

Escola Tècnica Superior d'Enginyeria Informàtica

Universitat Politècnica de València



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



etsinf

CONSUMO ENERGÉTICO

Proyecto I - Grado en
Ciencia de Datos

Autores

Gallego Andreu, Carlos | Cuberos Saus, Carles
Martínez Cabanes, Héctor | Oliver Belando, Daniel
Losa Brito, Alejandro | Santiago González, Jorge

Tutor

Carot Sierra, José Miguel

Curso 2019-2020

Presentación

Este trabajo ha sido desarrollado en la asignatura de *Proyecto 1: comprensión de datos*, del grado de Ciencia de Datos de la UPV, cuya finalidad es realizar un proyecto orientado a la descripción, visualización y comprensión de los datos de un problema real. En él, se aplican conocimientos adquiridos a lo largo del 1er curso del grado en Ciencia de Datos para trabajar con una base de datos real, en este caso sobre el consumo energético de la UPV.

Los datos con los que vamos a trabajar nos los ha proporcionado el Departamento de Medio Ambiente de la UPV, y recogen el consumo energético del Campus de Vera de varios meses. Se puede acceder a estos a través del siguiente enlace:

<https://www.dropbox.com/s/ubij5trdcu7l0n/energia.zip?dl=0>

Agradecimientos

En primer lugar, nos gustaría agradecer al profesor de la asignatura, José Miguel Carot Sierra, que nos ha supervisado durante todo el proyecto y nos ha aconsejado sobre cómo abordar muchas cuestiones.

También queremos destacar el papel del Departamento de Medio Ambiente de la UPV en todo este proyecto facilitándonos la base de datos sobre la que hemos trabajado. En especial mencionar a Álvaro y Cristina que nos explicaron cómo se organizaban los datos y cómo funcionaban todos los medidores.

Palabras claves

- Consumo energético
- Electricidad
- UPV

ÍNDICE:

1. [Introducción](#)
2. [Alcance del producto al servicio](#)
3. [Criterios de éxito del proyecto](#)
4. [Criterios de aceptación del producto](#)
5. [Objetivos](#)
6. [Programación y organización del proyecto](#)
7. [Estado del arte](#)
8. [Limpieza de la base de datos](#)
9. [Valores atípicos](#)
10. [Análisis](#)
 10. 1 [Dashboard](#)
11. [Conclusiones](#)

Introducción

La UPV es una institución de gran volumen, sin embargo, tiene un presupuesto y unos recursos limitados que se tienen que distribuir de manera óptima para satisfacer las necesidades de todas las personas, que de una forma u otra están relacionados con esta.

Dentro de estas necesidades una de las más importantes es la electricidad, ya que es imprescindible ya sea en funciones de climatización de edificios, de iluminación o energéticas para que funcionen todas las máquinas de los laboratorios. Es por esto, que una parte importante de presupuesto de la UPV se destina al consumo energético, en concreto, la universidad gasta aproximadamente 6.000.000€ de su presupuesto en consumo energético.

Sin embargo, ¿es todo este gasto necesario e imprescindible para el correcto funcionamiento de la UPV? Gracias a los datos recogidos por el Departamento de Medio Ambiente tenemos acceso al consumo energético del campus de Vera durante el ejercicio 2018-2019, con esto podemos ver cuánta energía consume cada edificio y cuando lo hace.

De esta forma, con este proyecto queremos analizar con detalle el consumo de gran parte de la UPV para determinar si este gasto energético puede ser más eficiente, y así rebajar esta parte del presupuesto, siempre sin perder de vista el objetivo primordial de la universidad, que es prestar sus servicios a todos sus integrantes cuando lo requieran.

Alcance del producto al servicio

Una vez el proyecto esté terminado, tendremos acceso al consumo de cada uno de los edificios del Campus de Vera de la Universitat Politècnica de Valencia. Podremos comparar estos consumos para determinar qué edificios son los que más consumen.

Además, de los edificios que más gasto generan tendremos acceso a su consumo/hora para evaluar cómo se reparte este gasto excesivo e intentar hacerlo más eficiente.

En definitiva, el resultado final de nuestro producto permitirá saber qué edificios de los campus que conforman la Universitat Politècnica de València son menos eficientes en cuanto a consumo energético se refieren y podremos proponer algunas soluciones para reducir el gasto durante las horas punta.

Criterios de éxito del proyecto

Los responsables del área medioambiental de la UPV en el Campus de Vera utilizarán nuestro análisis del consumo del Campus de Vera para hacer más eficiente el consumo de los edificios que más consumen y reducir el gasto general de consumo eléctrico del Campus de Vera.

Criterios de aceptación del producto

Con nuestro análisis se podrá acceder al gasto que produce cada edificio en un mes de manera sencilla, rápida y efectiva. Además, de los edificios que más consumen se tendrá acceso a su consumo por horas para evaluar porque gasta tanto en determinadas horas.

Nuestro producto permitirá ver el consumo tanto de manera visual a través de gráficos y mapas, como buscando a través de la base de datos.

Objetivos

El objetivo general del proyecto es realizar un análisis del consumo energético de la universidad. Entre otros factores, este análisis detectará los picos de consumo energético de la UPV, y servirá de soporte para la toma de decisiones enfocadas a reducir el mismo.

En base a este objetivo primordial, se derivan objetivos secundarios, que se corresponden con la descomposición del proyecto en tareas más sencillas y concretas. Estos objetivos secundarios son:

- Estudiar el consumo total de cada edificio de la UPV.
- Hacer un análisis exploratorio de los valores faltantes de la base de datos.
- Realizar una visualización de los datos resultantes para facilitar su comprensión.
- Ofrecer información que ayude a diseñar estrategias de reducción del consumo energético.

En los capítulos siguientes de la memoria se explicará todo el proceso de limpieza y reorganización de los datos, así como los métodos seguidos para conseguir una visualización gráfica eficiente de estos.

Programación y organización del proyecto

Para realizar este proyecto, el recurso más limitado que tenemos es el tiempo, por tanto, la organización y distribución de las tareas a realizar a lo largo del mismo es imprescindible para cumplir con las fechas de entrega marcadas. Para facilitar la distribución del tiempo, cada tarea se debe realizar en un número de sesiones determinado, por ejemplo, una tarea se empezaba la sesión del 28 de Febrero y tenía que estar acabada la sesión del 3 Marzo. Además, el responsable de cada tarea tiene libertad para trabajar en ese margen de tiempo. Las tareas a realizar en nuestro proyecto son las siguientes (en orden cronológico):

- Definir alcance, objetivos, criterios de éxito y planificación del proyecto (3 sesiones)
- Limpiar bases de datos (4 sesiones)
- Reordenar las bases de datos (4 sesiones)
- Explicar y describir el proceso seguido en la limpieza de los datos (2 sesiones)

- Preparar la segunda presentación del proyecto (2 sesiones)
- Analizar el consumo energético por edificio (3 sesiones)
- Analizar el consumo por horas, en edificios seleccionados (4 sesiones)
- Realizar la visualización y explicación del análisis y sus conclusiones (6 sesiones)
- Grabar el video explicativo del proyecto (7 sesiones)
- Preparar la presentación final (5 sesiones)
- Memoria entregable (se irá realizando conforme avance el proyecto)

Toda esta planificación está recogida en un diagrama de Gantt, de forma que sea más visible saber cuando empieza y termina cada actividad. Este diagrama de Gantt está disponible en el Anexo de esta memoria.

Además, para prevenir posibles contratiempos sobre todo al final del proyecto hay un margen de casi 2 semanas entre el final de la última tarea programada y la última entrega del proyecto. Esto nos permitirá poder trabajar con más calma si alguna tarea sufre un pequeño retraso.

Por otra parte, para mejorar la comunicación y el compromiso de todos los miembros del equipo hemos puesto algunas normas de grupo:

- Cumplir con los plazos asignados a cada miembro que forma parte del equipo
- Hacer una reunión exponiendo lo que tenemos cada semana
- Comunicación entre todos los participantes para lograr un objetivo común
- Si algún miembro no puede realizar la tarea que se le ha asignado, avisar con un mínimo de 3 días de antelación para poder tomar una decisión que no afecte a la programación del proyecto.

Incidencias y variaciones:

- Debido a la pandemia del COVID-19 durante la realización del proyecto, se acordó no realizar de forma obligatoria el vídeo. De esta forma, las sesiones planificadas a la grabación de este se dedicarán tanto al análisis de los datos, como a la realización de la presentación final. Si más adelante podemos grabar el vídeo de forma más eficiente, valoraremos esta posibilidad.
- Nuestra idea era ir trabajando los 6 en todas las actividades, dividiendo éstas en tareas más sencillas y en muchos casos doblando estas tareas para comparar resultados y minimizar errores. Como durante el proyecto no vamos a poder trabajar los 6 de forma presencial, hemos decidido trabajar por parejas en las actividades planificadas en el diagrama de Gantt, ya que es más fácil la comunicación constante entre 2 persona que entre 6. Esta decisión, unida a la anterior relacionada con la grabación del vídeo, ha hecho

que las actividades se realicen con bastante antelación a su plazo estimado y nos ha permitido avanzar de una manera más rápida en la limpieza, análisis y visualización de los datos.

Estado del Arte

Antes de comenzar a tratar con los datos, hemos realizado una búsqueda exhaustiva acerca del tema en cuestión para familiarizarnos y de esta manera tener la capacidad de adquirir información para poder extrapolar conclusiones más efectivas. Es decir, a través de los datos obtendremos información y gracias a esta búsqueda podremos transformarla en conocimiento mediante comparaciones y el uso de la razón, esta función sólo la puede realizar el ser humano.

A continuación, explicaremos lo más importante, a modo de resumen, y dejaremos en los anexos toda la información detallada que hemos encontrado.

En primer lugar, existen una serie de factores que influyen en el precio de la electricidad. Podríamos considerarlo como una cadena donde todo funciona en base al cliente, y se benefician de su facturación las compañías comercializadoras, compañías generadoras de electricidad, personal de administración y compañías distribuidoras.

No obstante, la factura eléctrica no cubre algunos costes como los derivados del agotamiento de recursos energéticos no renovables, o los de neutralización de impactos medioambientales como el cambio climático, el cual es un punto clave en el proyecto a tratar.

En segundo lugar, nos hemos centrado en el ámbito económico. Para ser más exactos, en qué gasta el presupuesto la UPV. La primera conclusión es que el presupuesto de gastos de mantenimiento viene resultando insuficiente, ya que en los últimos años ha superado al presupuesto inicial. Es decir, en 2017 el presupuesto inicial fue de 4.311.587 euros y el gasto liquidado de 4.839.664. En 2018 la UPV partió de 4.611.587 euros y la previsión de cierre fue de 5.011.587.

Para 2019 se propone destinar 5.046.641 euros. Esta cantidad va destinada en gran parte a los siguientes contratos: eléctrico (1.011.456 euros), climatización (1.568.548 euros), ascensores (151.016 euros), incendios (213.796 euros), asistencia técnica (81.360 euros), Winton (27.921 euros), Sik (23.424 euros), EDAR (11.616 euros). Por otra parte, se destinará a Voz y Datos (83.500 euros), excesos sobre franquicias y sobrevenidos contratos (401.953 euros), Mantenimiento general (1.457.050 euros) y revisiones técnico legales (15.000 euros).

En tercer lugar, nos interesa saber cómo controlamos la energía que consume la UPV. En este caso, se logra medir gracias a un nuevo sistema de monitorización energética en edificios que permite conocer en tiempo real y con total precisión su consumo de energía. Con este sistema de medida se puede determinar la causa de un excesivo gasto de consumo y ayudar a revertirlo.

Se trata de una serie de sensores inalámbricos que recogen diferentes parámetros ambientales, tales como la temperatura de la zona, la humedad relativa, la velocidad del aire, la luminosidad o la corriente del aire. Además, recoge también valores como la potencia que se consume en cada una de las aulas y pasillos. Todo esto es enviado a un servidor central y posteriormente, mediante un sistema de auditoría energética se obtiene el diagnóstico respecto a la eficiencia energética del edificio.

En último lugar, nos hemos informado sobre el precio del kilovatio-hora en España. Desde el año 2014, el precio del kWh del mercado regulado es establecido a través del mercado mayorista. Por su parte, el precio del kWh en el mercado libre es establecido por cada compañía. El mercado regulado de luz ofrece a sus clientes la tarifa de Precio Voluntario al Pequeño Consumidor (PVPC) y su precio es establecido por el mercado mayorista en función de la demanda del suministro eléctrico. En el caso del mercado libre existe la opción con discriminación horaria. Un ejemplo puede ser la Tarifa Plan Noche Iberdrola, en la que el kWh cuesta 0.1648 euros en hora punta y 0.0827 euros en hora valle. También ocurre en el mercado mayorista y es continuamente revisado por el Gobierno.

En conclusión, el precio de consumidores residenciales es más elevado que el de industrias, ya que en estas el precio es diferente debido al aumento del consumo. Gracias al proyecto, podremos identificar estas horas punta con la finalidad de que la UPV pueda ajustar su tarifa en función al consumo.

Limpieza de la base de datos

Debido a la dimensión de nuestra base de datos es fundamental la tarea dedicada a limpieza y tratamiento de los datos, ya que un solo valor erróneo puede llevarnos a conclusiones equivocadas al final del análisis. Por ello, hemos dedicado una gran parte del tiempo del que disponíamos a la limpieza de valores anómalos y atípicos del consumo de la UPV. A continuación, vamos a explicar brevemente qué tipos de valores hemos considerado oportunos eliminar y cuales pensamos que es mejor mantener porque son representativos del consumo real, la explicación más detallada junto con el código utilizado en Python y los gráficos hechos con Statgraphics o Tableau están disponibles en el Anexo.

Antes de empezar con los datos que hemos eliminado es importante explicar las variables que influyen en el consumo en nuestra base de datos. En primer lugar, la columna KWh nos indica el consumo acumulado de un edificio en un intervalo de tiempo de 15 minutos, es decir, si el primer valor que tenemos de un edificio es 100.000 KWh, todos los siguientes valores deben ser mayores o iguales que este, ya que a 100.000 se va añadiendo el consumo cada 15 minutos. Además, el consumo de un edificio no está determinado por un medidor en concreto si no por la relación entre varios medidores, por lo que si un medidor deja de medir durante un período de tiempo no nos encontramos con valores ausentes, si no con un valor que no es el consumo real (este valor puede ser un valor muy grande o incluso un valor negativo).

Para solucionar este problema de consumo acumulado, creamos una nueva columna que calculaba el consumo neto cada 15 minutos, ya que nuestro objetivo es calcular el consumo total de cada edificio y los valores acumulados nos dificultan mucho esta labor.

Finalmente, hay que distinguir entre 2 tipos de datos diferentes, los que miden el consumo general y los que miden solo los KWh destinados a la climatización de los edificios. Además, dentro de los datos de consumo general ya están incluidos los destinados a la climatización, por lo que, para alcanzar los objetivos propuestos al principio del proyecto, los datos que marcan el consumo de la climatización de edificios no nos son útiles.

Una vez explicadas las variables que influyen en el consumo energético de la universidad, pasamos a explicar que tipo de datos hemos eliminado y porque:

- **Faltantes:** Los valores faltantes (NaN), es decir, aquellos de los que no se tiene ningún registro en la columna del consumo acumulado los hemos eliminado ya que no nos aportan información alguna y afectan a la nueva columna creada para calcular el consumo neto cada 15 minutos. La proporción de filas que hemos eliminado a causa de estos valores es inferior al 2% de los datos, por lo que la muestra que se nos queda sigue siendo muy representativa.
- **Clima:** Los datos que nos proporcionaban información sobre el consumo empleado en la climatización de los edificios los hemos eliminado porque para calcular el consumo total de un edificio no nos son de utilidad, y mantenerlos podía llevar a confusiones a la hora de medir el consumo neto. De los datos que hemos eliminado, los de este tipo son los más abundantes, con una media de 150.000 casos por mes.
- **Mes siguiente:** El primer dato sobre el consumo de cada mes, aparece tanto en la base de datos del mes anterior como en la del propio mes. Es decir, este dato aparece duplicado y hemos eliminado el caso que corresponde al mes anterior para solucionar este problema. Por ejemplo, el consumo medido el día 1/12/2018 00:00h aparece tanto en la base de datos de diciembre como de noviembre, por lo que hemos eliminado este caso de la base de datos de noviembre para solucionar este problema de duplicación de datos.
- **Negativos (Consumo acumulado):** Como hemos explicado en los párrafos anteriores, los datos de esta columna deben ir aumentando progresivamente ya que nos indica el consumo acumulado, por lo que cualquier valor menor o igual a 0 en esta columna se debe a un error en los medidores o en la recogida de los datos, por lo que es necesario eliminarlos.
- **Negativos (Consumo neto):** En la nueva columna que hemos creado después de eliminar los datos anteriores, si encontramos algún valor negativo quiere decir que el consumo acumulado era erróneo puesto que ha disminuido de un valor al siguiente. En este caso, seguimos hablando de un error en los medidores y hemos considerado que estos casos deben ser eliminados ya que nos dan una información errónea.

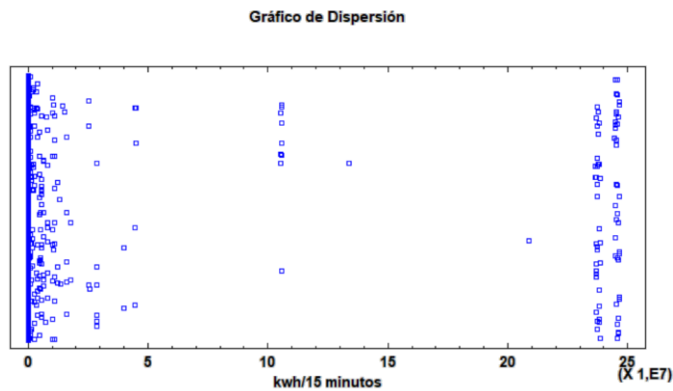
- Nulos: Datos cuyo valor en la casilla correspondiente a la columna KWH es 0. en esta columna partimos de un número (mayor que 0) y lo siguientes son iguales (si no se consume nada) o mayores (si consume). Luego, un valor 0 en esta columna es un dato incorrecto.
- Nulos (Consumo neto): Estos datos no son en su totalidad datos anómalos. No obstante, no podemos distinguir entre aquellos que son correctos y que significan que no se ha consumido nada en esos 15 minutos y entre aquellos incorrectos que aparecen como nulos por fallo de los sensores. La eliminación de éstos no influye en la representación del consumo total.
- Atípicos: Finalmente, una vez todos los datos de los que disponemos son, en teoría, correctos, debemos analizar si alguno de los datos no es posible que sea representativo porque excede demasiado en consumo respecto de los demás. La eliminación de estos valores ha ido dependiendo según el mes analizado, ya que dependía del funcionamiento que habían tenido los sensores durante el mismo. Al ser esta categoría más larga y costosa de explicar, la siguiente sección de la memoria está dedicada a el proceso de limpieza seguido para este tipo de datos en concreto.

Valores Atípicos

Como se ha explicado en el apartado anterior referido a la limpieza y calidad de los datos, el tratamiento de valores atípicos ha sido el proceso más largo y costoso del proyecto hasta la fecha. Esto se debe a que son bastante comunes los errores en la medición del consumo y nos encontramos con datos de consumo neto en un período de 15 minutos desorbitados. Además, podemos diferenciar 2 fases en este tratamiento de valores atípicos.

En primer lugar, entre Septiembre de 2018 y Febrero de 2019 nos encontramos con valores atípicos mucho más grandes y en mucha más cantidad que en los siguientes 6 meses. En principio, esto puede deberse a la mejora de los sensores de medición, aunque no tenemos evidencias de ello. Por esto, en los datos comprendidos en este período de tiempo hemos seguido un procedimiento diferente para eliminar dichos valores, que el seguido para los otros meses.

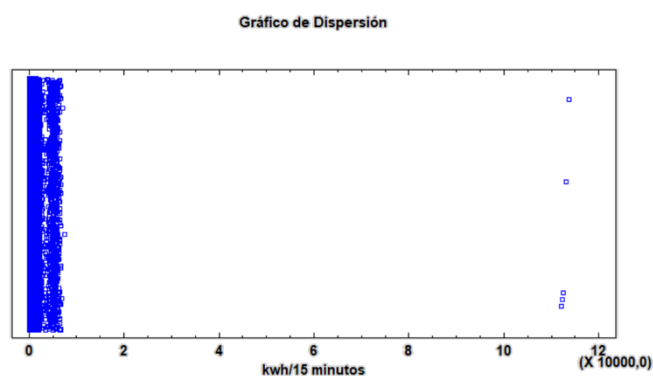
Para estos primeros meses, nos hemos basado en gráficos de aberrantes y en la diferencia entre los parámetros media aritmética y mediana. Si tanto el gráfico como la diferencias entre ambos parámetros nos mostraban la presencia de valores atípicos eliminamos los valores que se encuentran fuera del intervalo $[media \pm 2 \text{ desviaciones típicas}]$. A continuación, mostramos el resumen estadístico del mes de Septiembre antes de eliminar valores atípicos:



Resumen Estadístico para kwh/15 minutos	
Recuento	211688
Promedio	94045,6
Mediana	60,0
Media Recortada 5%	119,155
Media Winsorizada 5%	144,74
Desviación Estándar	4,527E6
DAM	40,0
Mínimo	10,0
Máximo	2,46587E8
Rango	2,46587E8
Sesgo Estandarizado	236058,
Curtosis Estandarizada	259334,

Como podemos observar en los gráficos anteriores, hay valores de consumo que son imposibles de consumir en 15 min. De hecho, algunos valores son tan grandes, que incluso después de eliminarlos con la técnica especificada en el párrafo anterior, la diferencias entre media y mediana seguía siendo muy grandes y se ha tenido que repetir el proceso varias veces.

El resultado final de esta limpieza de datos atípicos, siguiendo con el ejemplo del mes de Septiembre es el siguiente:



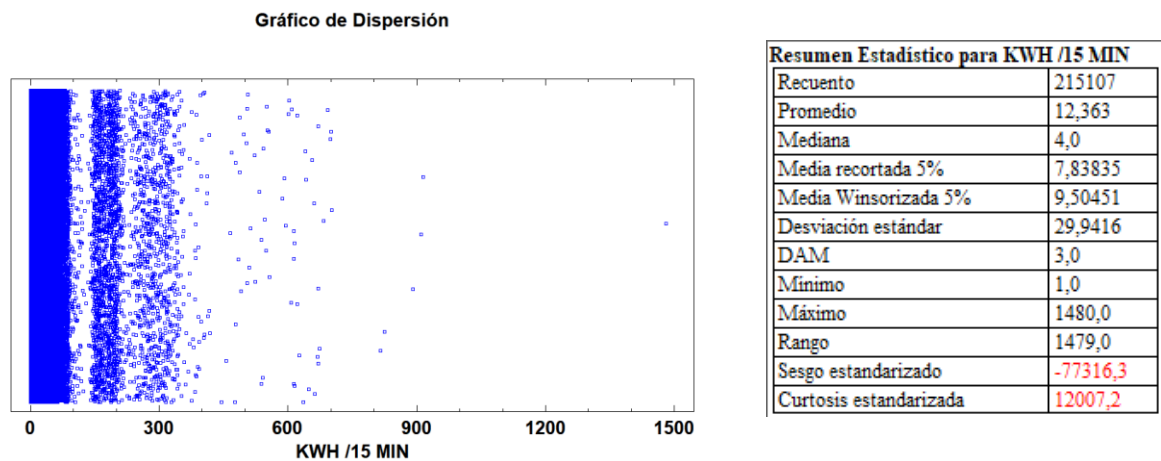
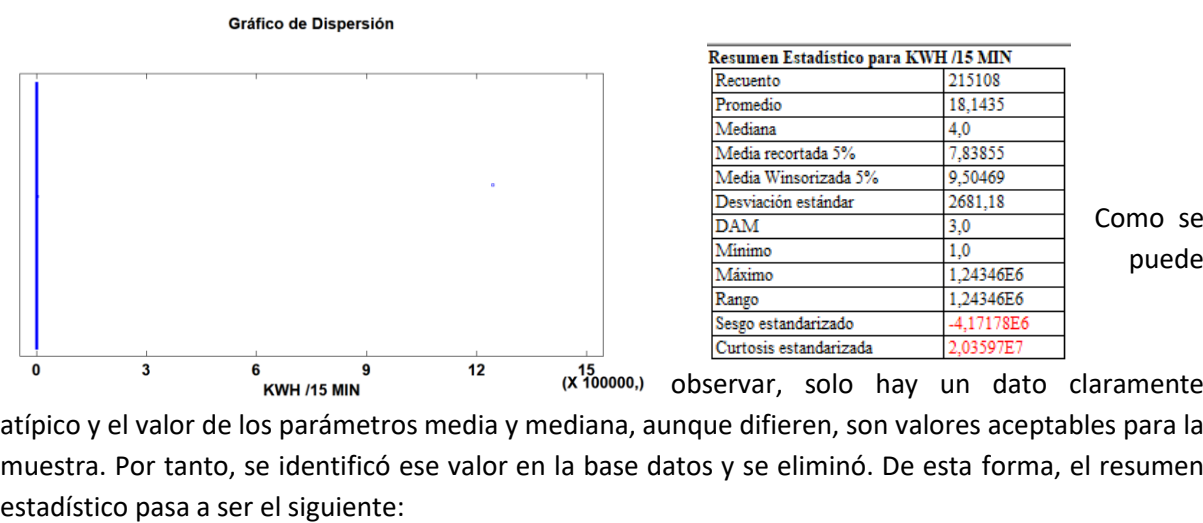
Resumen Estadístico para kwh/15 minutos	
Recuento	211446
Promedio	184,75
Mediana	60,0
Media Recortada 5%	118,401
Media Winsorizada 5%	143,561
Desviación Estándar	696,182
DAM	40,0
Mínimo	10,0
Máximo	113620,
Rango	113610,
Sesgo Estandarizado	485958,
Curtosis Estandarizada	1,51854E6

Los resultados de la tabla se han de dividir entre 10, puesto que el formato .0 presente en algunos meses causaba una multiplicación x10 de cada dato en Statgraphics.

En el anexo se especifica el proceso seguido en cada mes y los resúmenes estadísticos de estos, así como el código Python y los gráficos utilizados en la eliminación de valores atípicos. También se explica el criterio seguido para saber si un valor relativamente alto es un error de medición o es representativo de la muestra (como, por ejemplo, los 5 datos más separados en el gráfico de dispersión anterior).

Por otra parte, para los datos pertenecientes al intervalo Marzo-Agosto de 2019 se ha seguido un proceso similar aunque ligeramente diferente al explicado anteriormente ya que el número de valores atípicos es casi inexistente.

En este caso, se han identificado los datos atípicos mediante gráficos de aberrantes, como el recuento de estos datos era muy pequeño, se optó por buscar estos valores en la base de datos, confirmar su existencia y a que edificio pertenecían, y eliminarlos directamente. Por ejemplo, en el mes de Abril el resumen estadístico obtenido antes de la eliminación de valores atípicos es el siguiente:



Aunque algunos valores se alejan del consumo medio, no pueden ser considerados atípicos porque pueden ser representativos de la muestra. En el anexo, se especifica porque un valor relativamente elevado no lo consideramos atípico, y por tanto, lo mantenemos en la base de datos. Además, en este segundo intervalo, se especifica qué valores hemos eliminado y a que edificio pertenecían.

Una vez realizado todo este proceso de calidad de los datos, podemos pasar a la fase de análisis y visualización, que es donde vamos a poder sacar información valiosa para alcanzar la misión y los objetivos de proyecto.

ANÁLISIS

Esta fase es, posiblemente, la más determinante en base a cumplir los objetivos del proyecto, es por eso, que también es una de las fases a las que más recursos se le ha dedicado.

En este proyecto el equipo se ha centrado en tres puntos clave, con subíndices que hacen referencia al nombre de análisis hecho en el software Tableau. Con estos el equipo intenta lograr sus objetivos, con tal de abarcar todos los focos de interés y de poder tomar decisiones acertadas. Estos tres puntos son:

1. Identificar el gasto general en un periodo de tiempo.
 - a. Gasto Horas
 - b. Gasto Días
2. Descubrir los edificios que más gastos tienen diariamente.
 - c. Consumo
3. Señalar los sensores/edificios que pueden ser de interés debido a su inusual valor
 - d. Faltantes
 - e. Registros

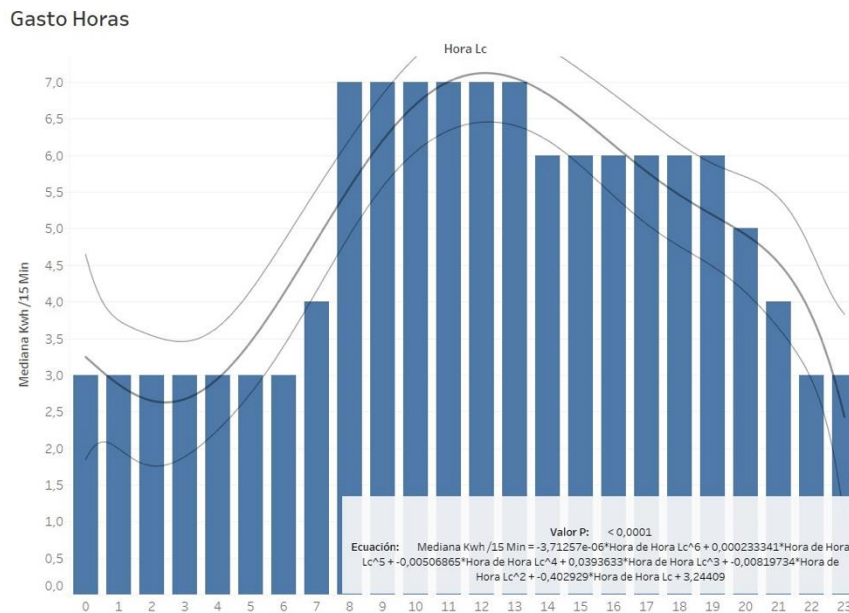
Antes de pasar a visualizar los datos, hay que mencionar que en este documento solo aparecerán los análisis más destacados y que nos aportan más información a modo de ejemplo. El resto de las visualizaciones y de los análisis realizados estarán disponibles en el Anexo correspondiente a dicha parte.

Gasto Horas

En este punto se relaciona el gasto en KWH cada 15 min con el tiempo en horas en un diagrama de barras. Se muestran las 24 horas del día en el eje X mientras que en el eje Y aparece la mediana del gasto global (no es de un edificio en específico). Se utiliza la mediana, dado que hay gastos atípicos/extremos en la base datos y con la media se obtendría un pico de gasto a las 2 de la madrugada.

En cuanto a la distribución, se puede apreciar como el gasto desde las 22:00h hasta las 5h es ínfimo, por debajo de la mediana que es alrededor de 6, dependiendo del mes. Mientras que desde las 6h hasta 13h-14h hay un aumento progresivo de este gasto llegando al punto más alto de esta gráfica, que normalmente es al mediodía. A partir de este tramo, empieza a disminuir de forma escalonada desde las 14h hasta las 21h que ya se estanca el gasto como hemos señalado antes. Teniendo en cuenta estos rasgos, podemos afirmar en gran medida que se sigue una distribución normal.

Importante dato sobre este análisis: en la mayoría de los meses, se puede describir la forma de este gráfico a través de una función polinómica. Dependiendo del mes, será de mayor grado debido a sus irregulares datos o de menos si el gasto crece y decrece de manera regular.



En este ejemplo, vemos como la distribución es bastante irregular y por eso la parábola no es muy exacta. No obstante, el consumo sigue la tendencia general en todos los meses como hemos explicado en el párrafo anterior.

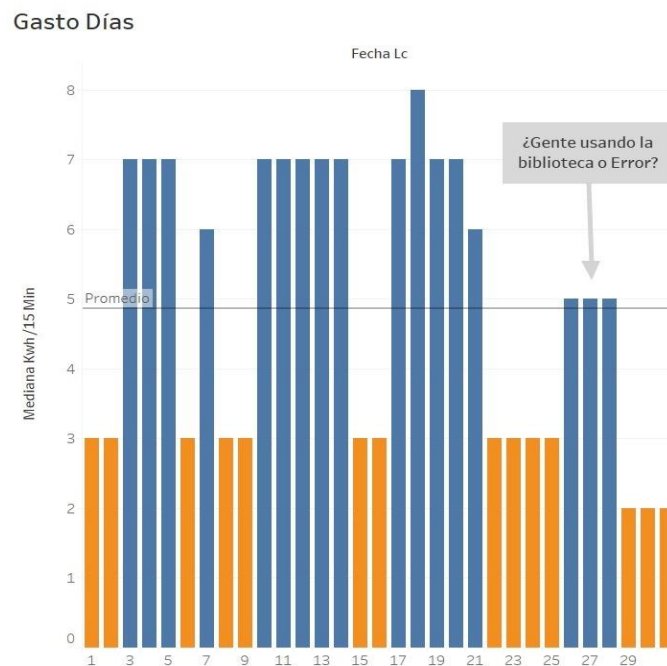
Gasto Días

Este análisis sigue los patrones del anterior punto. Únicamente cambia el parámetro del tiempo que pasa de ser en horas a ser en función de días, teniendo así todos los días del mes en el eje X. El eje Y sigue siendo de gasto, pero esta vez es el sumatorio de los gastos que tienen todos los edificios en función del día.

Es importante usar el sumatorio en este gráfico de barras, ya que la información que queremos extraer de este no es la distribución que sigue como ocurría en el anterior análisis, sino que queremos averiguar qué días son los que más y menos gasto acumulan y conocer la causa: fin de semana, festivos, entre semana, anómalo...

Para hacer más fácil la lectura de este, hemos utilizado la herramienta de clúster con la que conseguimos distinguir los grupos o conjuntos de datos que se comportan de manera similar. Se pueden observar dos grupos en general: el referente a días lectivos (clúster azul) y el de los días festivos, ya sean fines de semana o fiesta (clúster naranja).

Con tal de explotar lo máximo el rendimiento de este análisis, se ha añadido un detalle que aporta mucha información: el promedio. En todos los meses podemos saber cuál es el gasto medio de los días festivos sin tener en cuenta los días lectivos y viceversa, al igual que el gasto medio de todo el mes.



Este gráfico se corresponde al consumo por días en el mes de Diciembre, el dato más curioso lo tenemos en los días 26,27,28 en los cuales el consumo se asemeja a un día normal pero esos días solo estaban abiertas las bibliotecas. Por tanto, o las bibliotecas tuvieron un consumo eléctrico excesivo o se trata de algún fallo en las bases de datos.

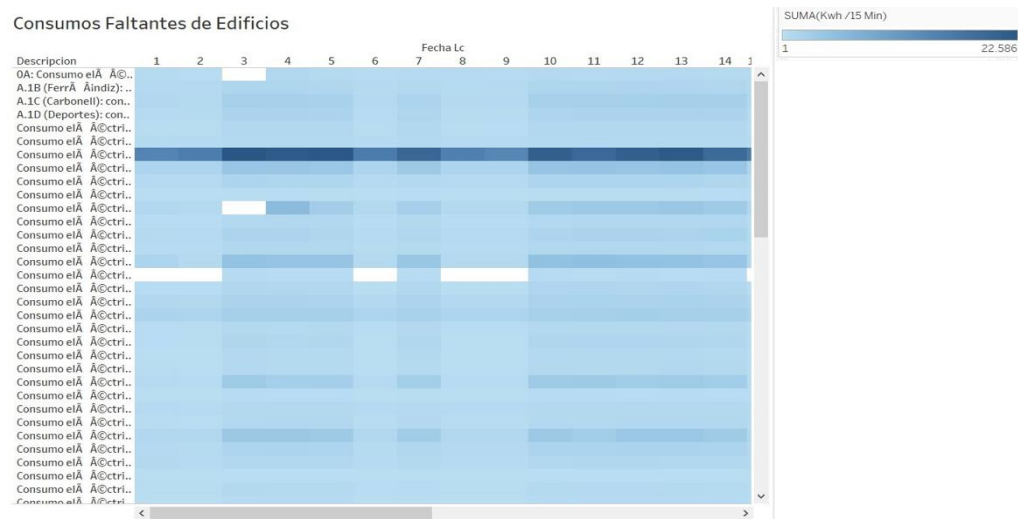
Faltantes

El siguiente estudio es uno de los más interesantes y a la vez el más simple de todos. Esto se debe a que únicamente se relacionan los días del mes con las descripciones (nombre) de los diferentes edificios, además de tener como 'marca' la suma del gasto kwh/15min de cada día, en función del edificio, con esto se consigue mapear el mes y el gasto de los edificios.

Este mapa resalta los días en los que un edificio no tiene ningún gasto; información muy relevante en cuanto a fallos de sensores y errores en el sistema. También destaca cuando hay menor gasto (color claro) hasta cuando hay mayor (color oscuro).

Cabe destacar del mapa, su aportación en relación con no solo detectar cuándo hay un error o datos erróneos, sino que, asimismo, se obtienen hipótesis de a qué se deben estos errores. Esto es posible debido al uso del parámetro 'Día' que ayuda a entender si se debe el error a la fecha (el sensor falla en un día en concreto) o se trata de fallos aislados.

La clave de su utilidad está en el uso de señales visuales, en este caso una gama de azules. Esto consigue que sea un análisis sencillo de entender y en el que consigues de manera muy rápida información, sin necesidad de entender o de enfocar la atención en este.

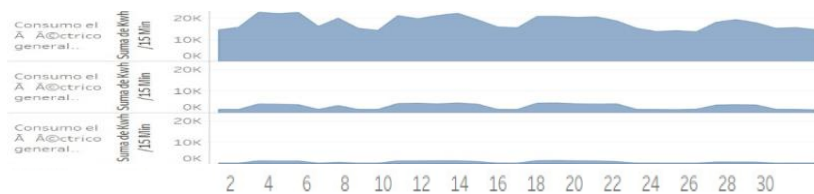


En este gráfico llama la atención los 7 u 8 paneles blancos que nos indican que ese día un edificio no consumió ningún Kwh, esto puede deberse a que los sensores no midieron el consumo ese día o que midieron mal y esos datos fueron eliminados durante el proceso de limpieza de la base de datos. Además, es interesante observar que, en general, el consumo de cada edificio se mantiene constante a lo largo del mes (misma gama de azul), es decir, un edificio consume lo mismo todos los días, por lo que es más fácil detectar situaciones atípicas para detectar el consumo por días.

Consumo

Respecto a la evaluación del consumo, se han utilizado los días del mes, el nombre de cada edificio (['Descripción']) y la suma total del gasto Kwh/15 min. De esta manera, se obtienen tantos gráficos de área como edificios se disponen en la base de datos, además de señalar el gasto total que tiene cada uno de los días del mes.

El principal aspecto positivo del estudio del consumo es tener la capacidad de averiguar qué edificios tienen gastos extremadamente altos o, por el contrario, de que este gasto sea ínfimo en alguno de los edificios estudiados, y con los gráficos obtenidos se visualiza fácilmente. También cabe destacar la importante función que aporta el software Tableau en este análisis, el cual da la posibilidad de excluir los edificios cuyo gasto entre dentro de lo común para los diferentes parámetros que tenemos en cuenta, de tal forma que sólo quedarían los edificios de los cuales realmente se tienen dudas sobre si están bien medidos sus gastos, tienen algún tramo sin gasto alguno u otro rasgo del que desconfiar. Por el contrario, también existe un factor que convierte a este estudio un quebradero de cabeza. Este factor es la cantidad de gráficos que se crean, ya que hay un importante número de edificios en este estudio. Es por ello que no se pueden visualizar todos de una forma relativamente rápida ni tampoco identificar rápidamente los detalles señalados anteriormente.



Al igual que en el gráfico anterior, el consumo es bastante constante todos los días, con pequeños bajones que suelen coincidir con fines de semana o días festivos, en los que obviamente, el consumo es inferior a los días lectivos. Cuando un edificio no tiene datos de consumo en un día aparece un vacío en la gráfica que nos lo indica.

Registros

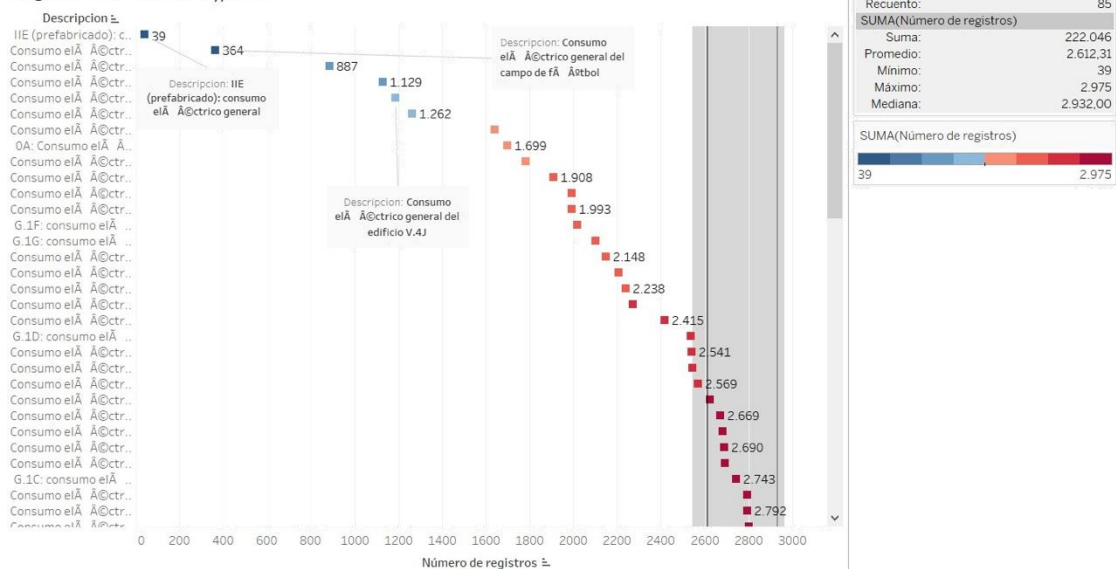
Por último, el análisis de los registros se enfoca en la cantidad de registros que tiene cada edificio. Para ellos se hace uso tanto del nombre de cada uno de los edificios como de la cantidad total de registros que tiene este a lo largo del mes, todo esto en un diagrama de cuadrados.

Al igual que en el anterior estudio, no se puede visualizar todo fácilmente pero no se puede señalar como defecto, ya que este diagrama es muy limpio y, al estar ordenado de menor a mayor, se pueden detectar anomalías rápidamente.

Una de las herramientas utilizadas para llevar a cabo el análisis es la 'Línea de promedio' que resalta cuál es la media de registros; también se hace uso de la 'Mediana con Cuartiles' de tal forma que se puede observar la desviación que puede llegar a tener la cifra de registros y comparar la media y la mediana con el fin de averiguar si hay valores atípicos o aberrantes.

Con la intención de ayudar visualmente, se ha añadido un filtro que resalta cada uno de los cuadrados, pasando desde un azul, que hace referencia a la poca cantidad de registros, hasta un color rojo oscuro, que indica la gran cantidad de registros que tiene el edificio en cuestión.

Registros de Edificios // Mes



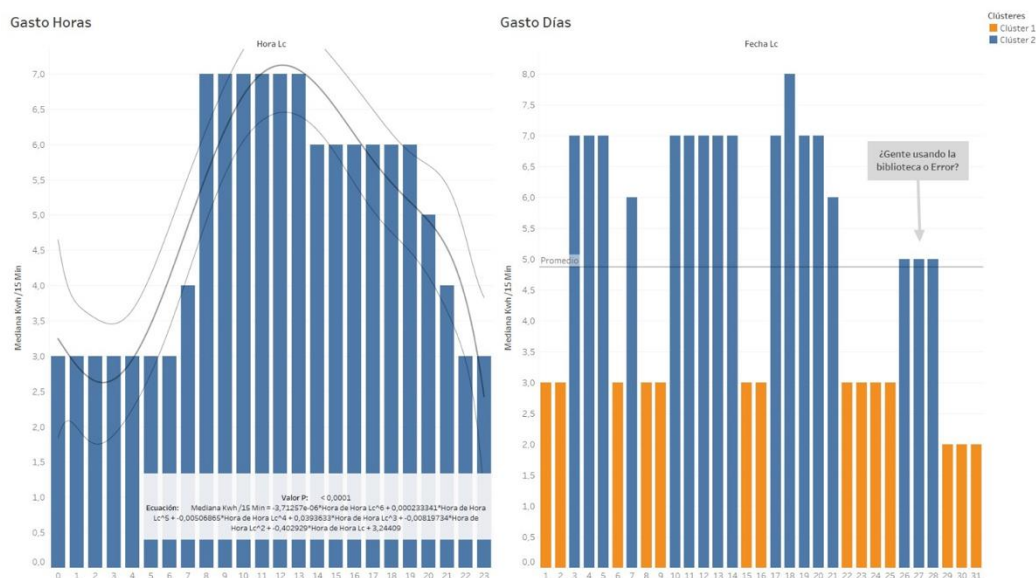
De este modo, observamos como hay algunos edificios que tienen muy pocos registros por mes. Aunque los 2 edificios que tienen menos registros, no son edificios como tal, sino que se corresponde con el Consumo Eléctrico General y con el Alumbrado del Campo de Fútbol, de forma que es posible que haya alguna relación entre esto y que ambos tengan pocos registros en comparación con los edificios de la UPV.

Dashboard

Teniendo como objetivo simplificar y crear un medio lo más visual posible, el equipo ha creado dos cuadros de mandos con los anteriores estudios explicados:

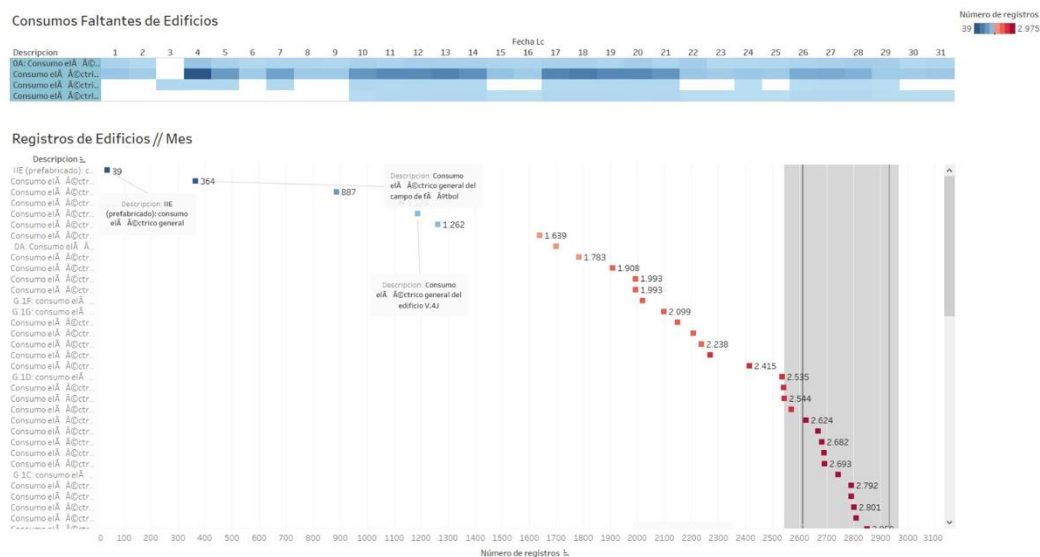
1. Diario - Mensual

Se utilizan los estudios de 'Gasto Horas' y 'Gasto Días' para tener la capacidad de observar la distribución que sigue el gasto a diario y mensualmente.



2. Relación Registros/Faltantes

Para este cuadro de mandos, se hace uso de las hojas de trabajo 'Faltantes' y 'Registros' porque de esta manera, gracias a la funcionalidad que se ha destacado en la que se puede excluir los edificios que no resultan interesantes para este estudio, se pueden comparar los datos de ambos estudios y sacar conclusiones de tal forma que se puede obtener información sobre si únicamente es casualidad o error.



Reiteramos que estos cuadros de mandos únicamente unen las hojas de trabajo por su utilidad juntas y por la sencillez que pueda aportar, pero no existe ninguna diferencia de estas que se muestran a las que se encuentran individualmente.

Además, debido al gran tamaño de nuestras bases de datos, nos es imposible mostrar todos los gráficos en este documento. Si bien es cierto que hemos proporcionado un ejemplo de cada tipo de gráfico, y los más representativos los pondremos también en las conclusiones, el resto del análisis estará disponible en el anexo correspondiente a esta parte. Mencionar además, que a diferencia de los otros anexos, este no es un documento escrito, sino un libro de Tableau. Hemos elegido esta opción ya que, con este formato que hemos implementado permite al usuario interactuar con los filtros preparados para seleccionar la información que más interese en cada momento. Además, para una visión más general hemos implementado una aplicación donde se puede ver el consumo medio de cada edificio del Campus de Vera de la UPV a través de un mapa interactivo.

En este mapa, se resuelve uno de los objetivos planteados, es decir, la representación del consumo total de cada edificio de la UPV. Además, se puede interactuar, de manera que se puede localizar cualquier edificio y, con un click, saber su consumo total y el promedio del consumo por mes.



Conclusiones

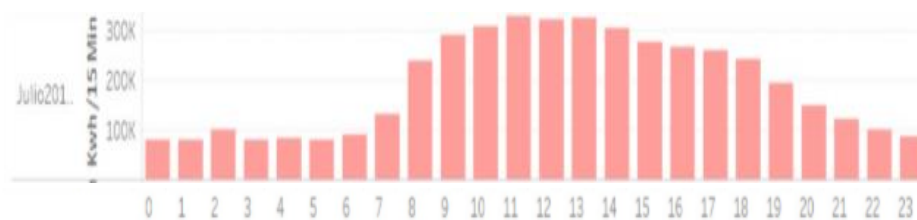
Se pueden separar en función de dos parámetros, según los meses o los edificios.

Primeramente, vamos a explicar la información extraída del trabajo relacionada con los meses: podemos observar dos gráficos relacionados con el análisis llevado a cabo. Cada uno de ellos nos aporta una información útil a la hora de sacar conclusiones.

La primera de ellas, apreciable en la imagen siguiente, va encaminada hacia el tema del gasto total. Cada mes es independiente del resto y es por ello por lo que nos llama la atención el gran consumo que se realizó en el mes de Julio. Dicho mes tiene unas características muy específicas, las cuales son que no se imparten clases, no se realizan exámenes o son mínimos y los trabajadores son menos de lo habitual, supuestamente. Por lo tanto, podemos sacar varias hipótesis de ello: hay errores de medición a causa de un fallo en los sensores, hubo un fenómeno inusual que causó este enorme dispendio o simplemente es un mes en el que se hace gran uso de las instalaciones y es normal la cifra de consumo.



Además, entrando en detalle, en Julio existe una clara diferencia del consumo en función de las horas como se puede ver en el siguiente gráfico.



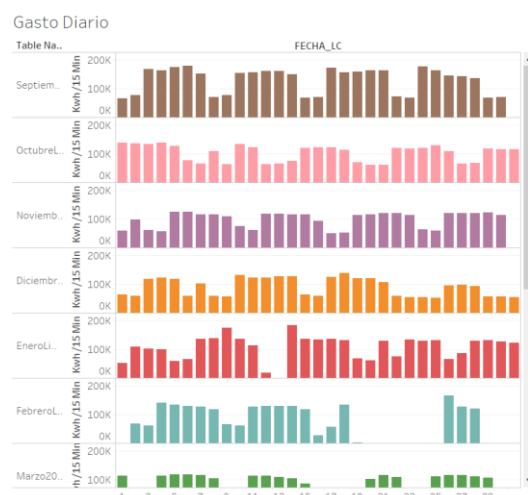
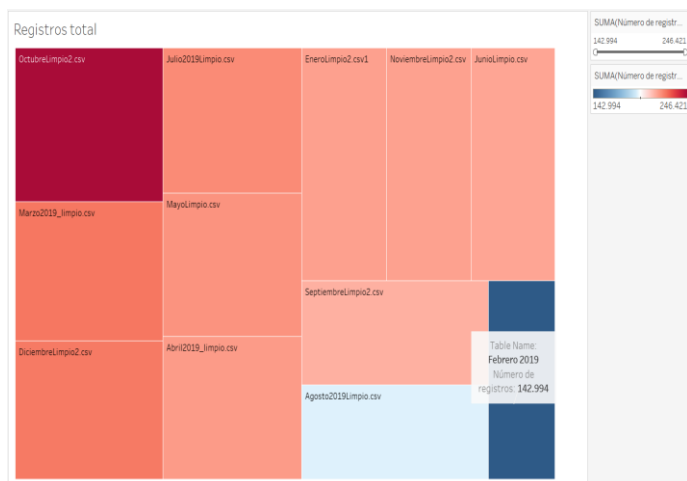
A continuación, vamos a acabar explicando las conclusiones extraídas de los edificios, dejando de lado los meses. Eso significa que no hemos tenido en cuenta los edificios en los que sucedía algo peculiar en un mes aislado, sino que hemos sacado hipótesis de aquellos que nos aportan información extraordinaria, indiferentemente del mes.

Empezamos detectando los casos de exceso, en qué edificios o grupos de estos había un consumo desmesurado. Para ello, hemos tenido en cuenta el análisis 'Consumo', en el cual hemos podido observar varios gráficos de un mismo conjunto de edificios pero de diferentes meses en el que era muy llamativo, en comparación con el resto, su gasto. A continuación, el gráfico que se muestra es uno de ellos, en el que podemos apreciar el enorme gasto que se tiene en el bloque de edificios 8G, 8E, 8B, 8M.

Una de las ideas que se nos ocurren sería hacer un seguimiento más en profundidad de este bloque, de tal manera que podríamos averiguar si es algo normal o, por el contrario, se pudiese reducir en gran medida este consumo. En caso de poder menguar en gasto, tendríamos la oportunidad de ahorrar una gran cantidad de electricidad, que se traduce en dinero, para poder invertirlo en otras necesidades de la universidad.



Otra conclusión viene dada del gráfico de abajo a la izquierda, en el que se observan la cantidad de registros obtenidos de los sensores a lo largo de cada mes. Gracias a los colores, podemos distinguir un caso llamativo en el mes de Febrero, en el cual hay un número de registros escaso en comparación con el resto de meses. Pero, con tal de conseguir una conclusión sólida, comprobamos que ocurría en este mes a partir de un análisis que muestra el gasto diario en función del mes. De esta manera, podemos afirmar que hay errores en los medidores, ya que deja de haber registros en los días 20-25. En definitiva, Febrero fue un mes en el que se tuvieron fallos con los sensores y no sabemos realmente el gasto que hubo. De todas formas, se puede apreciar en los días de los que sí tenemos registros, que su consumo no era anormal si lo relacionamos con los demás meses.

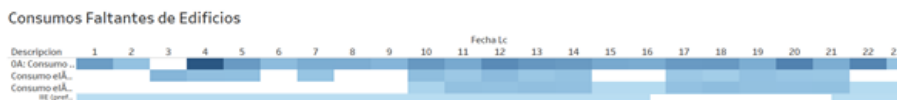
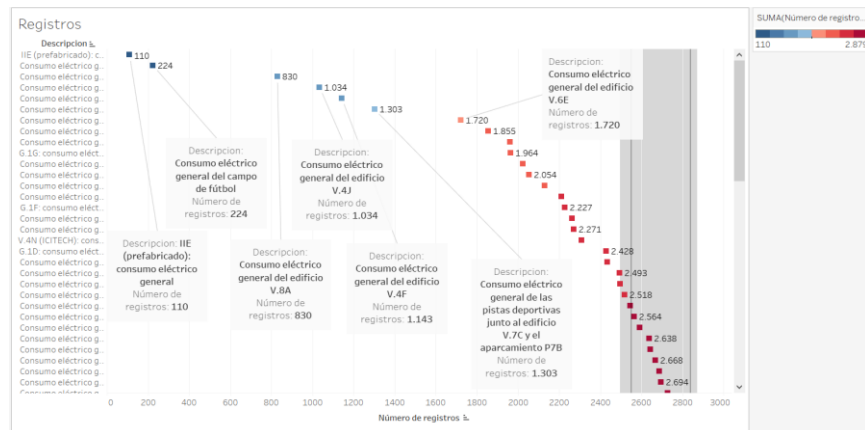


Por otro lado, hemos conseguido recoger información de los análisis 'Registros' y 'Faltantes', de tal manera que conseguimos relacionar información de ambas consiguiendo así una mejor hipótesis de sobre los resultados.

Antes de nada, utilizando el análisis 'Registros', nos hemos percatado de la poca cantidad de registros que tenían unos edificios en concreto, sin importar los meses. Estos edificios, que se muestran en la imagen situada abajo, son los siguientes: 'IIE (prefabricado): consumo eléctrico general'*, 'Consumo eléctrico general del campo de fútbol'*, 'Consumo eléctrico general del edificio V.8A', 'Consumo eléctrico general del edificio V.4J'*, 'Consumo eléctrico general del edificio V.4F', '0A: Consumo eléctrico general del alumbrado público'*, 'Consumo eléctrico general de las pistas deportivas junto al edificio V.7C y el aparcamiento P7B', 'Consumo eléctrico general del edificio V.6E'.

Con tal no sacar conclusiones precipitadas, encontramos indicios de faltantes en el análisis 'Faltantes'. Como se puede observar en la segunda imagen de abajo, 4 de los edificios que tienen pocos registros aparecen en el otro análisis con días faltantes, lo que quiere decir que va relacionado el hecho de tener pocos registros con fallos en los sensores. Los otros 4 edificios que no tienen faltantes, pero sí pocos registros son en lo que hay que fijarse más, ya que de estos pueden haber diversas hipótesis, como que los sensores dejen de funcionar en un determinado momento del día, que estén programados para no recoger el consumo cada 15 min sino cada más tiempo o quizá sea por un error a la hora de contar los registros del sensor.

Con estos gráficos, cerramos otro de los objetivos planteados, es decir, el análisis de los datos faltantes en función de cada edificio. Cabe añadir, que este objetivo es explicado con mayor detalle dentro de los anexos.



En definitiva, podemos concluir el análisis con varias ideas principales: la primera de ellas es que en el gasto total debemos de hacer un análisis más en profundidad del mes de Julio con tal de reducir su gasto, además de controlar en Febrero su ínfimo gasto respecto a los demás meses.

En relación con los edificios, tratar de menguar el consumo que realiza el conjunto de edificios 8G, 8E, 8B, 8M. Este bloque de edificios se corresponde con todos los dedicados a institutos y centros de investigación y a las Start-Up de IDEAS UPV, esto quiere decir que en este bloque se concentran gran cantidad de laboratorios y maquinaria que muy probablemente tengan un consumo elevado de energía. Por tanto, una posible solución a reducir el consumo de este bloque sería averiguar qué máquinas o laboratorios en concreto son los responsables de este consumo y estudiar si de algún modo se puede reducir sin afectar a sus prestaciones y su rendimiento eficiente.

Finalmente, realizar un seguimiento de los demás edificios señalados anteriormente para tratar de reducir los errores que pueda haber con sus respectivos sensores o simplemente indicar si esos registros son usuales.

En síntesis, quedan cumplidos los objetivos, es decir, el consumo total por edificio queda representado con el mapa interactivo y con los gráficos expuestos, llegando a profundizar en función del día o de la hora. Los datos faltantes, por su parte, han sido tratados con diferentes gráficos que se encuentran en ampliación en los anexos y en la presente memoria la información más importante en forma de dashboard. A todo esto, se suman los resultados que proporciona el apartado análisis. En este último caso, hemos explicado un ejemplo de un posible mes, pero es necesario saber que se han analizado los 12 meses, es decir, las 12 bases de datos. Por último, toda la información presentada puede servir para ayudar a diseñar estrategias de reducción del consumo energético en la UPV.