ver en este gráfico puesto que el tipo NaN los eclipsa, quitaremos este y analizaremos el resto.



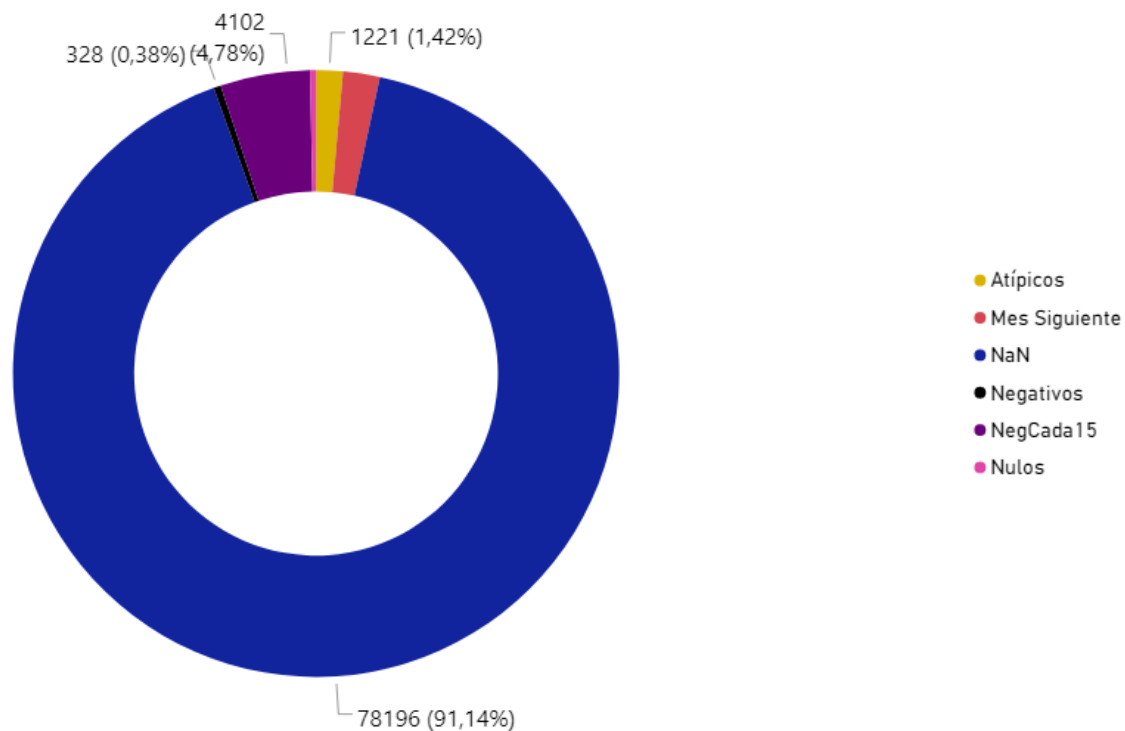
No obstante hemos eliminado estos 6 tipos de anómalos y los datos de 'CLIMA' de la columna 'Uso_Energia' y los nulos de la columna 'KWH /15' . Estos los hemos quitado puesto que no ayudaban a conseguir nuestro objetivo. Sus cantidades eran muy grandes como podemos ver a continuación. Eclipsan al resto de tipos.



En este gráfico se encuentran todos los datos borrados separados por meses.

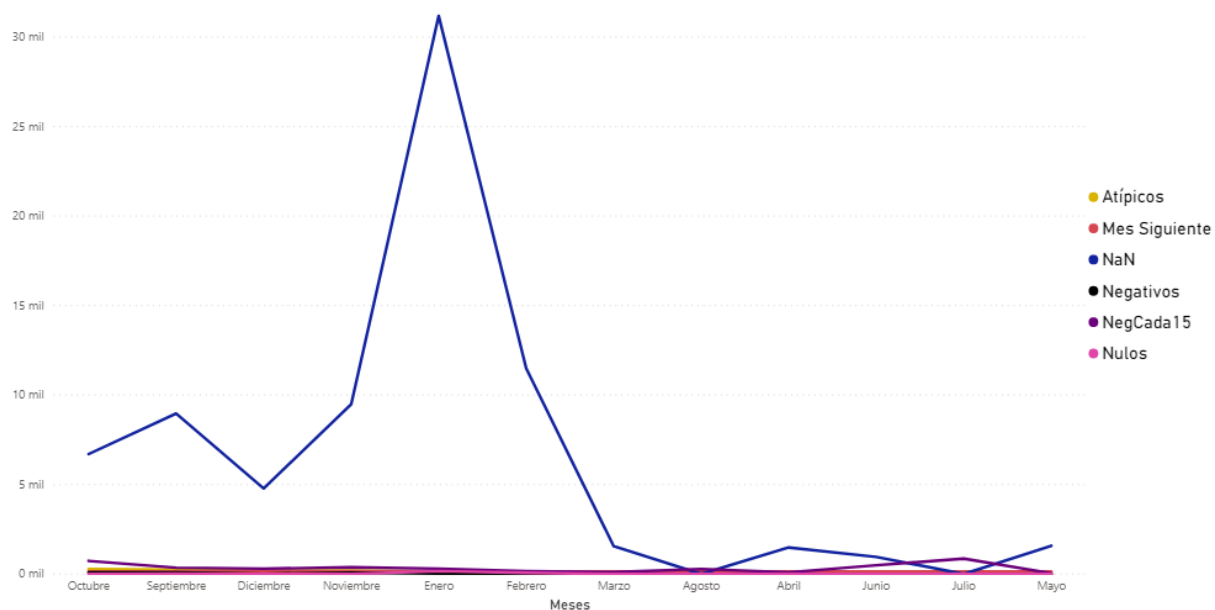


Este gráfico representa todos los tipos eliminados. Como se puede ver, los de clima y nulos en la columna creada de consumo neto son los que aparecen en más frecuencia. No obstante, estos no se producen por error de los sensores, pero no ayudan a conseguir nuestro objetivo.



Respecto a los fallos por los sensores, la mayoría son de tipo Nan, con un recuento total de 78196 y representa el 91,14% de todos los fallos producidos.

Con el objetivo de identificar qué ha pasado con los tipo Nan, hemos desglosado esta información por meses.



Como veníamos comentando, esta gran cantidad se remarca en los primeros 6 meses y se corrige en los siguientes.