



Edgard Leonardo Castañeda García  
Johan José Castañeda Vega  
John Alexander Cárdenas

## Problema 01

### Caracterización del consumo del edificio 471-Medicina UN

#### Índice

<b>1. Análisis del proyecto</b>	<b>1</b>
<b>2. Metodología</b>	<b>3</b>
<b>3. Análisis y auditoría de datos</b>	<b>4</b>
3.1. Muestreo . . . . .	4
3.2. Tensiones por fase (V) . . . . .	4
3.3. Corrientes por fase (A) . . . . .	5
3.4. Factor de potencia . . . . .	7
3.5. Potencia (kW) . . . . .	8
3.6. Potencia por fase (kW) . . . . .	10
3.7. Energía Activa (kWh) y Energía Reactiva (kvar) . . . . .	12

#### 1. Análisis del proyecto

A continuación, se expone el análisis y la valoración de la calidad de los datos de mediciones de 220 paneles solares instalados en el edificio de medicina 417 del campus UN, reportados por un Instrumento de Medición Avanzado (AMI) y almacenados al sistema de monitoreo central. La universidad tiene como plan a mediano plazo mejorar el uso de la energía eléctrica en cada uno de los edificios mediante un monitoreo permanente con el objetivo de identificar consumos anómalos para su corrección oportuna, para llevar a cabo este fin a los datos proporcionados por La Oficina de Gestión Ambiental (OGA) de la sede Bogotá, se les da una valoración de calidad y la identificación de datos y valores extraños o no existentes.

Las mediciones suministradas vienen de un AMI modelo Pilot SPM33, reportando variables de Energía activa, energía reactiva, potencia activa, potencia aparente, potencias, tensiones y corrientes por fase,



---

el factor de potencia, frecuencia y la distorsión armónica total en tensión y corriente.

Para tener una vista más amplia del problema desde la perspectiva de la energía y la informática se muestra el análisis de las siguientes dimensiones

1. ***Dominio***

El problema que se aborda es la caracterización del consumo de energía eléctrica del edificio de medicina 417 del campus universidad nacional sede Bogotá. Esto ubica el dilema en la parte del operador de red de la cadena de valor, ya que este sería el sistema central de monitoreo que recibe y almacena las mediciones dadas por la AMI, analizándose desde la perspectiva del prosumidor que es el edificio en estudio el cual produce y consume a través de los 220 paneles solares.

2. ***Impacto efectivo***

El funcionamiento de los 220 paneles solares instalados en 2018 en la facultad de medicina reduce el gasto entre un 30 y 50 % del consumo en la hora pico de esta facultad dando como ahorro aproximadamente 30 millones de pesos al año. En ese año el edificio estaba consumiendo aproximadamente unos 120 kwh en consumo pico por lo que estimando una entrega de 60 kwh por los paneles se supliría el 50 % de la demanda pedida de la red con una amortización de siete años y una vida útil restante de 13.

3. ***Ciclo de vida***

El proyecto de implementación de energías renovables en la U.N se encuentra en la etapa de operación ya que empezó su funcionamiento en el año de 2018 con una estimación de veinte años de vida útil, también se le debe realizar mantenimiento de tres a cuatro veces por año el cual consiste en la limpieza de suciedad y polvo.

4. ***Flujo de datos***

Los datos son provenientes de las mediciones reportadas por el AMI y almacenadas en el sistema de monitoreo central del campus, estos no tienen ningún procesamiento hasta el momento.

5. ***Categoría del problema de datos***

La categoría del problema de datos al que se enfrenta esta etapa del proyecto es la de análisis de desviación ya que el enfoque es identificar y explicar los datos atípicos o anómalos que hay en el base de datos suministrada puesto que esta no ha tenido ningún procesamiento.

6. ***Método de solución***

El método de solución que se emplea para resolver el problema de caracterización de consumo del edificio de medicina es el de descubrimiento de patrones ya que se trabaja con variables eléctricas que tienen un comportamiento conocido y con este método es una buena herramienta para encontrar subgrupos de un comportamiento diferente.



## 2. Metodología

Para realizar el análisis de los valores atípicos o anómalos de la base de datos proporcionada por la OGA se tiene en consideración que el equipo de medida es el AMI modelo Pilot SPM33 y su ficha técnica que se muestra en el cuadro 1.

Parámetro	Exactitud [%]	Rango de medición
Voltaje	0.5	Linea-Linea : 0 - 650 V
		Linea- Neutro : 0 -400 V
		PT lado primario : Hasta 650 Kv
		PT lado secundario : 100- 400 V (L-N)
Corriente	0.5	Cada fase: 0-65000 A
		Secuencia cero: 0 -65000 A
Factor de potencia	0.5	-1/1
Potencia activa	0.5	0-99999999.9 W
Potencia reactiva	1	0-99999999.9W
Energía Activa	1	0-99999999.9 kWh
Energía Reactiva	2	0-99999999.9 kVarh
Desbalance de tensión trifásico	Class B	0 %-100 %
Desbalance de corriente trifásico	Class B	0 %-100 %
THD	Class B	0 %-100 %

Tabla 1: Ficha técnica AMI Pilot SPM33. Disponible en <http://en.pmac.com.cn/images/download/20160425dmihru.pdf>

Inicialmente, se cargaron los datos entregados por el medidor dentro de un programa de python para realizar un primer análisis. Lo primero que salta a la vista es la ausencia total de datos en algunas columnas de parámetros de medición:

- Potencia reactiva (kVAR).
- Distorsión armónica de corriente y tensión.
- La potencia aparente (kVA).
- Frecuencia (Hz).
- Generación de energía.

Estas columnas fueron removidas para poder realizar una visualización más cómoda de los parámetros restantes.

Posteriormente, se realizó la caracterización de cada uno de los parámetros medidos por el equipo y se sugirieron posibles acciones que ayudarían a contar con un conjunto de datos de mayor calidad.



### 3. Análisis y auditoría de datos

#### 3.1. Muestreo

El análisis de las muestras tomadas cada día durante todo el rango de fechas en la base de datos no tienen una distribución uniforme. Dicha información resulta evidente en el histograma de frecuencias del muestreo.

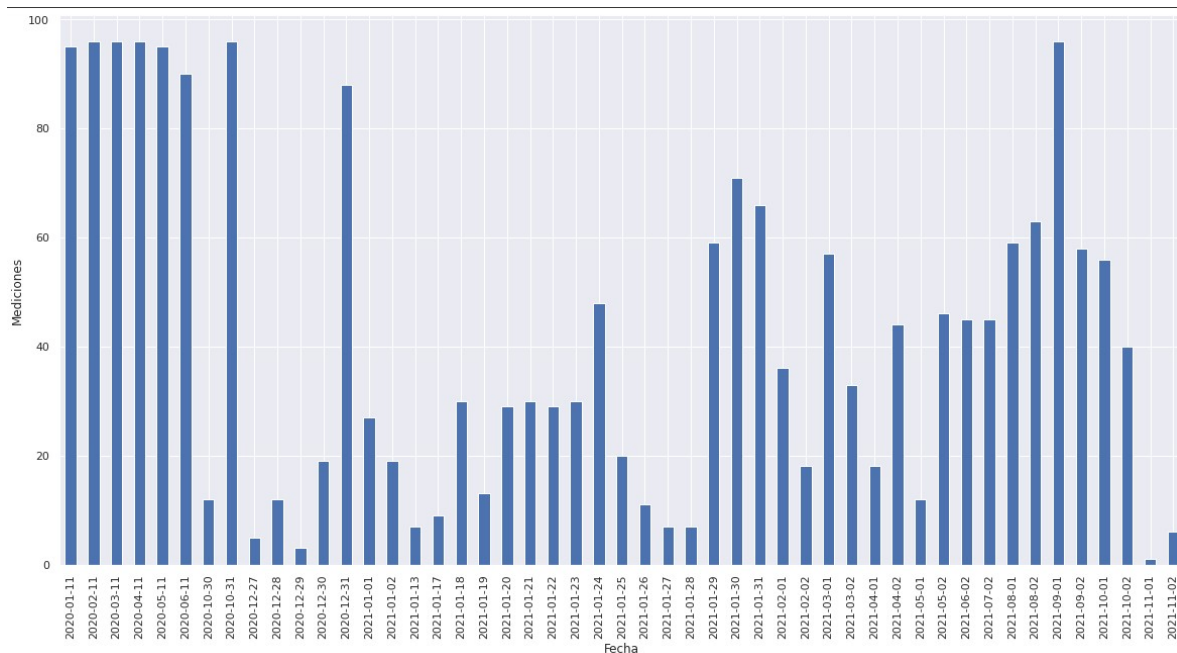


Figura 1: Histograma de frecuencia de muestreo.

	Muestras
Media	41.79
Mínimo	1
Máximo	96

Tabla 2: Caracterización del muestreo.

Se observa que la cantidad de muestras obtenidas del AMI tienen una alta variabilidad en el tiempo. Por que que es recomendable hacer un remuestreo de los datos para poder obtener información consistente desde esta base de datos.

#### 3.2. Tensiones por fase (V)

Analizando los datos de tensión por fase se obtuvo la siguiente caracterización:



	Tensión fase a	Tensión fase b	Tensión fase c
Cantidad de datos	2048	2048	2048
Media (V)	127.41767	127.35065	126.94504
Min (V)	123.58	123.51	123.13
Max (V)	130.33	130.07	129.76

Tabla 3: Caracterización de Tensiones por fase

De los datos expuestos en el cuadro 2 se observa a primera vista que no hay datos anómalos ya que el máximo y el mínimo están muy cerca a la media. Para confirmar se realiza el histograma de la tensión por cada fase.

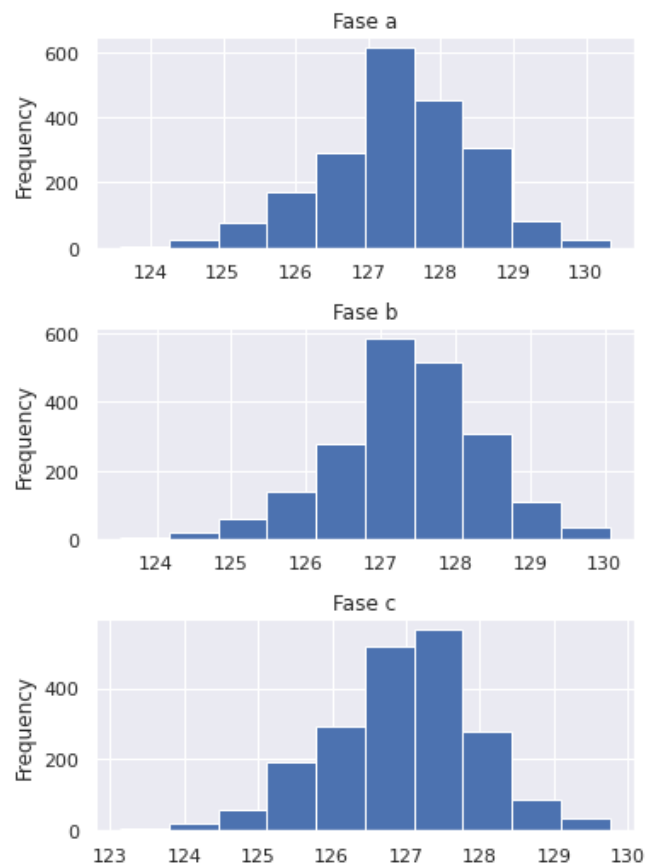


Figura 2: Histograma de los datos de tensiones por fase.

### 3.3. Corrientes por fase (A)

Al realizar la caracterización de las corrientes por fase se obtuvo lo siguiente:



	Fase A	Fase B	Fase C
Cantidad de datos	2048	2048	2048
Media (A)	5,54	4,85	4,27
Mínimo (A)	1,51	1,29	1,12
Máximo (A)	9,36	8,12	7,57

Tabla 4: Caracterización de las corrientes de fase.

A primera vista no parece haber datos anómalos, pues ningún dato se encuentra por fuera del rango de medición del equipo. Esto se confirma cuando se observa el histograma de los datos de corriente por fase:

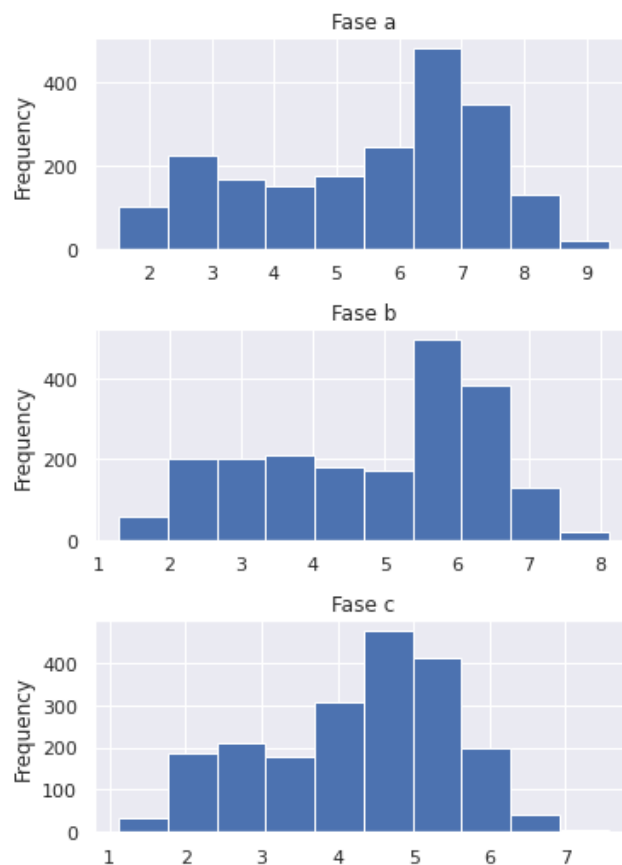


Figura 3: Histograma de corrientes por fase.



Se observa una agrupación de datos que no presenta una distribución anómala. Observando esto, se propone trabajar con estos datos, pues no se considera necesario realizar una limpieza de los mismos.

### 3.4. Factor de potencia

La caracterización de los datos de factor de potencia medidos por el equipo se muestra a continuación:

- **Cantidad de datos:** 2048.
- **Media:** 6,71.
- **Mínimo:** 0,004.
- **Máximo:** 65,53.

Se observa que hay datos anómalos en este conjunto de mediciones, pues tanto el máximo como el valor medio de las mediciones corresponden a un valor mayor a 1. Esto es aún más evidente cuando se observa el histograma de los datos del medidor:

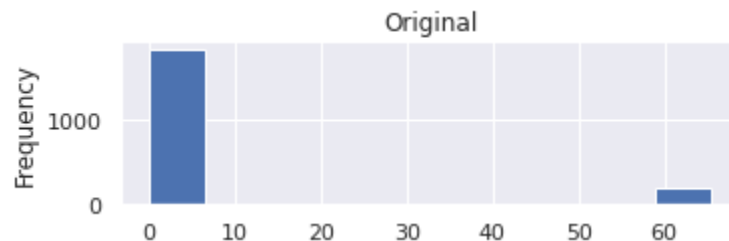


Figura 4: Histograma de los datos originales de factor de potencia.

En la figura 4 es evidente que hay una cantidad considerable de datos con valores cercanos a 60. Para mejorar la calidad de datos disponibles se sugieren 3 posibles acciones: Eliminar los datos mayores a 1, otorgarles un valor igual a cero o asignarles el valor de *NONE*. Un resumen de los resultados obtenidos al realizar estas acciones se muestra a continuación:

	Original	Eliminación	Reasignación a 0	Reasignación a NONE
<b>Cantidad de datos</b>	2048	1862	2048	1862
<b>Media</b>	6,71	0,89	0,81	0,89
<b>Mínimo</b>	0,004	0,004	0	0,004
<b>Máximo</b>	65,53	0,992	0,992	0,992

Tabla 5: Caracterización de los datos de factor de potencia originales y limpiados.

Una vez realizada la limpieza de los datos, se obtienen los siguientes histogramas:

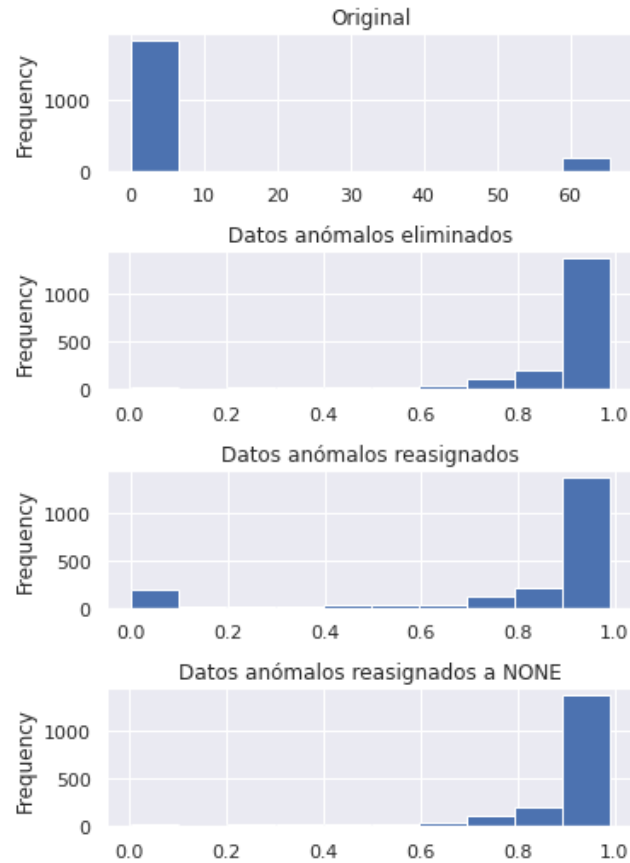


Figura 5: Histograma de los datos de factor de potencia limpiados.

Se observa que una vez limpiados los datos de factor de potencia, se obtiene una mejor distribución de los mismos. También, se puede concluir que es mejor eliminar los datos anómalos o asignarles un valor igual a *NONE*, pues no se insertan mediciones falsas iguales a cero.

### 3.5. Potencia (kW)

Al realizar la visualización de los datos de potencia a través de python, se obtuvo la siguiente caracterización de las mediciones dadas por el equipo:

- Cantidad de datos: 2048.
- Media (kW):  $3,9 \cdot 10^7$ .
- Mínimo (kW): 0,2.
- Máximo (kW):  $4,29 \cdot 10^8$





En este punto se notó que habían algunos datos anómalos, pues el máximo de potencia de los datos disponibles, corresponde a un valor por encima del rango de medición del equipo. Esto se confirmó al realizar el histograma de los datos disponibles:

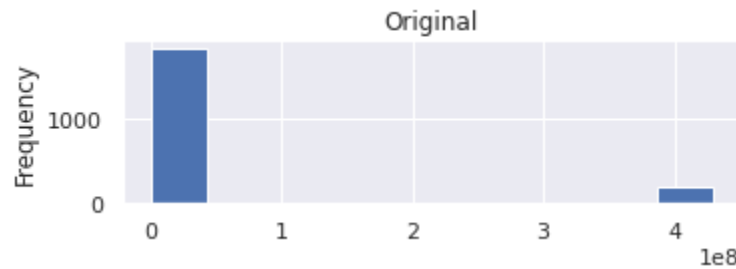


Figura 6: Histograma de los datos originales de potencia dados por el medidor.

Se puede observar en la parte derecha de la gráfica, una cantidad de datos cercano a los  $4 \cdot 10^8$  kW. Lo cual, como se mencionó anteriormente corresponden a datos anómalos. Para mejorar la calidad de los datos de potencia disponibles, se sugieren las siguientes acciones sobre los datos cuyo valor se encuentra por fuera del rango de medición del equipo: Eliminar estos valores, asignarles un valor de cero y otorgarles el valor de *NONE*. A continuación, se muestra un resumen de los resultados obtenidos con cada una de estas acciones y se comparan con los valores originales:

	Original	Eliminación	Resignación a 0	Reasignación a NONE
Cantidad de datos	2048	1862	2048	1862
Media (kW)	$3,9 \cdot 10^7$	180,65	164,24	180,65
Mínimo (kW)	0,2	0,2	0	0,2
Máximo (kW)	$4,29 \cdot 10^8$	287,6	287,6	287,6

Tabla 6: Características de los datos de potencia originales y los limpiados.

Se observa que cuando se eliminan o reasignan los valores anómalos, se obtiene un valor máximo de potencia de 287,6. Esto también altera considerablemente el valor medio de las mediciones y también mejor considerablemente la calidad de los datos. Lo anterior se confirma cuando se visualizan los nuevos histogramas de potencia:

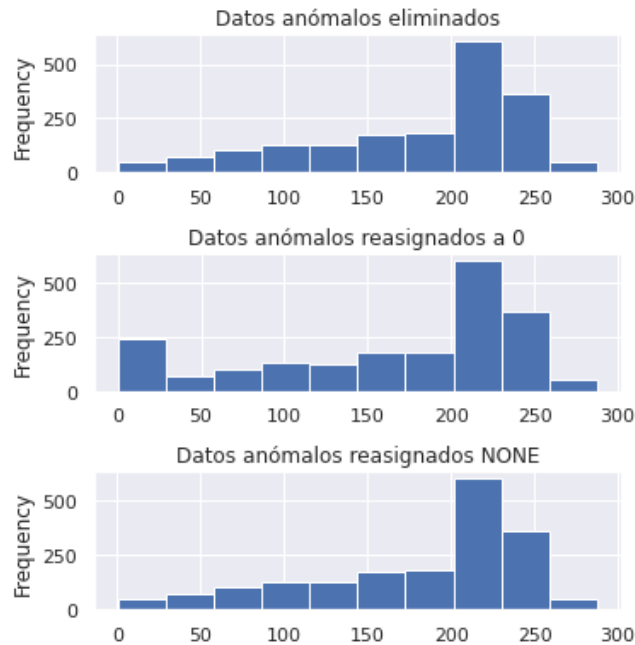


Figura 7: Histogramas de los datos de potencia limpiados.

Se observa que cuando se limpian los datos de potencia, se dispone de un conjunto con mejor calidad, pues los valores de potencia disponibles presentan una mejor distribución. También, se puede observar que para realizar la limpieza de estos datos, es mejor eliminar dichos datos o asignarles un valor de *NONE*, pues no se insertan valores falsos (mediciones iguales a cero).

### 3.6. Potencia por fase (kW)

La potencia por fase tiene también valores anómalos por encima de la precisión del dispositivo. Como ya se observó anteriormente, la mejor alternativa es eliminar estos valores o cambiarlos a tipo *NONE* o *NaN*. En la fase A este criterio funciona apropiadamente para obtener una distribución más acotada. Sin embargo, para las fases B y C, el panorama es diferente, como puede observarse en los histogramas adjuntos.

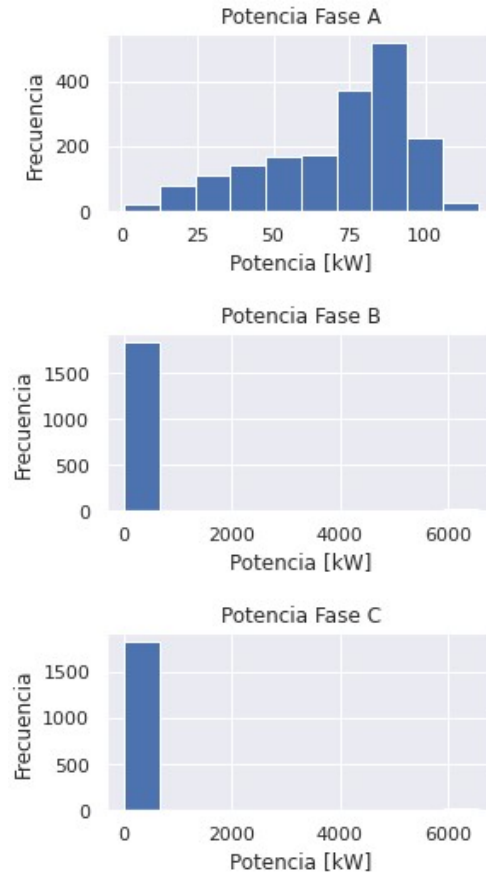


Figura 8: Histogramas de los datos de potencia por fase limpiados.

Los valores anómalos, por encima de 6000kW, en las fases B y C pueden ser resultado de altas potencias demandadas o de errores de medición. Debido a esta incertidumbre, no se eliminan estos datos dado que pueden representar variaciones relevantes en el estudio de la demanda de dichas fases en un análisis posterior.

	Potencia fase A [kW]	Potencia fase B [kW]	Potencia fase C [kW]
Media	71.1	707.1	715.6
Mínimo	0.9	0.0	0.1
Máximo	117.4	6553.1	6553.3

Tabla 7: Caracterización de la potencia por cada fase después de cambiar valores anónimos por NONE.



### 3.7. Energía Activa (kWh) y Energía Reactiva (kvar)

Al realizar la caracterización de los datos de Energía activa (kWh) y energía reactiva (kvar) se obtuvieron los siguientes resultados

	Energía (kwh)	Energía reactiva (kvar)
Cantidad de datos	2.048	2.048
Media	194.234	414.718
Mínimo	189.297	401.000
Máximo	198.158	426.000

Tabla 8: Caracterización de la energía activa y reactiva

A primera vista estos resultados se encuentran en el rango de medición del equipo, también los valores máximos y mínimos se encuentran muy cerca de la media. Para verificar se representan estos mismos en unos histogramas

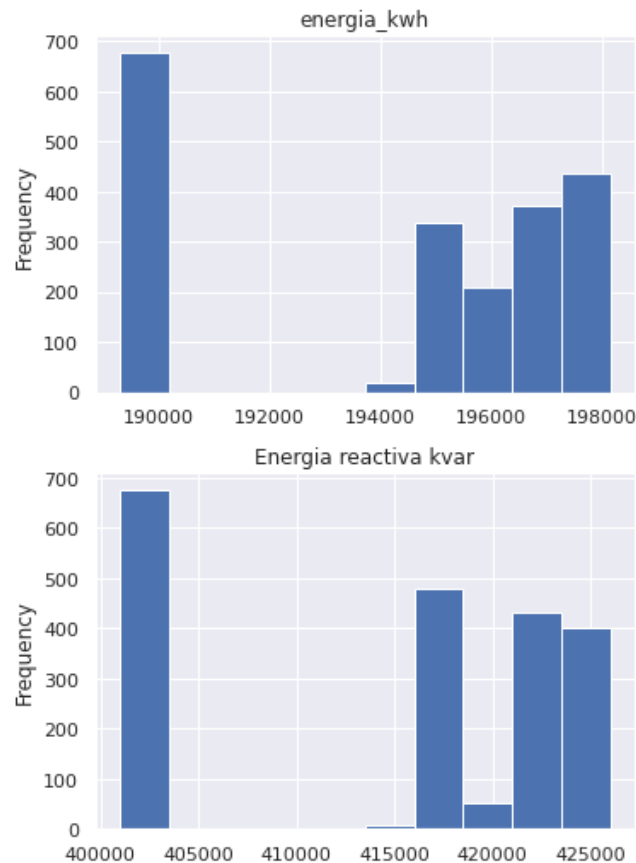


Figura 9: Histograma de la energía activa y reactiva.