

Universidad Internacional de La Rioja
Escuela Superior de Ingeniería y Tecnología

Grado en Ingeniería Informática

Análisis del porfolio de clientes de
electricidad mediante curvas de carga
reales para mejorar su factura

| | |
|--|---|
| Trabajo fin de estudio presentado por: | Laia Casulà Medina |
| Director/a: | Juan Carlos Martínez Ferreiro |
| Fecha: | 09/07/2025 |
| Repositorio del código fuente: | https://github.com/laiacasula17/codigo-TFG-laia |

Agradecimientos

Quiero agradecer a todas las personas que me han acompañado durante la realización de este Trabajo de Fin de Grado. En primer lugar, a mi director, Juan Carlos Martínez Ferreiro, por su orientación técnica, sus comentarios precisos y su paciencia en todo el proceso. También a mis compañeros de profesión y a mi equipo en Audax Renovables, por brindarme acceso a los datos y por sus valiosos comentarios desde la experiencia práctica.

A mi familia, por su apoyo incondicional, y a mis amistades, por los ánimos y la comprensión en los momentos de más carga de trabajo.

Finalmente, a la Universidad Internacional de La Rioja, por ofrecer un entorno flexible, exigente y estimulante que ha hecho posible compatibilizar mi vida laboral con la finalización de estos estudios.

Resumen

El presente Trabajo de Fin de Grado desarrolla una solución analítica orientada a comparar curvas de carga reales de clientes domésticos frente a los perfiles estándar utilizados en su facturación, con el fin de identificar desviaciones significativas y potenciales oportunidades de ahorro.

La solución se enfoca en el ámbito doméstico en España, donde la facturación eléctrica se basa mayoritariamente en perfiles estándar definidos por Red Eléctrica Española (REE). Aunque esta metodología de cálculo de factura, facilita la operativa del sistema, puede generar desviaciones relevantes en la facturación individual si el patrón de consumo de un cliente no se ajusta al perfil medio aplicado.

En la compra de energía, los desajustes se compensan globalmente entre clientes. No obstante, a nivel individual, el uso de perfiles estándar puede generar diferencias significativas en la factura eléctrica de los clientes. Por este motivo, se plantea la posibilidad de analizar el perfil real de consumo y determinar si conviene aplicar otro tipo de estrategia, ya sea tarifaria, de autoconsumo o de facturación horaria.

La implementación de la herramienta se apoya en tecnologías de análisis y visualización de datos como Python, SQL Server y Looker Studio, permitiendo la exploración de datos históricos y la identificación de patrones de consumo atípicos. El análisis incluye la comparación detallada entre el consumo facturado y el consumo real horario, así como la simulación del impacto económico asociado.

Los resultados de este estudio contribuyen a mejorar la personalización de ofertas para los clientes y optimizar la gestión de la demanda por parte de las comercializadoras eléctricas. Como ejemplo, en uno de los casos identificados, un cliente llegó a pagar 11,87 € más en un solo mes, debido a la diferencia entre su consumo real y el perfil estándar asignado; lo que pone de manifiesto el impacto potencial del análisis propuesto.

Palabras clave: consumo eléctrico, cliente doméstico, facturación, ahorro energético, análisis de datos.

Abstract

This Final Degree Project develops an analytical solution aimed at comparing the actual load curves of residential electricity customers against the standard profiles used in their billing. The goal is to detect significant deviations and identify potential savings opportunities.

The proposed solution focuses on the residential sector in Spain, where electricity billing is mostly based on standard profiles published by Red Eléctrica Española (REE). Although this method simplifies operations, it can lead to significant mismatches in billing if a customer's consumption pattern differs from the average profile.

While energy purchases balance out globally across customers, at an individual level, standard profile billing may generate substantial discrepancies. Therefore, this study proposes analyzing real consumption curves to assess whether customers would benefit from alternative strategies such as time-of-use billing, solar self-consumption, or other tailored approaches.

The tool is built using Python, SQL Server, and Looker Studio to perform data cleansing, analysis, and visualization of historical consumption data. It includes the comparison of billed versus actual consumption and estimates the financial impact of the observed differences.

The findings contribute to improving tariff personalization for customers and optimizing demand-side management for electricity retailers. In one of the analyzed cases, a customer overpaid by €11.87 in a single month due to the mismatch between their real consumption and the assigned standard profile.

Keywords: electricity consumption, residential customer, billing, energy savings, data analysis.

Índice de contenidos

| | |
|--|-----|
| Glosario Técnico | XII |
| 1. Introducción | 1 |
| 1.1. Motivación | 1 |
| 1.1.1. Beneficios de la solución propuesta..... | 3 |
| 1.1.2. Obstáculos identificados. | 3 |
| 1.2. Caso práctico de comparación entre curva real y facturación estándar..... | 3 |
| 1.3. Planteamiento del trabajo | 4 |
| 1.4. Alcance del proyecto: | 4 |
| 1.5. Consideraciones adicionales..... | 5 |
| 1.6. Estructura del trabajo | 6 |
| 2. Contexto y estado del arte | 7 |
| 2.1. Estructura del sistema eléctrico español..... | 7 |
| 2.2. El flujo de información: Desde la medida hasta la factura | 8 |
| 2.3. Limitaciones del uso de perfiles estándar | 8 |
| 2.4. Comparativa con el modelo industrial..... | 9 |
| 2.5. Soluciones existentes y posibles alternativas..... | 9 |
| 2.6. Marco Normativo y Regulatorio Ampliado..... | 10 |
| 3. Objetivos y metodología de trabajo..... | 11 |
| 3.1. Objetivo general..... | 11 |
| 3.2. Objetivos específicos | 11 |
| 3.3. Metodología de trabajo | 13 |
| 3.3.1. Fases del proyecto:..... | 13 |
| 3.3.2. División de sprints | 14 |
| 3.3.3. Cronograma de ejecución de sprints..... | 15 |

| | | |
|--------|--|----|
| 3.3.4. | Casuísticas resueltas durante los sprints..... | 16 |
| 3.4. | Justificación metodológica..... | 16 |
| 3.4.1. | Consideraciones éticas y de confidencialidad..... | 17 |
| 3.5. | Evaluación del cumplimiento de los objetivos | 18 |
| 4. | Desarrollo de la Herramienta..... | 19 |
| 4.1. | Definición Funcional | 19 |
| 4.1.1. | Requisitos funcionales..... | 19 |
| 4.1.2. | Requisitos no funcionales..... | 20 |
| 4.1.3. | Casos de Uso..... | 21 |
| 4.1.4. | Matriz de Trazabilidad entre OE, RF y CU..... | 23 |
| 4.2. | Análisis | 24 |
| 4.2.1. | Obtención de datos | 24 |
| 4.2.2. | Tratamiento y limpieza de datos..... | 24 |
| 4.2.3. | Análisis comparativo..... | 25 |
| 4.2.4. | Descripción del problema abordado | 26 |
| 4.2.5. | Evolución de la lógica de desarrollo | 26 |
| 4.2.6. | Herramientas utilizadas..... | 26 |
| 4.2.7. | Estructura técnica de la solución..... | 27 |
| 4.2.8. | Metodología del clustering y visualización de los resultados | 27 |
| 4.2.9. | Conexión con el objetivo 7: Visualización final | 29 |
| 4.3. | Visualización..... | 30 |
| 4.3.1. | Fuentes de datos | 30 |
| 4.3.2. | Objetivo de la visualización | 31 |
| 4.3.3. | Conexión entre el análisis y la visualización..... | 31 |
| 4.3.4. | Subida de datos a Google Sheets (Load.py) | 31 |

| | | |
|--------|--|----|
| 4.3.5. | Construcción del dashboard..... | 32 |
| 4.3.6. | Métricas mostradas en Looker Studio..... | 32 |
| 4.3.7. | Alineación con el objetivo 7 | 37 |
| 4.4. | Escalabilidad y mejoras técnicas..... | 37 |
| 4.5. | Casos de uso del dashboard en negocio..... | 38 |
| 4.6. | Código | 38 |
| 4.6.1. | Repositorio GIT | 38 |
| 4.6.2. | Link Looker Studio | 39 |
| 4.6.3. | Esquema Componentes..... | 39 |
| 4.7. | Componentes del sistema | 39 |
| 4.7.1. | Extract.sql: Origen de datos: SQL Server..... | 39 |
| 4.7.2. | Componente de transformación: Transform.py | 40 |
| 4.7.3. | Componente de carga: Load.py..... | 42 |
| 4.7.4. | Visualización: Looker Studio..... | 42 |
| 5. | Evaluación y Validación | 43 |
| 5.1. | Validación técnica de los cálculos..... | 43 |
| 5.2. | Análisis estadístico de la muestra..... | 43 |
| 5.3. | Validación visual con Looker Studio | 44 |
| 5.4. | Coherencia con hipótesis iniciales | 44 |
| 5.5. | Análisis de sensibilidad y validación cruzada..... | 45 |
| 5.5.1. | Umbral de calidad de datos..... | 45 |
| 5.5.2. | Agrupación mensual vs. estacional | 45 |
| 5.5.3. | Conclusión del análisis..... | 45 |
| 5.5.4. | Limitaciones del análisis y calidad del dato..... | 46 |
| 6. | Conclusiones y trabajo futuro | 47 |

| | | |
|--------|---|----|
| 6.1. | Conclusiones del trabajo..... | 47 |
| 6.2. | Conclusiones en detalle | 47 |
| 6.2.1. | CNAEs..... | 48 |
| 6.2.2. | Municipios con desviaciones elevadas..... | 48 |
| 6.2.3. | Tarifas | 48 |
| 6.2.4. | Productos..... | 49 |
| 6.2.5. | Ahorro global y eficiencia estimada | 49 |
| 6.2.6. | Viabilidad técnica y económica | 49 |
| 6.3. | Limitaciones y Riesgos del Análisis | 51 |
| 6.4. | Líneas de trabajo futuro | 51 |
| 6.4.1. | Ampliar el análisis a otras tarifas y segmentos | 51 |
| 6.4.2. | Integrar nuevas variables predictivas..... | 52 |
| 6.4.3. | Evolución hacia modelos supervisados | 52 |
| 6.4.4. | Construcción de nuevos perfiles estándar | 52 |
| 6.4.5. | Despliegue como herramienta de negocio | 52 |
| 6.4.6. | Evaluación del impacto ambiental y social..... | 52 |
| 6.4.7. | 6.4.7. Diseño técnico futuro de la herramienta | 53 |
| 6.5. | Reflexión estratégica | 53 |
| 6.6. | Implicaciones regulatorias y técnicas | 54 |
| 6.7. | Propuesta de implementación real | 54 |
| 6.8. | Comparativa con otras innovaciones del sector..... | 55 |
| 6.9. | Segmentación como herramienta de negocio..... | 55 |
| 6.10. | Alineación con la estrategia del sector energético..... | 56 |
| 6.11. | Riesgos de no actuar | 57 |
| 6.12. | Visión de futuro: Herramienta integral de gestión energética..... | 57 |

| | | |
|-------|---|----|
| 6.13. | Recomendaciones para la implantación en una comercializadora real | 58 |
| 6.14. | Sostenibilidad y contribución al empoderamiento energético | 59 |
| 6.15. | Posibilidades de integración con sistemas corporativos | 59 |
| 7. | Reflexión estratégica y visión a largo plazo | 61 |

Índice de Ilustraciones

| | |
|--|----|
| Ilustración 1 Diagrama UML. Fuente: Elaboración propia | 21 |
| Ilustración 2 Centroides de los clústeres. Fuente: Elaboración propia..... | 29 |
| Ilustración 3 Métricas analizadas. Fuente: Elaboración propia | 33 |
| Ilustración 4 Análisis Facturación Curvas vs Perfiles. Fuente: Elaboración propia | 34 |
| Ilustración 5 Análisis detalle 20TD. Fuente: Elaboración propia..... | 35 |
| Ilustración 6 Análisis detalle 30TD. Fuente: Elaboración propia..... | 36 |
| Ilustración 7 Esquema componentes. Fuente: Elaboración propia | 39 |
| <i>Ilustración 8 Consumos Facturados</i> . Fuente: Elaboración propia..... | 40 |
| Ilustración 9 Consumos Curvas de las fechas de las facturas. Fuente: Elaboración propia..... | 40 |
| Ilustración 10 Creación de tabla resultado de consumos facturados y consumos curvas. Fuente: Elaboración propia | 40 |
| Ilustración 11 Limpieza de datos extraídos que no formarán parte del análisis. Fuente: Elaboración propia..... | 41 |
| Ilustración 12 Segmentación de CUPS por ahorro. Fuente: Elaboración propia | 41 |
| Ilustración 13 Fichero de salida anonimizado. Fuente: Elaboración propia | 41 |
| Ilustración 14 Configuración para exportar CSV a Sheets. Fuente: Elaboración propia..... | 42 |
| Ilustración 15 Subida a Sheets. Fuente: Elaboración propia..... | 42 |

Índice de tablas

| | |
|---|----|
| Tabla 1 Cronograma Sprints | 15 |
| Tabla 2 Evaluación de Objetivos..... | 18 |
| Tabla 3 Matriz de Trazabilidad entre OE, RF y CU..... | 23 |
| Tabla 4 Ahorro por trimestre..... | 45 |
| Tabla 5 Coste estimado implantación | 50 |

Glosario Técnico

ATR (Acceso de Terceros a la Red): Ficheros estándar que contienen las medidas de consumo eléctrico, enviados por las distribuidoras a las comercializadoras.

Centroide: Punto central que representa un clúster en técnicas de agrupamiento como el clustering. Se calcula como el promedio de los valores de todas las variables para los elementos del grupo. Sirve como referencia para describir y comparar la posición relativa de los distintos clústeres dentro del espacio de datos.

Clustering: Técnica de aprendizaje automático no supervisado que permite agrupar elementos con características similares en conjuntos denominados clústeres. Su finalidad es identificar patrones comunes dentro de un conjunto de datos, sin necesidad de categorías predefinidas. Es útil para detectar comportamientos o perfiles representativos que faciliten el análisis y la toma de decisiones.

CRM (Customer Relationship Management): Sistema de gestión que permite organizar, automatizar y sincronizar las interacciones con usuarios o partes interesadas. Su finalidad es mejorar la relación, el seguimiento y la fidelización, centralizando la información relevante sobre contactos, comunicaciones y actividades comerciales o de soporte.

CUPS (Código Universal del Punto de Suministro): Identificador único de cada punto de suministro eléctrico o de gas.

Curva de carga horaria: Registro horario del consumo eléctrico real de un cliente durante un periodo determinado.

Dashboard: Entorno visual interactivo que permite representar de forma resumida y estructurada indicadores clave, métricas y datos relevantes. Su objetivo es facilitar el análisis, el seguimiento y la toma de decisiones mediante gráficos, tablas y otros elementos visuales. También se conoce como cuadro de mando o panel de control.

Dataset: Conjunto estructurado de datos que se utiliza como base para el análisis, el entrenamiento de modelos o la visualización de resultados. Suele organizarse en forma de tabla con filas (observaciones) y columnas (variables). Un dataset puede provenir de bases de datos, archivos CSV, sensores, formularios u otras fuentes digitales.

ERP (Enterprise Resource Planning): Sistema de planificación de recursos empresariales que integra en una única plataforma los procesos clave de una organización, como finanzas, compras, inventario, producción o recursos humanos. Facilita la coordinación, el control y la toma de decisiones a partir de datos centralizados y actualizados.

ETL (Extract, Transform, Load): Proceso de extracción, transformación y carga de datos utilizado en análisis y almacenamiento de información.

K-means: Algoritmo de agrupamiento no supervisado que segmenta observaciones en clústeres según su similitud.

Perfil estándar: Modelo de consumo medio definido por REE y utilizado para la facturación cuando no se dispone de curva real.

PNIEC: El Plan Nacional Integrado de Energía y Clima (PNIEC) se constituye como la herramienta de orientación estratégica nacional que integra la política de energía y clima con un horizonte temporal a 2030, de acuerdo con la normativa nacional y europea

Python: Lenguaje de programación utilizado en este trabajo para tratar y analizar datos energéticos.

SCRUM: Marco de trabajo ágil utilizado en la gestión de proyectos (Cockburn, 2006), especialmente en desarrollo de software. Se basa en ciclos iterativos y colaborativos llamados sprints, con roles definidos (Scrum Master, Product Owner y equipo de desarrollo) y eventos estructurados como reuniones diarias (daily), planificación de sprint y retrospectivas. SCRUM promueve la entrega incremental de valor, la adaptación continua y la mejora constante del proceso.

SPRINT: Intervalo de tiempo fijo dentro del marco de trabajo Scrum durante el cual se desarrolla un incremento del producto con valor añadido. Suele tener una duración de entre una y cuatro semanas

SQL Server: Sistema de gestión de bases de datos relacionales empleado como origen de datos para el análisis.

Looker Studio: Herramienta de visualización que permite construir paneles dinámicos sobre datos estructurados.

Scikit-learn: Librería de Python para aprendizaje automático, utilizada en este trabajo para aplicar clustering.

1. Introducción

La introducción de este Trabajo de Fin de Grado se estructura para contextualizar la problemática, justificar la motivación del estudio y definir claramente el objetivo del proyecto. La evolución del sector eléctrico, la creciente digitalización de los contadores y la necesidad de una mayor transparencia en la facturación, son elementos que han propiciado la búsqueda de nuevas herramientas que aporten precisión y valor tanto para los usuarios como para las comercializadoras.

En este sentido, se plantea una solución basada en el análisis de curvas de carga reales en contraposición al uso de perfiles estándar. El modelo vigente, aunque funcional a nivel global, presenta desviaciones que pueden tener un impacto económico relevante a nivel individual. Estas diferencias surgen porque los perfiles estándar reflejan un consumo medio generalizado que no necesariamente coincide con el comportamiento real de cada cliente.

Además, se observa una oportunidad para mejorar los procesos de compra de energía por parte de las comercializadoras, quienes podrían beneficiarse de una mejor alineación entre el consumo real y la previsión de demanda. En consecuencia, la herramienta propuesta no solo aporta valor al usuario final, sino que también constituye un instrumento estratégico para el sector eléctrico.

El presente capítulo introduce los elementos clave que guiarán el desarrollo del proyecto y proporciona una visión general de los capítulos siguientes. Se abordan los antecedentes normativos, tecnológicos y operativos que sustentan el trabajo, así como el enfoque metodológico adoptado para su implementación.

1.1. Motivación

La experiencia profesional del autor en el sector eléctrico ha permitido identificar ineficiencias relevantes en la forma en que se gestiona la facturación para clientes domésticos. A diferencia de los clientes industriales, cuya energía se compra y factura por curvas horarias reales, los clientes domésticos son mayoritariamente facturados con perfiles estándar, definidos por Red

Eléctrica Española (REE). Esta simplificación operativa, aunque adecuada para el volumen, puede resultar perjudicial en términos de equidad y precisión.

El principal inconveniente radica en que un perfil estándar representa un consumo medio generalizado, lo cual invisibiliza patrones individuales que pueden diferir considerablemente. Clientes que concentran su consumo en horarios de menor coste —por ejemplo, durante el mediodía con alta generación solar— no ven reflejado ese beneficio si el perfil estándar agrupa el periodo completo. Como consecuencia, pagan precios medios por periodos, sin que su comportamiento eficiente sea premiado.

Además, el sistema actual, Ministerio de Industria, Energía y Turismo. (2012) también impone limitaciones desde la perspectiva de la comercializadora. Al no disponer de curvas completas en el corto plazo (solo se alcanza un 90 % de disponibilidad de curvas pasados seis meses), las empresas se ven obligadas a aplicar perfiles estándar, retrasando así cualquier posibilidad de facturación horaria. Este desfase impacta en la experiencia del cliente, y complica la planificación financiera.

Asimismo, la normativa vigente (Ley 24/2013, de 26 de diciembre), si bien permite la facturación por curva real cuando los datos están disponibles, no obliga a su uso en el segmento doméstico, manteniéndose el modelo de perfil estándar. Esto representa una oportunidad de mejora en la que la analítica de datos puede jugar un papel clave.

El presente trabajo surge de esta necesidad detectada: identificar de forma sistemática aquellos clientes cuyo consumo real se desvía del perfil estándar aplicado, estimar el impacto económico de esa diferencia y, a partir de ahí, proponer estrategias que permitan una facturación más justa, eficiente y personalizada. A largo plazo, se plantea incluso la posibilidad de construir nuevos perfiles representativos para subgrupos con hábitos comunes, como pueden ser los usuarios con autoconsumo o los hogares electrificados con vehículos eléctricos.

En definitiva, la motivación de este proyecto se enmarca en la búsqueda de una mayor transparencia, eficiencia y equidad en la relación entre la comercializadora eléctrica y el consumidor final, aprovechando el potencial de las tecnologías actuales de tratamiento y visualización de datos.

1.1.1. Beneficios de la solución propuesta.

- Para las comercializadoras: Identificar clientes cuyo consumo real, difiere significativamente del perfil estándar y permitir ajustar estrategias de compra y facturación, optimizando la gestión de la demanda y mejorando la competitividad en el mercado, mejorando la eficiencia en la demanda y minimizando el coste de los desvíos.
- Para los clientes: Aquellos con patrones de consumo que se desvían del perfil estándar podrían beneficiarse de tarifas más precisas respecto a su consumo real, lo que se traduciría en ahorros económicos y una facturación más justa (Ministerio para la Transición Ecológica y el Reto Demográfico. (2020).

1.1.2. Obstáculos identificados.

- Disponibilidad de datos: La falta de datos completos de curvas de carga reales es un desafío. En ocasiones, faltan horas e incluso días de datos, lo que dificulta un análisis preciso
- Percepción del cliente: Implementar una facturación basada en datos reales podría retrasar la emisión de facturas, acumulando varias en un solo periodo. Esto podría generar confusión y descontento en clientes no acostumbrados a gestionar pagos acumulados.
- Necesidad de educación: Para cambiar el sector hacia una facturación más precisa, es esencial educar a los consumidores sobre los beneficios y funcionamiento de este sistema, promoviendo una mayor comprensión y aceptación.

1.2. Caso práctico de comparación entre curva real y facturación estándar

En un análisis realizado con datos reales de un cliente doméstico, se ha comparado el coste derivado del uso de curvas reales de consumo frente al coste facturado con perfil estándar. Por ejemplo, durante el mes de noviembre de 2023, el cliente registró un consumo real de 146,83 kWh, cuyo coste estimado fue de 21,75 €. Sin embargo, fue facturado por 226,93 kWh, lo que supuso un coste extra para el cliente de 33,62 €, generando una diferencia de +11,87 € a su cargo. En otros meses, el efecto puede invertirse, y el cliente termina pagando menos de lo que correspondería con su curva real, lo que evidencia el desequilibrio del sistema de

perfiles estándar. Esta variabilidad justifica la necesidad de contar con herramientas de análisis que permitan identificar estas situaciones y ajustar la estrategia de facturación.

Esta herramienta contribuiría también a una planificación energética más eficiente desde la perspectiva de la comercializadora, al poder basar su previsión de compra en patrones de consumo reales y no promedios, reduciendo así los desvíos entre energía contratada y consumida. Por lo tanto, reduciendo los sobre costes que provoca la falta de precisión, que sistemáticamente implica la compra de más o menos energía de la necesaria.

1.3. Planteamiento del trabajo

Este proyecto plantea el desarrollo de una herramienta analítica que permita comparar de forma masiva las curvas de carga reales de clientes domésticos con los perfiles estándar que se utilizan para su facturación. Se parte de datos históricos anonimizados y extraídos de la base de datos SQL Server de una comercializadora eléctrica, donde se incluyen tanto curvas horarias de consumo como valores facturados reales.

La propuesta se centra en construir un proceso ETL (extracción, transformación y carga) que depure y prepare los datos para su análisis, y que integre capacidades de visualización e identificación automática de desviaciones relevantes.

El objetivo es evaluar el impacto económico que tiene para el cliente doméstico el hecho de ser facturado por perfil en lugar de por curva real. A partir de los resultados, se busca definir reglas de negocio que permitan detectar casos en los que sería conveniente facturar con curva, ajustar el producto o incluso rediseñar los perfiles aplicados.

El alcance funcional incluye el desarrollo de la lógica analítica, el tratamiento de los datos y la presentación visual de resultados mediante Looker Studio, sin desarrollar un producto final automatizado para el cliente, pero con potencial de escalabilidad e integración futura.

1.4. Alcance del proyecto:

El proyecto se acota a los siguientes elementos:

- Análisis de datos históricos reales, con curvas horarias y datos de facturación de una muestra representativa de clientes domésticos.
- Enfoque centrado en clientes denominados Consumidores. Con tarifa 20TD o 30TD.
- Exclusión de datos incompletos o con registros faltantes, con justificación de criterios de calidad mínima.
- Simulación de impacto económico cliente a cliente.
- Identificación de patrones comunes entre clientes con mayores desviaciones.

Posibles aplicaciones:

- Identificar Patrones por grupo de Perfil.
- Identificar clientes a lo que sugerir cambios de patrón para ahorro.

No se incluye el desarrollo de un motor de facturación, ni la integración con sistemas productivos. Tampoco se contemplan datos meteorológicos externos, aunque se deja como posible mejora futura.

1.5. Consideraciones adicionales

Debido a la naturaleza del análisis, se requiere un tratamiento riguroso de la calidad de los datos. La validación, limpieza y transformación de los registros de curvas y facturación es crítica para asegurar que los resultados sean representativos.

Asimismo, debe tenerse en cuenta que el sistema actual de facturación con perfiles estándar se sustenta en una operativa regulatoria. Por tanto, cualquier propuesta que derive de este proyecto tendría que considerarse dentro del marco normativo vigente y, en su caso, promoverse como mejora o innovación sectorial.

Se incluirá también una reflexión sobre el posible impacto ambiental y económico del ajuste entre consumo real y facturación, especialmente en un contexto donde la generación distribuida, el autoconsumo y la flexibilidad de la demanda serán cada vez más relevantes.

Este tipo de ajustes, además de mejorar la eficiencia y sostenibilidad del sistema, podrían contribuir a reducir el riesgo de desequilibrios como los observados durante la interrupción de suministro eléctrico del 28 de abril, donde se evidenció la necesidad de una previsión más precisa y distribuida del consumo.

1.6. Estructura del trabajo

El trabajo se estructura en siete capítulos principales:

Capítulo 1: Introducción. Se describe el problema, la motivación, el planteamiento del trabajo, el alcance, las consideraciones metodológicas iniciales y la estructura general del documento.

Capítulo 2: Contexto y estado del arte. Se analiza el funcionamiento del sistema eléctrico español, los agentes implicados, el flujo de datos, las limitaciones del perfil estándar y las soluciones actuales en el mercado.

Capítulo 3: Objetivos y metodología. Se detallan los objetivos generales y específicos del proyecto, y se presenta la metodología de desarrollo, incluyendo el enfoque ágil basado en SCRUM y sus iteraciones.

Capítulo 4: Desarrollo de la herramienta. Se documentan las fases de toma de requisitos, análisis, diseño e implementación de la solución técnica, incluyendo fragmentos de código, lógica de transformación y visualizaciones.

Capítulo 5: Evaluación y validación. Se presentan los resultados obtenidos y su validación comparativa con datos reales de facturación.

Capítulo 6: Conclusiones y líneas futuras. Se resumen las aportaciones del proyecto, se reflexiona sobre sus limitaciones y se plantean posibles mejoras o desarrollos futuros.

Capítulo 7: Conclusión estratégica y visión a largo plazo: Presenta una reflexión estratégica sobre el impacto a largo plazo de la solución desarrollada, destacando su contribución a un modelo energético más personalizado, justo y centrado en el cliente.

Cada uno de los capítulos busca abordar los distintos aspectos del proyecto de manera progresiva y coherente, desde la identificación del problema hasta la validación de los resultados y la propuesta de evolución futura.

2. Contexto y estado del arte

Este capítulo presenta el marco conceptual y técnico en el que se desarrolla el trabajo. Se describe el contexto sectorial, se revisan las prácticas actuales en la gestión y facturación energética, y se analizan estudios previos y tecnologías relevantes relacionadas con el análisis de curvas de carga y técnicas de clustering aplicadas a datos de consumo.

2.1. Estructura del sistema eléctrico español

El sector eléctrico en España está estructurado en varios niveles y agentes que intervienen en la generación, transporte, distribución, comercialización y consumo de energía. Entre los actores principales destacan:

- **Red Eléctrica de España (REE):** Es el operador del sistema eléctrico y el responsable del transporte de la electricidad en alta tensión. REE publica, entre otras cosas, los perfiles de consumo estándar utilizados en la facturación cuando no se dispone de curvas de carga reales. También gestiona el sistema de medidas y validación de datos de consumo.
- **Distribuidoras:** Son las empresas encargadas de llevar físicamente la electricidad hasta los puntos de consumo. Además, son responsables de la lectura de los contadores y de proporcionar los datos de consumo a las comercializadoras mediante los ficheros ATR (Acceso de Terceros a la Red), conforme a lo establecido por el regulador.
- **Comercializadoras:** Compañías que venden la electricidad a los consumidores finales. Compran la energía en los mercados mayoristas, OMIE. (2024) y la facturan a sus clientes, ya sea aplicando precios indexados, fijos o personalizados. En el caso de clientes domésticos, la gran mayoría son facturados mediante perfiles estándar.
- **Clientes domésticos:** Consumidores con potencias contratadas inferiores a 15 kW, cuyas facturas suelen calcularse usando perfiles estándar si no disponen de contadores con telemedida activa o si sus datos de curvas horarias no están completos.

Este sistema presenta un importante hándicap: Cuando no se dispone de curvas reales completas en tiempo y forma, se aplica un perfil medio estándar para toda la facturación, lo que puede provocar desviaciones significativas entre lo consumido y lo cobrado.

La compra de energía se realiza por parte de la comercializadora en base a previsiones, lo que introduce incertidumbre en la valoración posterior del margen real. Por tanto, el sistema actual penaliza tanto al cliente como a la empresa cuando el perfil asignado no refleja adecuadamente el consumo real.

2.2. El flujo de información: Desde la medida hasta la factura

El proceso de facturación eléctrica se basa en la medida del consumo horario recogido por los contadores digitales. La distribuidora valida y remite estos datos a las comercializadoras mediante los ficheros ATR, que incluyen:

- Datos identificativos del suministro (CUPS, tarifa, entre otros.).
- Períodos de medida y consumo horario (cuando está disponible).
- Perfil estándar aplicado (en caso de no tener curva real).

La comercializadora genera la factura agrupando los consumos por periodos tarifarios (P1, P2, y P3). Si el cliente no dispone de curva horaria, se aplica el perfil estándar para repartir el consumo total mensual entre los distintos periodos. Este procedimiento, aunque regulado, puede generar desajustes importantes si el patrón real del cliente difiere del perfil medio.

2.3. Limitaciones del uso de perfiles estándar

El uso de perfiles estándar simplifica la gestión operativa de las comercializadoras, pero presenta limitaciones relevantes:

- **Inexactitud individual:** El perfil estándar es una media generalizada, lo que penaliza a clientes con hábitos diferentes. Por ejemplo, quienes concentran su consumo en horas solares pueden acabar pagando como si consumieran en horas punta.
- **Invisibilización de tecnologías emergentes:** El perfil estándar no contempla patrones derivados de autoconsumo, almacenamiento energético o consumo flexible.
- **Riesgo económico:** Las desviaciones pueden suponer sobrecostes relevantes para el cliente, y pérdida de eficiencia para la comercializadora si sus previsiones de compra no se ajustan al consumo real.

Este desfase afecta tanto a clientes como a comercializadoras, que pueden ver comprometido su margen de beneficio.

2.4. Comparativa con el modelo industrial

En cambio, los clientes industriales (habitualmente con consumos superiores a 1 GWh/año) están obligados a ser facturados por su curva de carga real. Esto permite:

- Una mayor precisión en la imputación de costes.
- Ajustes más finos en las tarifas contratadas.
- Gestión activa de la demanda y optimización de potencia contratada.

Este modelo más transparente y detallado se ha ido ampliando con el despliegue de contadores inteligentes. Sin embargo, su aplicación masiva en doméstico aún tiene barreras técnicas, económicas y sociales.

2.5. Soluciones existentes y posibles alternativas

Actualmente, las principales comercializadoras como Iberdrola, Endesa o Naturgy no ofrecen simuladores de facturación basados en curva real para el segmento doméstico. Algunos portales permiten consultar curvas registradas o realizar comparaciones, pero no incorporan este análisis en su estrategia comercial o en la facturación estándar.

Las principales barreras detectadas son:

- **Disponibilidad de datos:** Las curvas horarias no siempre están disponibles de forma completa y en tiempo para la facturación. Según datos del sector, sólo pasados seis meses se alcanza una cobertura del 90 %.
- **Complejidad tecnológica:** Implementar una facturación basada en curva real implica integrar nuevos procesos ETL, lógicas de cálculo y gestión de incidencias.
- **Escasa rentabilidad:** Los clientes domésticos fuera del perfil estándar son minoría. A nivel agregado, el modelo actual compensa desviaciones, por lo que el incentivo empresarial es bajo.

A nivel internacional, países como Reino Unido o Países Bajos han empezado a explorar perfiles más personalizados o facturación horaria para clientes con contadores inteligentes. En España, el despliegue de la telemedida y el avance de la digitalización energética abren nuevas oportunidades que este trabajo busca aprovechar desde un enfoque técnico y analítico.

2.6. Marco Normativo y Regulatorio Ampliado

El marco regulatorio del sistema eléctrico español establece el uso de perfiles estándar como mecanismo para facilitar la facturación de los consumidores domésticos cuando no se dispone de medidas horarias completas. Este modelo ha sido clave para el desarrollo del sector, pero también plantea limitaciones que se han abordado en el presente trabajo.

Ley 24/2013 (Jefatura del Estado, 2013) del Sector Eléctrico: Esta ley define el suministro eléctrico como un servicio de interés económico general, y establece los derechos de los consumidores, entre ellos el de recibir una facturación basada en consumos reales cuando éstos estén disponibles. Esta disposición es clave para justificar la validez del modelo alternativo basado en curvas reales.

Real Decreto 1718/2012 (Ministerio de Industria, 2012): Este RD regula el procedimiento de lectura y facturación. Su artículo 6 establece que "cuando se disponga de medida horaria, la facturación se realizará sobre dicha medida". Sin embargo, no obliga a su aplicación en el segmento doméstico, dejando margen para que las comercializadoras elijan si utilizar perfiles o curvas, lo cual genera desigualdad entre clientes.

PNIEC 2021–2030: El Plan Nacional Integrado de Energía y Clima (Ministerio para la Transición Ecológica y el Reto Demográfico, 2020), establece como línea estratégica la digitalización del sistema energético y el empoderamiento del consumidor. Esta visión encaja plenamente con el uso de datos horarios para mejorar la transparencia y la eficiencia en la facturación.

Este trabajo se sitúa en ese contexto regulatorio y aporta una solución técnica viable para avanzar hacia modelos más justos y adaptados al consumo real, en línea con los objetivos del marco normativo actual.

3. Objetivos y metodología de trabajo

Este capítulo tiene como propósito definir, de manera estructurada y precisa, tanto los objetivos generales como los específicos del trabajo, enmarcados en un contexto técnico y aplicado. Además, se detalla la metodología empleada para la ejecución del proyecto, la cual se basa en principios de desarrollo ágil, garantizando una implementación incremental, validada por etapas y orientada a la obtención de resultados tangibles.

Se presentan las distintas fases del análisis, que abarcan desde la adquisición inicial de los datos hasta su transformación, validación, análisis multivariable y posterior visualización. A lo largo del desarrollo se han ido tomando decisiones metodológicas justificadas por la naturaleza de los datos, el propósito del análisis y las limitaciones operativas del entorno en el que se ha llevado a cabo.

3.1. Objetivo general

El objetivo general de este Trabajo de Fin de Grado es desarrollar una herramienta analítica capaz de comparar de manera automatizada las curvas de carga reales de clientes domésticos con los perfiles estándar utilizados habitualmente en su facturación. A partir de esta comparativa, se pretende detectar desviaciones significativas y estimar su impacto económico, con el fin de proponer estrategias de optimización tarifaria y de facturación más justas, precisas y personalizadas para cada cliente.

Esta herramienta busca ser un apoyo técnico y estratégico para las comercializadoras eléctricas, permitiéndoles identificar oportunidades de mejora en la gestión de la demanda, fidelización de clientes y eficiencia operativa. Asimismo, promueve una mayor transparencia para el consumidor final, alineándose con los objetivos de digitalización del sistema energético definidos a nivel nacional y europeo

3.2. Objetivos específicos

Para lograr el objetivo general planteado, se han definido siete objetivos específicos que cubren de forma secuencial y complementaria las distintas etapas del proceso analítico. Cada

uno de ellos está orientado a generar un resultado tangible que contribuya a la consecución global del proyecto.

- **OE1.** Obtener, al menos, una muestra representativa del 10 % de la población total de puntos de suministro (alrededor de 30.000 CUPS), asegurando que los datos cumplen criterios de calidad mínimos (por ejemplo, ≥ 80 % de horas disponibles por mes y facturas asociadas completas).
- **OE2.** Comparar, cliente a cliente y mes a mes, el consumo horario real registrado frente al consumo facturado en base a perfil estándar, cuantificando la diferencia tanto en energía (kWh) como en términos económicos (€). Este análisis permitirá establecer la magnitud y recurrencia de las desviaciones.
- **OE3.** Estimar el impacto económico del modelo actual de facturación por perfil, calculando el ahorro potencial individual para cada cliente y el ahorro agregado para toda la muestra. Se espera identificar un ahorro total superior a los 400.000 € anuales en el conjunto de puntos de suministro analizados.
- **OE4.** Detectar al menos tres patrones comunes de desviación mediante la aplicación de técnicas de clustering no supervisado (K-means), con una separación inter-clúster suficiente (silhouette score $> 0,3$), que permita segmentar a los clientes en perfiles operativamente útiles para negocio.
- **OE5.** Formular, con base en el análisis anterior, un conjunto de reglas de decisión que permitan recomendar la facturación por curva en lugar de perfil estándar cuando se superen determinados umbrales, como un ahorro estimado superior a 50 € anuales o una desviación energética relativa superior al 10 %.
- **OE6.** Explorar la viabilidad de definir nuevos perfiles estándar agrupados por comportamiento real de consumo, identificando al menos dos grupos diferenciados, tales como usuarios con autoconsumo fotovoltaico diurno o consumidores con predominio de uso nocturno.
- **OE7.** Construir un conjunto de dashboards interactivos en Looker Studio, alimentados desde un pipeline ETL automatizado, que permitan visualizar y filtrar los resultados por variables como zona geográfica, producto, tipo de cliente o nivel de ahorro, asegurando una actualización completa de datos en menos de 5 minutos.

En conjunto, estos objetivos aseguran que la solución propuesta sea útil, medible y escalable, y que permita transformar datos brutos en información procesable para la toma de decisiones comerciales y estratégicas por parte de una comercializadora eléctrica.

3.3. Metodología de trabajo

La metodología adoptada para el desarrollo del trabajo es una adaptación del marco ágil SCRUM (Beck et al., 2001). A pesar de que se trata de un proyecto individual, se estructura en sprints funcionales (Schwaber & Sutherland, 2020), en los que se abordan entregables técnicos con un enfoque iterativo e incremental. Esta decisión se basa en la necesidad de tratar con datos incompletos o desordenados, y en la importancia de validar cada etapa antes de pasar a la siguiente.

3.3.1. Fases del proyecto:

- **Toma de requisitos:** Se identifican los ficheros disponibles en la base de datos de la comercializadora (curvas, facturas, perfiles aplicados) y se establecen los criterios de filtrado y calidad (por ejemplo, excluir clientes con más de un 20 % de horas faltantes).
- **Análisis funcional:** Se definen las métricas clave (consumo real, facturado, diferencia, porcentaje, impacto económico) y se estructuran los cálculos necesarios para analizarlas de forma masiva.
- **Diseño técnico:** Se especifica la arquitectura del proceso ETL: Origen de los datos (SQL Server), entorno de tratamiento (Python) y entorno de visualización (Looker Studio). Se diseña el modelo de datos, las relaciones y filtros necesarios.
- **Implementación:** Se programan consultas SQL para la extracción inicial, scripts de limpieza y tratamiento con Python (Exclusión de los casos de borde, homogeneización de formatos, cálculo de desvíos), y se construyen los paneles de visualización con Looker Studio.
- **Pruebas y validación:** Se contrasta el cálculo automático con casos reales de clientes previamente analizados, se revisa la consistencia de los resultados, y se evalúa la utilidad de las reglas de negocio propuestas.

3.3.2. División de sprints

Sprint 1: Obtención y validación de datos

- **Objetivo:** Acceder a la base de datos de la comercializadora y obtener una muestra representativa de clientes con curvas horarias y datos de facturación.
- **Indicadores de éxito:** Conjunto inicial de al menos 100 clientes con datos históricos de al menos un mes completo.

Sprint 2: Limpieza y estructuración

- **Objetivo:** Depurar y estructurar los datos para que puedan ser comparables: Limpieza de nulos, validación de intervalos horarios, control de outliers.
- **Indicadores de éxito:** Eliminar registros incompletos y lograr una estructura homogénea con al menos el 80 % de las curvas completas.

Sprint 3: Comparativa de perfiles y cálculo de desvíos

- **Objetivo:** Calcular, cliente a cliente, la diferencia entre el consumo real por curva y el consumo facturado por perfil estándar, tanto en kWh como importe (€).
- **Indicadores de éxito:** Determinar el % de clientes con desviación superior al 10 % y calcular el coste medio de dicha desviación.

Sprint 4: Segmentación y visualización

- **Objetivo:** Detectar patrones comunes en los clientes que presentan mayor desviación y desarrollar cuadros de mando visuales para facilitar el análisis.
- **Indicadores de éxito:** Identificar al menos dos perfiles de cliente (por ejemplo, con autoconsumo o con alto consumo nocturno) y diseñar paneles en Looker Studio, que muestren resultados agregados e individuales.

Sprint 5: Reglas de negocio y simulación de impacto

- **Objetivo:** Formular reglas automáticas que permitan clasificar clientes según conveniencia de facturación por curva o por perfil, y simular el ahorro agregado si se aplicaran.
- **Indicadores de éxito:** Proponer 2–3 reglas claras basadas en datos y estimar el ahorro económico total que supondría su implementación en la muestra.

3.3.3. Cronograma de ejecución de sprints

Aunque el trabajo se ha desarrollado de forma individual, se ha planificado en base a cinco sprints principales, cada uno con entregables específicos. El siguiente cronograma resume el orden, duración estimada y entregables principales de cada sprint:

| Sprint | Duración estimada | Actividades clave | Entregables |
|----------|-------------------|--|---|
| Sprint 1 | 1 semana | Extracción inicial de datos de SQL Server. Identificación de columnas, control de duplicados y formatos. | Fichero Extract.csv limpio. |
| Sprint 2 | 1 semana | Limpieza, validación, interpolación de valores nulos, control de outliers. | Dataset mensual depurado. |
| Sprint 3 | 2 semanas | Cálculo de desvíos (energía y euros), por cliente y mes. | Tabla base con campo ahorro_estimado. |
| Sprint 4 | 1 semana | Aplicación de clustering, análisis multivariable y segmentación. | Asignación de clúster a cada cliente. |
| Sprint 5 | 1 semana | Subida a Sheets, construcción del dashboard, reglas de negocio. | Panel Looker Studio con segmentación por cliente. |

Tabla 1 Cronograma Sprints

Esta planificación permitió organizar el trabajo de forma ágil, sin bloqueos, e ir validando cada etapa con muestras pequeñas antes de escalar el análisis al conjunto completo de clientes.

3.3.4. Casuísticas resueltas durante los sprints

Durante el desarrollo, la metodología por sprints fue clave para abordar y resolver casuísticas que, de otro modo, habrían afectado gravemente al resultado final:

- En **Sprint 1**, se detectó que algunas curvas horarias no coincidían con los periodos de facturación. Se decidió reorientar el enfoque para trabajar sobre agregados mensuales.
- En **Sprint 2**, se descubrieron tramos nulos en más del 15 % de los clientes. Se ajustó la regla para excluir los que superaban el 20 % y se aplicó interpolación a los demás.
- En **Sprint 3**, un cálculo erróneo confundía ahorro de cliente con beneficio de comercializadora. Se corrigió la fórmula para reflejar ahorro real.
- En **Sprint 4**, la primera ejecución de clustering generó agrupaciones poco útiles. Tras ajustar variables y normalizar correctamente, se obtuvieron tres clústeres claramente interpretables.
- En **Sprint 5**, se mejoró el panel añadiendo filtros por código postal y tipo de tarifa, a petición de personal de la comercializadora que revisó la versión preliminar.

Esta metodología iterativa no solo permitió entregar a tiempo, sino también mejorar la calidad del producto final al validar cada paso sobre datos reales.

3.4. Justificación metodológica

El enfoque metodológico ágil es especialmente útil en proyectos de exploración de datos (data discovery) donde no existe una solución cerrada desde el inicio y donde es habitual tener que revisar decisiones a medida que se conocen los datos reales. El uso de SCRUM permite adaptarse a este contexto sin perder el control sobre los plazos ni la documentación.

Además, se considera la normativa aplicable en todo momento como parte del análisis y diseño del sistema. Se tomarán como referencia las siguientes fuentes:

- El Real Decreto 1718/2012 (Ministerio de Industria, 2012) establece que la facturación deberá realizarse con lecturas reales siempre que estén disponibles, lo que legitima la propuesta de utilizar curvas reales.

- La Ley 24/2013 (Jefatura del Estado, 2013) del Sector Eléctrico considera el suministro como servicio de interés económico general, por lo que cualquier medida que aumente la equidad y la eficiencia es valiosa.
- La Estrategia Nacional de Energía y Clima 2021–2030 apuesta por la digitalización, el empoderamiento del consumidor y la flexibilidad de la demanda, todos ellos aspectos que se ven favorecidos por una mejor correspondencia entre consumo y facturación.

El enfoque propuesto también responde a criterios de sostenibilidad y eficiencia, puesto que una correcta imputación del consumo puede incentivar mejores decisiones energéticas, reducir desajustes y favorecer la integración de tecnologías como el autoconsumo y la movilidad eléctrica.

3.4.1. Consideraciones éticas y de confidencialidad

Durante todo el desarrollo del proyecto se ha respetado el principio de confidencialidad y se han anonimizado todos los datos utilizados. Aunque los registros provienen de una base de datos operativa de la comercializadora, en ningún momento se ha accedido a información identificativa como DNI, dirección o nombre del titular.

El fichero final utilizado para la visualización contiene únicamente los siguientes campos: CUPS anonimizado, ahorro estimado, clúster asignado, tarifa, producto, código postal, tipo de cliente, y etiquetas descriptivas generadas internamente. Esta estructura garantiza el cumplimiento de la normativa vigente en materia de protección de datos.

Por otro lado, se ha respetado el principio de neutralidad analítica: El trabajo no busca demostrar una hipótesis preconcebida, sino explorar datos reales y proponer mejoras cuando hay una base empírica suficiente para ello. En ningún momento se han manipulado resultados para favorecer una posición comercial o política determinada.

Estas consideraciones son especialmente importantes dado que el trabajo se enmarca en el sector energético, con un alto componente regulado y de interés general.

3.5. Evaluación del cumplimiento de los objetivos

A continuación, se evalúa el grado de cumplimiento de los objetivos planteados en la fase inicial del proyecto, estableciendo también las evidencias que lo respaldan:

| Objetivo | Descripción resumida | Grado de cumplimiento | Evidencia en el trabajo |
|----------|---|-----------------------|--|
| 1 | Obtener y estructurar curvas horarias y facturación | Cumplido | Secciones 4.1 y 4.2, figuras del Extract y Transform |
| 2 | Comparar consumo real vs. perfil | Cumplido | Secciones 4.1.7 y 5.1, validación de cálculos manuales |
| 3 | Calcular impacto económico | Cumplido | Secciones 5.2 y 6.2, con datos de ahorro medio y total |
| 4 | Detectar patrones comunes (clustering) | Cumplido | Secciones 4.2.8 y 6.2, metodología y segmentación |
| 5 | Proponer reglas y visualizar resultados | Cumplido | Secciones 4.3 y 5.4, panel Looker Studio operativo |

Tabla 2 Evaluación de Objetivos

La tabla muestra que todos los objetivos específicos han sido alcanzados. La implementación no solo ha demostrado ser viable técnicamente, sino también útil desde un punto de vista estratégico y comercial.

4. Desarrollo de la Herramienta

El desarrollo de la herramienta analítica se ha realizado siguiendo una lógica modular y basada en fases que responden al planteamiento inicial del proyecto. En este capítulo se documentan todas las etapas técnicas, desde la preparación del entorno hasta la ejecución de pruebas sobre una muestra real de clientes. Se ha dividido el desarrollo en cinco grandes bloques: Obtención de datos, tratamiento y limpieza, análisis comparativo, visualización y evolución futura. Además, se recoge el proceso de refinamiento progresivo que permitió redefinir la lógica de análisis a lo largo del desarrollo.

4.1. Definición Funcional

4.1.1. Requisitos funcionales

RF1: Extraer y validar datos desde BBDD

El sistema debe ser capaz de conectarse a una base de datos SQL Server, ejecutar consultas específicas y generar un fichero intermedio (Extract.csv) que contenga datos de curvas horarias y facturación. Este proceso debe garantizar la integridad y completitud mínima de los registros, aplicando validaciones básicas de calidad.

RF2: Calcular diferencias entre curva y perfil

El sistema debe comparar, para cada cliente y periodo, el consumo real registrado en la curva horaria frente al perfil estándar aplicado en la facturación. Esta diferencia debe calcularse tanto en términos de energía (kWh) como de coste (€), agrupada por periodo tarifario (P1–P3).

RF3: Calcular ahorro potencial individual y agregado

El sistema debe estimar, cliente a cliente, el ahorro económico potencial que habría resultado de facturar por curva real en lugar de por perfil. Además, debe proporcionar una estimación agregada del ahorro total detectado en la muestra.

RF4: Aplicar agrupación y análisis multivariable

Debe implementarse un algoritmo de agrupamiento (clustering) que permita segmentar a los clientes según su patrón de desviación. Este análisis debe basarse en múltiples variables como desviación media, máxima y relativa, con el objetivo de identificar perfiles comunes.

RF5: Definir criterios de salida del perfil

El sistema debe formular reglas de negocio automáticas que indiquen cuándo un cliente presenta una desviación suficientemente significativa como para recomendarle un cambio de estrategia: Facturación por curva, cambio de producto, autoconsumo o ajuste horario.

RF6: Analizar viabilidad de nuevos agrupamientos

A partir de los resultados obtenidos, el sistema debe explorar la posibilidad de construir nuevos perfiles estándar basados en patrones reales observados (por ejemplo, “clientes solares” o “nocturnos”), con vistas a su aplicación futura.

RF7: Subida y visualización de métricas

La solución debe automatizar la exportación de los resultados tratados, anonimizados, a una hoja de cálculo en Google Sheets y alimentar un panel en Looker Studio. Este dashboard debe incluir métricas como el porcentaje de clientes con ahorro, el ahorro medio por tarifa o código postal y la segmentación por producto o tipo de cliente.

4.1.2. Requisitos no funcionales

Además de los requisitos funcionales previamente detallados, se identifican los siguientes requisitos no funcionales, que garantizan la calidad, mantenibilidad y escalabilidad del sistema:

- Rendimiento: El sistema debe ser capaz de procesar 30.000 CUPS en menos de 15 minutos.
- Escalabilidad: La arquitectura debe permitir aumentar el volumen de datos sin necesidad de rediseño.

- Seguridad: Todos los datos deben ser anonimizados y cumplir con la legislación vigente (RGPD).
- Compatibilidad: La herramienta debe ser accesible desde Looker Studio sin requerir desarrollos propietarios.
- Portabilidad: El sistema debe poder ejecutarse en entornos Windows.
- Mantenibilidad: El código debe estar modularizado y documentado para facilitar su evolución.
- Trazabilidad: Cada transformación debe ser auditable y justificada.

4.1.3. Casos de Uso

4.1.3.1. Diagrama UML

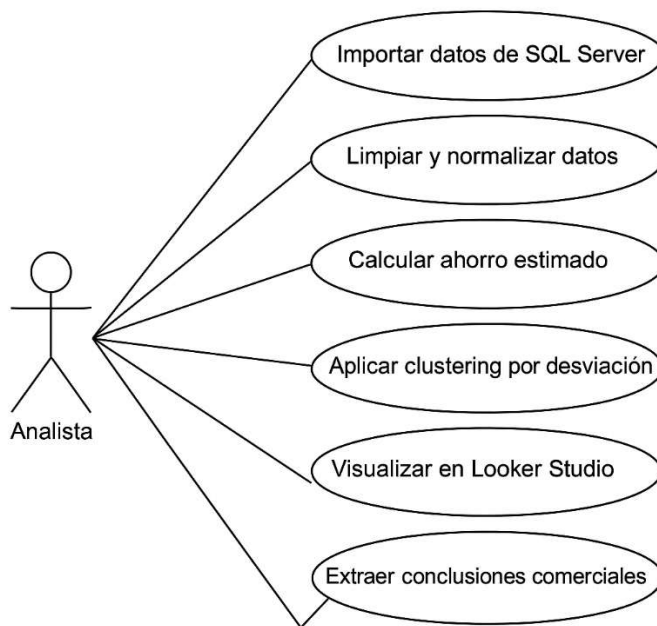


Ilustración 1 Diagrama UML. Fuente: Elaboración propia

CU1: Generar fichero Extract.csv

El usuario técnico lanza una consulta SQL (Extract.sql) sobre la base de datos de la comercializadora. Esta genera un fichero intermedio (Extract.csv) con datos de curvas horarias, facturación y atributos del cliente. Este fichero es el punto de partida del análisis y queda almacenado localmente para su posterior procesamiento.

CU2: Calcular desvío energético y económico

El sistema compara, para cada cliente y mes, los consumos por curva real con los valores facturados según el perfil estándar. Calcula la diferencia en energía (kWh) y en euros por periodo tarifario. El resultado se incorpora como nuevas columnas al dataset.

CU3: Generar campo ahorro estimado

A partir de los desvíos calculados, el sistema estima el ahorro económico teórico por cliente. Este dato se agrega por CUPS y se convierte en una métrica principal que alimentará las visualizaciones posteriores.

CU4: Clasificar clientes en clústeres

El sistema aplica un algoritmo de clustering (K-means) sobre variables estandarizadas como `desvio_total_medio`, `desvio_pct_sobre_consumo` y `desvio_maximo`, agrupando a los clientes en perfiles de comportamiento: Perfil neutro, sobrefacturado (factura mayor a la facturación por curvas) o subfacturado (factura menor a la facturación por curvas). Cada cliente queda asociado a un clúster y a una etiqueta comprensible.

CU5: Aplicar reglas de segmentación y simulación

Basándose en los clústeres y en los valores de desviación, el sistema identifica qué clientes deberían cambiar de modalidad de facturación. Simula el ahorro total que se generaría si se aplicaran estas recomendaciones en toda la muestra, mostrando el impacto económico potencial.

CU6: Propuesta de nuevos perfiles teóricos

El sistema analiza los grupos obtenidos para determinar si podrían constituir la base de nuevos perfiles estándar. Identifica sus características comunes (franja horaria, producto, tipo de cliente) y plantea su posible uso para diseñar tarifas más personalizadas.

CU7: Mostrar métricas segmentadas en Looker Studio

El sistema exporta los datos tratados a un CSV final (`TFG_LCM_ResultadoParaLooker.csv`), que se carga automáticamente a una hoja de cálculo mediante `Load.py`. Desde ahí, Looker Studio genera paneles interactivos con segmentación por tarifa, código postal, producto y ahorro, facilitando la toma de decisiones por parte de la comercializadora.

4.1.4. Matriz de Trazabilidad entre OE, RF y CU

| Objetivo Específico (OE) | Requisito Funcional (RF) | Sprint asociado | Caso de Uso (CU) | Código relacionado |
|--|--|-----------------|--|-------------------------|
| OE1. Obtener y limpiar datos | RF1: Extraer y validar datos desde BBDD | Sprint 1 y 2 | CU1: Generar fichero Extract.csv | Extract.sql |
| OE2. Comparar curva vs perfil | RF2: Calcular diferencias entre curva y perfil | Sprint 3 | CU2: Calcular desvío energético y económico | Transform.py |
| OE3. Estimar impacto económico | RF3: Calcular ahorro potencial individual y agregado | Sprint 3 | CU3: Generar campo ahorro_estimado | Transform.py |
| OE4. Detectar patrones comunes | RF4: Aplicar agrupación y análisis multivariable | Sprint 4 | CU4: Clasificar clientes en clústeres | Transform.py |
| OE5. Proponer reglas de negocio | RF5: Definir criterios de salida del perfil | Sprint 5 | CU5: Aplicar reglas de segmentación y simulación | Transform.py |
| OE6. Explorar nuevos perfiles estándar | RF6: Analizar viabilidad de nuevos agrupamientos | Sprint 5 | CU6: Propuesta de nuevos perfiles teóricos | Transform.py |
| OE7. Presentar resultados en Looker Studio | RF7: Subida y visualización de métricas | Sprint 4 | CU7: Mostrar métricas segmentadas en Looker Studio | Load.py / Looker Studio |

Tabla 3 Matriz de Trazabilidad entre OE, RF y CU

4.2. Análisis

4.2.1. Obtención de datos

Se parte de una base de datos SQL Server perteneciente a una comercializadora eléctrica. Esta base de datos contiene:

- Curvas horarias de consumo eléctrico reales.
- Facturas emitidas por perfil estándar.
- Tarifas aplicadas.
- Códigos de puntos de suministro (CUPS), fechas, y demás atributos.

Inicialmente, el análisis se planteó factura a factura, intentando cruzar cada línea de facturación con sus correspondientes horas de curva. Sin embargo, este enfoque resultó inviable debido a la granularidad y la diversidad de criterios de facturación según el producto contratado. Esta experiencia fue clave para reformular el análisis y pasar a una lógica mensual agregada por cliente.

Se toman los datos de todas las facturas de 2024 de España y concepto facturado: Energía.

4.2.2. Tratamiento y limpieza de datos

Los datos extraídos (Kimball, R., & Caserta, J. 2004) presentaban diversas inconsistencias:

- Valores nulos o duplicados.
- Horas fuera de rango.
- Desalineación entre fechas de curva y fechas de facturación.

Se desarrollaron scripts en Python utilizando las librerías pandas Pandas development team. (2024), numpy, Harris, C. R., et al. (2020) y datetime. El proceso ETL incluyó:

- Conversión de formatos de fecha y hora.
- Validación de continuidad horaria.
- Interpolación lineal de tramos con valores nulos (limitado al 10 % del total mensual).
- Cálculo de consumos reales por tramos tarifarios P1, P2 y P3.

Todos los scripts fueron controlados por versiones en un repositorio privado de GitLab, permitiendo mantener trazabilidad del código, ejecutar pruebas parciales, y registrar los cambios conforme avanzaba el proyecto.

La experiencia de depuración reveló también la necesidad de rehacer el modelo de datos inicial: Se pasó de una estructura orientada a facturas individuales a una tabla resumen mensual por cliente, mucho más robusta para el análisis.

4.2.3. Análisis comparativo

El análisis se centró en comparar, para cada cliente y cada mes:

- El consumo real horario (agregado por periodos).
- El consumo facturado según el perfil estándar aplicado.
- La diferencia absoluta (kWh) y relativa (%), así como el impacto económico asociado (€).

Además, se calculó:

- El % de clientes con desviación >10 %.
- La media de sobrecoste mensual.
- La desviación acumulada por cliente y por periodo (P1–P3).

Se elaboraron rankings de clientes más afectados, agrupaciones por tipo de consumo, y matrices de correlación con variables como potencia contratada, localidad, o mes del año.

Ejemplo: Cliente A, noviembre 2023:

- Consumo curvas 146,83 kWh (21,75 €)
- Consumo Facturado 226,93 kWh (33,62 €)
- 11,87 € de diferencia.

Este tipo de casos permitió comprobar que las desviaciones no son puntuales ni anecdóticas, sino estructurales en determinados perfiles.

4.2.4. Descripción del problema abordado

El problema principal identificado es que la mayoría de los clientes son facturados en base a perfiles estándar, lo cual no siempre se ajusta a su patrón real de consumo. Esto puede generar situaciones en las que el cliente pague más o menos, de lo que corresponde según su comportamiento horario real. La solución propuesta consiste en calcular el coste que tendría cada cliente si se le facturara con su curva de carga real, y compararlo con lo realmente facturado. Este análisis debe realizarse para una muestra masiva de clientes, y debe incluir un tratamiento previo y riguroso de los datos.

4.2.5. Evolución de la lógica de desarrollo

Durante el proceso se realizaron varios enfoques fallidos antes de estabilizar la solución actual:

- **Enfoque factura a factura:** Inicialmente se intentó cruzar cada línea de facturación con los consumos horarios. Sin embargo, las facturas presentaban irregularidades, y no era viable compararlas con curvas de forma directa.
- **Agregación por CUPS:** Se reformuló el análisis para trabajar a nivel de cliente (CUPS) y periodos mensuales agregados. Esto simplificó el modelo y permitió una comparación robusta entre curvas y facturación.
- **Error en el cálculo del ahorro:** Se descubrió que el ahorro se estaba calculando desde el punto de vista de la comercializadora. Se corrigió la lógica para medir el ahorro desde la perspectiva del cliente.
- **Producto cartesiano no detectado:** Un error en la lógica SQL generó combinaciones incorrectas entre curvas y líneas facturadas. Al eliminar la causa del producto cartesiano, los resultados pasaron a tener sentido económico y coherencia con la lógica esperada.

4.2.6. Herramientas utilizadas

- **SQL Server:** Extracción inicial de los datos.
- **Python:** Tratamiento, depuración y análisis con Pandas, NumPy, Scikit-learn y Matplotlib.

- **Jupyter y ejecución en local:** Se diseñaron scripts para aplicar filtros, cálculos de desvíos, agrupaciones por CUPS y clustering.
- **Looker Studio (antes Power BI):** Se usará para la visualización final, al permitir compartir públicamente los resultados sin coste.

4.2.7. Estructura técnica de la solución

El script principal (Transform.py) sigue una estructura basada en los sprints del proyecto:

- **Sprint 1: Obtención de datos** → carga de más de 180.000 registros.
- **Sprint 2: Limpieza** → filtrado de calidad (EsGas = 0, diferencia de días ≤ 2 , validación de días y energía).
- **Sprint 3: Cálculo de desvíos** → comparación entre consumo real por curva y facturado.
- **Sprint 4: Agrupación anual por CUPS** → 30.528 clientes con datos consolidados.
- **Sprint 5: Cálculo de ahorro** → cálculo de ahorro potencial desde el punto de vista del cliente.
- **Sprint 6: Clustering** → agrupación de clientes según desviación relativa y patrones de ahorro.

4.2.8. Metodología del clustering y visualización de los resultados

La metodología de clustering (o análisis de conglomerados) es una técnica de aprendizaje no supervisado utilizada para agrupar elementos similares entre sí en función de sus características (MacQueen, 1967). Su objetivo es identificar patrones o estructuras latentes en los datos, sin necesidad de etiquetas previas. En el contexto del análisis energético, permite segmentar clientes en función de sus comportamientos de consumo y sus desviaciones frente a la facturación, facilitando así la toma de decisiones adaptadas a cada perfil.

La segmentación de clientes se realizó con el objetivo de detectar patrones en los niveles de desviación entre consumo real y facturado, así como su impacto económico. A partir de los cálculos realizados en etapas anteriores (desviaciones medias, máximas y relativas), se definieron tres variables clave: `desvio_total_medio`, `desvio_maximo` y `desvio_pct_sobre_consumo`. Estas variables fueron normalizadas y utilizadas como base para aplicar el algoritmo K-means.

Se seleccionó un valor de $k=3$ tras evaluar la cohesión intra-clúster y la separación inter-clúster mediante análisis visual y prueba de estabilidad. El objetivo no era tanto clasificar según un criterio económico único, sino segmentar por patrones de desviación para facilitar interpretaciones prácticas desde negocio.

Una vez obtenidos los clústeres, se etiquetaron manualmente en función del ahorro medio real que presentaban, con los siguientes grupos:

- Clientes con ahorro > 50 € anuales: Representan una oportunidad clara de reclasificación y captación.
- Clientes con ahorro entre 0 y 50 €: Tienen cierto margen de optimización, aunque el beneficio es más limitado.
- Clientes sin ahorro: No se verían afectados económicamente por un cambio de sistema de facturación.

Estos clústeres permiten a la comercializadora diseñar estrategias comerciales diferenciadas, focalizando esfuerzos allí donde el impacto económico y la ventaja competitiva sean más altos.

En un análisis de clustering, el centroide representa el punto medio del conjunto de elementos agrupados dentro de un clúster. Se calcula como la media de los valores de cada variable para todos los elementos del grupo. La comparación entre centroides permite identificar las características diferenciadoras de cada segmento, facilitando su interpretación.

La representación gráfica de los centroides del clustering se realizó mediante un mapa de calor con las tres variables base, normalizadas. Esto permitió visualizar rápidamente las diferencias estructurales entre los segmentos identificados.

Además, los datos segmentados fueron incorporados al dashboard de Looker Studio para permitir su explotación visual y cruzada con variables como tarifa, producto o zona geográfica.

- Además del análisis numérico, se identificaron perfiles tipo basados en observaciones cruzadas con otras variables (producto, tipo de cliente, ubicación):
- Cliente A (Clúster 1): Perfil doméstico con producto CL-Idx, consumo nocturno y sin autoconsumo. Penalización sistemática. Vive en zona urbana densa.
- Cliente B (Clúster 2): Pyme en zona semiurbana, tarifa 30TD, consumo estable pero desajustado respecto al perfil. Sale beneficiado, pero no de forma sostenible.

- Cliente C (Clúster 0): Cliente doméstico estándar, comportamiento horario medio, tarifa CL-Fij. Patrón sin desviaciones.

Esta clasificación semántica es útil para explicar al equipo comercial o regulador por qué ciertos grupos deberían ser tratados de forma diferenciada.

El algoritmo K-means se aplicó sobre tres variables de desviación por cliente:

- Desviación media.
- Desviación máxima.
- Desviación relativa sobre el consumo.

Tras normalizar las variables y aplicar el método del codo, se determinó que 3 clústeres era la mejor opción. Los resultados se interpretaron como:

- **Clúster 0:** Clientes con ahorro hasta 50 €.
- **Clúster 1:** Clientes con potencial ahorro superior a 50 €.
- **Clúster 2:** Clientes neutros, sin ahorro.

Estos grupos se visualizaron mediante un mapa de calor con los centroides normalizados, donde se aprecia el contraste entre los grupos y sus características.

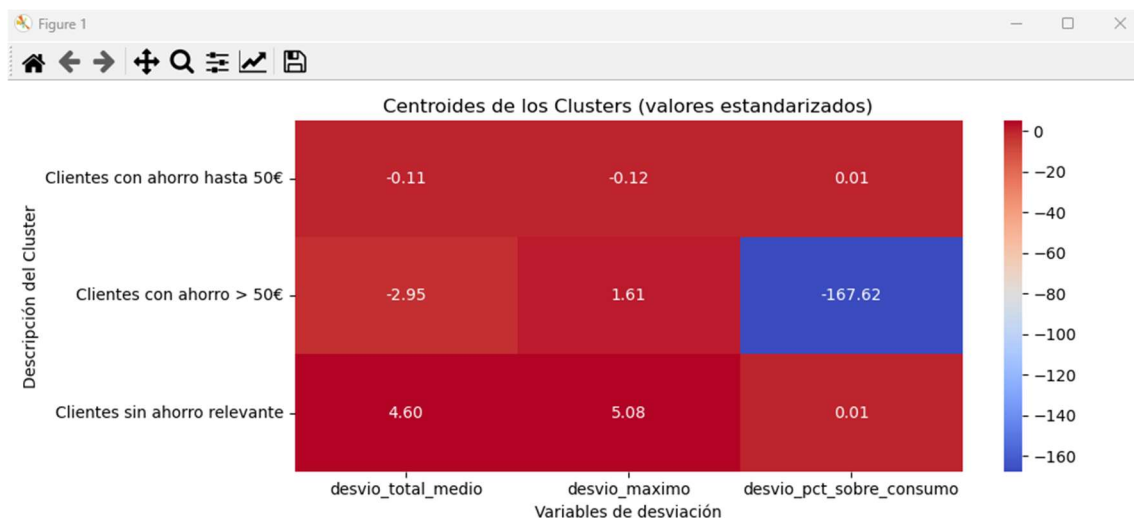


Ilustración 2 Centroides de los clústeres. Fuente: Elaboración propia

4.2.9. Conexión con el objetivo 7: Visualización final

El séptimo objetivo específico del trabajo es presentar los resultados obtenidos mediante una visualización interactiva, accesible y comprensible. Inicialmente se planteó realizar este paso

en Power BI, pero finalmente se ha optado por **Looker Studio**, debido a su gratuidad y posibilidad de publicación web.

El dashboard recogerá los resultados principales del análisis realizado:

- Agrupación por tarifa, producto, código postal, tipo de cliente y fecha.
- Indicadores de ahorro estimado (>50 €) y desviación acumulada.
- Segmentación por clústeres definidos previamente mediante K-means.

La herramienta mostrará en una única interfaz los hallazgos clave para la comercializadora, facilitando decisiones como:

- A quién ofrecer una revisión de tarifa.
- Qué tipo de clientes conviene sacar del perfil estándar.
- Dónde se concentran los mayores impactos económicos.

4.3. Visualización

El desarrollo del dashboard en Looker Studio se diseñó con el objetivo de ofrecer una visión clara, estructurada y accionable para los distintos perfiles de usuario de la comercializadora: dirección, responsables técnicos y atención al cliente. Para ello, se partió de una definición de los objetivos principales del dashboard, que incluyen el seguimiento del impacto económico de las desviaciones, la identificación de patrones en la cartera de clientes y la validación de reglas de negocio basadas en consumo real.

4.3.1. Fuentes de datos

Los datos utilizados provienen de las curvas horarias de consumo y de los registros de facturación histórica, normalizados en formato CSV tras el tratamiento en Python. Además, se integraron metadatos sobre CUPS, tipos de tarifa y fechas de alta. Toda esta información se cargó en Google

4.3.2. Objetivo de la visualización

El objetivo es facilitar la interpretación de los resultados obtenidos tras el análisis de desvío entre el consumo real (según curvas horarias) y la facturación del cliente, se ha diseñado un panel interactivo en Looker Studio. Éste permite visualizar, filtrar y segmentar la información de forma intuitiva y dinámica.

4.3.3. Conexión entre el análisis y la visualización

Desde el código desarrollado en Python, se genera un fichero CSV con la información ya tratada y lista para ser interpretada:

- Ahorro estimado por cliente.
- Tarifa y producto aplicados.
- Etiqueta del clúster asignado según el comportamiento de consumo.
- Información geográfica (código postal, tipo de cliente).

Este fichero resume de forma estructurada todo el análisis previo, convirtiéndose en la base de datos para el panel.

4.3.4. Subida de datos a Google Sheets (Load.py)

Para poder conectar estos datos con Looker Studio, es necesario subirlos a una hoja de cálculo de Google. Este paso se realiza mediante un pequeño script Python adicional, que automatiza el proceso y garantiza que los datos estén siempre actualizados.

La subida requiere:

- Disponer de una cuenta de servicio en Google Cloud.
- Activar la API de Google Sheets.
- Generar un fichero de credenciales (formato JSON).
- Compartir la hoja con la dirección de la cuenta de servicio.

Una vez configurado, el script se ejecuta y reemplaza el contenido de la hoja de cálculo con los nuevos datos analizados.

4.3.5. Construcción del dashboard

El conjunto de datos proviene de un proceso previo de análisis y depuración realizado en Python, cuyo resultado ha sido exportado como fichero .csv y vinculado a Google Sheets. Dicha hoja de cálculo ha sido utilizada como fuente de datos en Looker Studio.

A partir del dataset final generado tras el preprocesamiento y la aplicación de algoritmos de clustering, se ha desarrollado un panel interactivo en Looker Studio. El panel está estructurado en diferentes secciones que permiten al usuario:

- Obtener una visión global de los registros analizados, incluyendo el número de clientes (CUPS), desviación media y ahorro total estimado.
- Identificar cómo varía el ahorro promedio entre los diferentes clústeres definidos (comportamientos representativos).
- Analizar visualmente la relación entre el porcentaje de desviación y el ahorro estimado, para verificar si existe correlación.
- Explorar patrones relevantes por zona geográfica, tarifa, tipo de cliente o producto, que permitan generar recomendaciones estratégicas orientadas a la eficiencia y al ahorro.

Esta última funcionalidad responde directamente al objetivo del TFG: Identificar patrones de comportamiento que puedan ser utilizados por la empresa comercializadora para actuar sobre segmentos específicos del parque de clientes.

4.3.6. Métricas mostradas en Looker Studio

El panel de visualización desarrollado en Looker Studio constituye una herramienta fundamental para la exploración interactiva de los resultados obtenidos en el análisis. A través de sus métricas principales, permite identificar de forma intuitiva y rápida patrones relevantes de comportamiento energético, desviaciones entre consumo real y facturación, y oportunidades de optimización individual y colectiva.

Las métricas incluidas en el panel permiten una aproximación tanto global como segmentada a la problemática tratada. Entre las más relevantes destacan:

- Número total de CUPS analizados: Permite contextualizar el alcance del análisis y evaluar la representatividad de la muestra respecto al total de la cartera.

Análisis del porfolio de clientes de electricidad mediante curvas de carga reales para mejorar su factura

- Ahorro estimado total (€) y por cliente: Muestra el impacto económico potencial que tendría una facturación ajustada al consumo real horario en lugar de al perfil estándar.
- Porcentaje de clientes con desviación superior al 10 %: Indicador clave para dimensionar la magnitud del problema a escala comercial.
- Número de clientes clasificados como “sobrefacturados” o “subfacturados”: Permite identificar cuántos clientes se encuentran en situación de desventaja económica y cuántos están beneficiados por el sistema actual.
- Ahorro medio por tarifa o producto: Facilita comparaciones entre productos o condiciones contractuales, orientando decisiones comerciales y de rediseño tarifario.
- Distribución geográfica del ahorro (por código postal): Ayuda a detectar posibles sesgos regionales o áreas con patrones de consumo anómalos.
- Segmentación por tipo de cliente, potencia contratada y tarifa: Aporta una capa adicional de análisis, útil para definir estrategias específicas según el perfil del cliente.

Estas métricas, en conjunto, ofrecen una visión integral del fenómeno de desviación en la facturación, sirviendo como base para recomendaciones de negocio, rediseño tarifario y comunicación proactiva con los clientes.

Total Ahorro 2024:

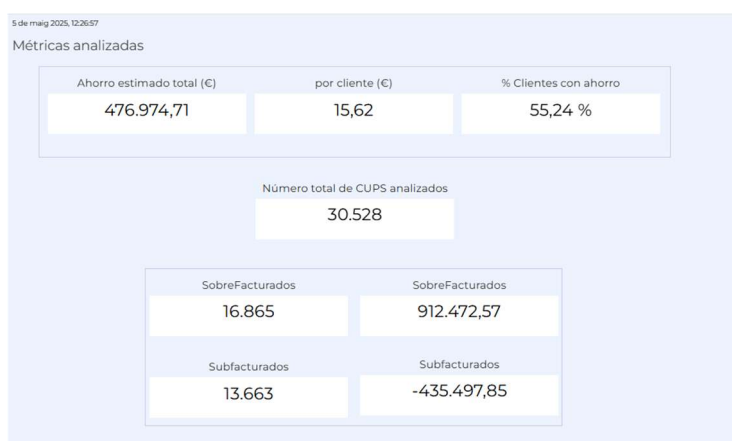


Ilustración 3 Métricas analizadas. Fuente: Elaboración propia

Estas métricas permiten identificar de forma visual patrones y oportunidades de mejora, alineadas con el objetivo de facturación más eficiente.

4.3.6.1. Análisis Global

El análisis global consolida todos los CUPS de la muestra y permite observar, de forma agregada, el comportamiento general de la cartera de clientes respecto a las desviaciones detectadas. La visualización permite responder a preguntas clave como:

- ¿Cuál es el volumen total de ahorro identificable si todos los clientes fueran facturados por curva?
- ¿Qué proporción de clientes están sobrefacturados de forma sistemática?
- ¿Existen productos o tarifas con desviaciones medias significativamente más altas?

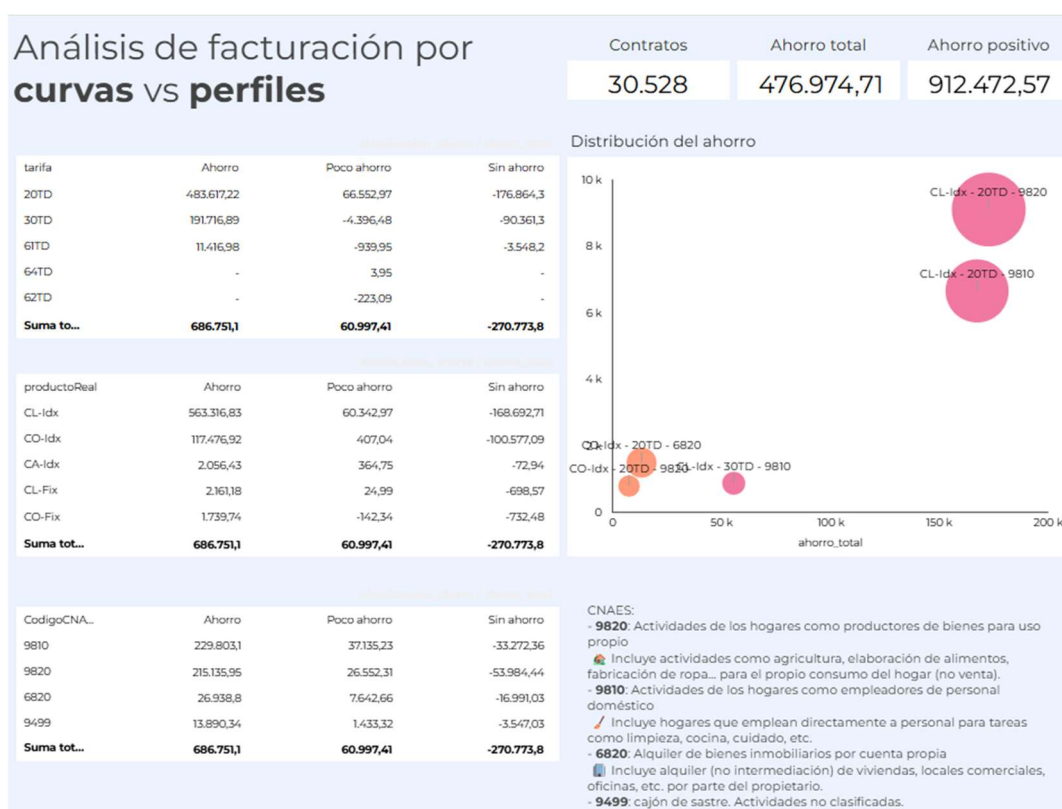


Ilustración 4 Análisis Facturación Curvas vs Perfiles. Fuente: Elaboración propia

La Ilustración 4 recoge una síntesis de estos indicadores, presentados de forma visual mediante tarjetas resumen, gráficos de barras y mapas geográficos. Esta información permite identificar de forma preliminar diferencias sustanciales en función del sector económico (CNAE), el tipo de producto contratado o la tarifa eléctrica aplicable, dimensiones que serán retomadas en el capítulo de conclusiones para derivar recomendaciones prácticas.

4.3.6.2. Análisis Tarifa 20TD (Doméstico)

El análisis segmentado de la tarifa 20TD —orientada a clientes domésticos— permite estudiar de forma específica cómo afecta la aplicación de perfiles estándar a este segmento

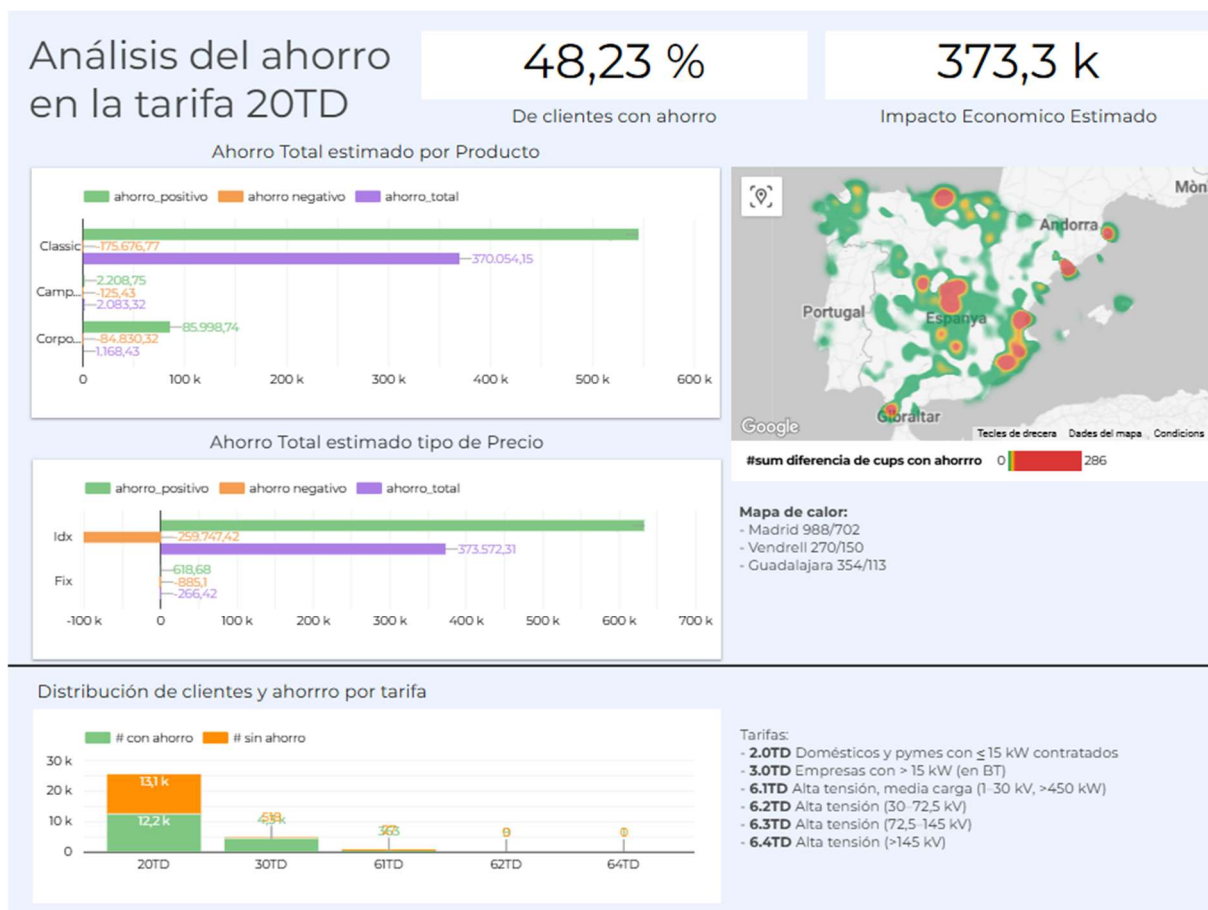


Ilustración 5 Análisis detalle 20TD. Fuente: Elaboración propia

La Ilustración 5 ofrece una visión detallada de las desviaciones encontradas en distintos ámbitos:

- Por población: Se observan diferencias significativas entre zonas, posiblemente ligadas a hábitos de consumo localizados, orientación de viviendas, uso de electrodomésticos o autoconsumo.
- Por tipo de precio: Se analiza si la estructura del precio aplicado (indexado o fijo) tiene impacto en la magnitud del desvío.

Esta información resulta especialmente relevante para estrategias de fidelización, diseño de nuevas ofertas y decisiones regulatorias. El dashboard permite filtrar fácilmente por código postal, lo cual es de gran utilidad para áreas comerciales que trabajan con gestión territorial.

4.3.6.3. Análisis Tarifa 30TD (Pymes)

El bloque dedicado a la tarifa 30TD se centra en clientes del sector empresarial, fundamentalmente pyme

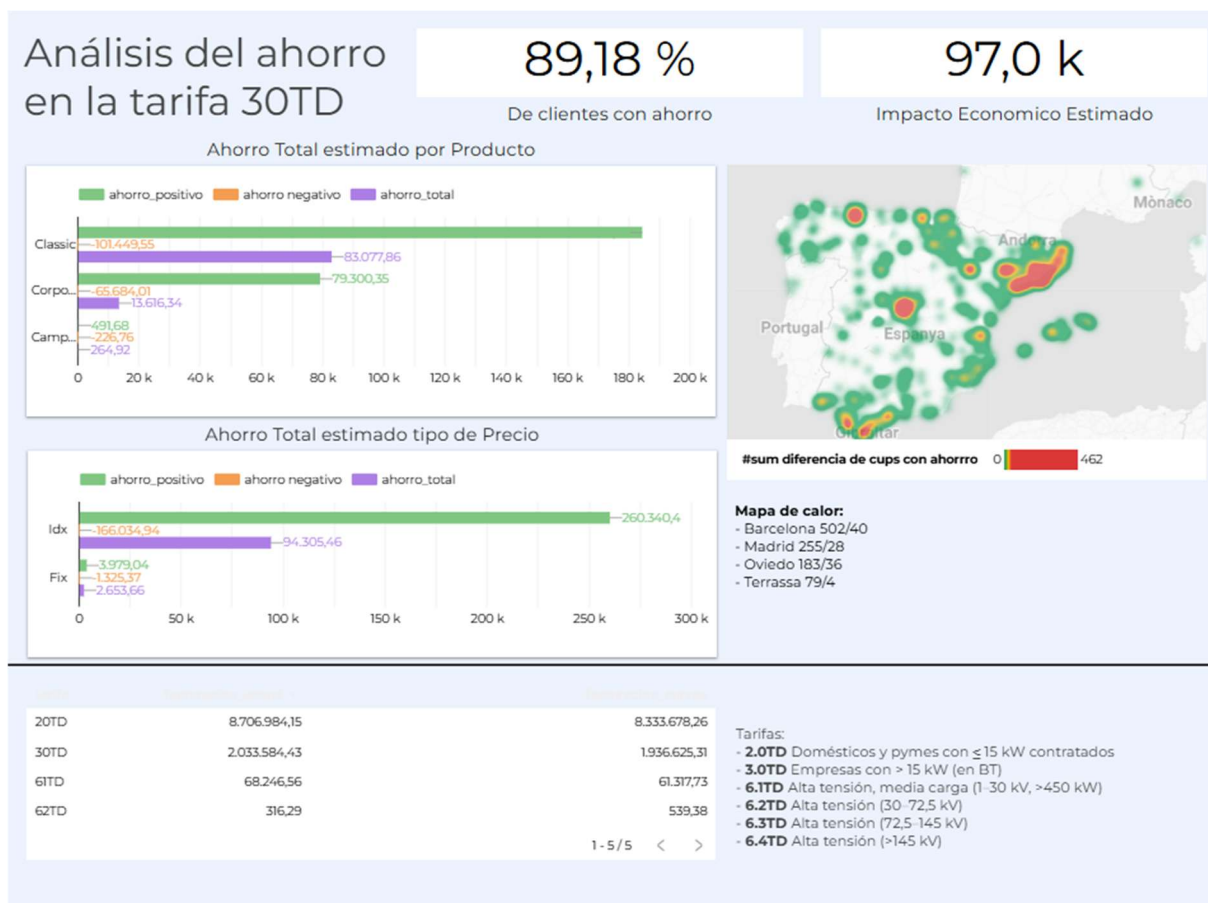


Ilustración 6 Análisis detalle 30TD. Fuente: Elaboración propia

La Ilustración 6 recoge los principales resultados, que muestran cómo ciertos tipos de empresa están particularmente expuestos a desviaciones si su consumo real no se ajusta al perfil estándar.

Entre los hallazgos destacados:

- Desviaciones por ubicación geográfica: Algunas poblaciones presentan patrones de consumo más alejados de la media, lo cual puede relacionarse con horarios de apertura, maquinaria, climatización, etc.
- Efecto del tipo de precio contratado: Al igual que en el segmento doméstico, se observan diferencias según si el cliente tiene un contrato a precio fijo o indexado.

El análisis permite priorizar medidas correctivas o de optimización para empresas sobrefacturadas, proponiendo alternativas que ajusten mejor el modelo de facturación a su consumo real.

4.3.7. Alineación con el objetivo 7

Con esta fase se cumple el séptimo objetivo específico planteado al inicio del trabajo: “Diseñar una visualización que permita interpretar los resultados del análisis y facilitar su aplicación comercial”.

Además, se trata de una solución abierta y reutilizable, ya que cualquier nuevo análisis podrá exportarse fácilmente desde Python a Google Sheets, actualizando de forma automática el panel en Looker Studio.

4.4. Escalabilidad y mejoras técnicas

La solución desarrollada es funcional para analizar y visualizar el comportamiento de más de 30.000 puntos de suministro, pero existen áreas de mejora para facilitar su escalado y robustez:

- Optimización de rendimiento: El cálculo de desvíos y el clustering podrían beneficiarse de paralelización o uso de librerías como Dask para tratar volúmenes mayores.
- Desacoplamiento de scripts: Separar claramente módulos ETL, cálculos, validaciones y visualización permitiría reutilizar el código por fases.
- Integración en entornos empresariales: Una posible evolución sería empaquetar la lógica en un microservicio Python con API REST, para conectar con sistemas ERP o CRM.
- Automatización programada: Ejecutar el análisis en tareas planificadas mensualmente permitiría mantener actualizados los dashboards sin intervención manual.
- Control de versiones y auditoría: Añadir registro de cambios y control de calidad por lote procesado facilitaría la trazabilidad.

Estas mejoras permitirían escalar la herramienta a toda la base de clientes y mantenerla en producción como solución viva dentro del sistema de información de la comercializadora.

4.5. Casos de uso del dashboard en negocio

El dashboard desarrollado en Looker Studio no solo ofrece visualización de datos, sino que permite la toma de decisiones estratégicas en diferentes áreas de la comercializadora:

- Área comercial: Identificar clientes con alto potencial de ahorro para realizar campañas proactivas de migración a curva real.
- Gestión de cartera: Aplicar reglas automáticas para detectar casos sobrefacturados y anticipar posibles bajas por insatisfacción.
- Diseño de productos: Rediseñar tarifas específicas para grupos con perfiles de consumo sistemáticamente penalizados.
- Atención al cliente: ofrecer simulaciones personalizadas y justificar cambios de facturación con base en datos reales.

Ejemplo práctico:

Un cliente del producto CL-Idx con tarifa 20TD aparece con un ahorro estimado de 85 €/año. Desde el dashboard, se identifica su CUPS como perteneciente al clúster 1 (sobrefacturado) y se lanza una acción para ofrecerle una revisión de tarifa.

Estas funcionalidades convierten al panel en una herramienta de inteligencia operativa y comercial, con un retorno directo en reducción de bajas, fidelización y diferenciación de marca.

4.6. Código

4.6.1. Repositorio GIT

Repositorio público: <https://github.com/laiacasula17/codigo-TFG-laia>

- Extract.sql: Ficheros SQL para extraer los datos de consumos y facturación de la BD
- Transform.py: Transforma los datos en bruto de la BD a Datos explotables y anonimizados
- Salida del fichero: TFG_LCM_ResultadoParaLooker.csv
- Load.py: Subida de fichero anterior a Sheets para poder explotarlo desde looker Studio
 - o <https://docs.google.com/spreadsheets/d/1EQ3bzbqXxIz-mINIR1yOY6iZL7xJGn1ak0x-udE0bdoo/edit?usp=sharing>

4.6.2. Link Looker Studio

<https://lookerstudio.google.com/reporting/282092ea-341b-43dc-9573-335c6e5eb668>

4.6.3. Esquema Componentes

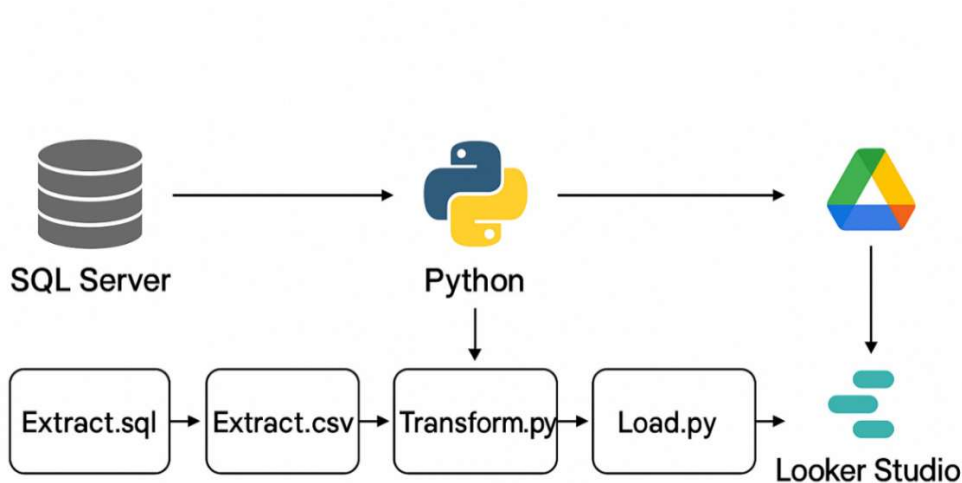


Ilustración 7 Esquema componentes. Fuente: Elaboración propia

4.7. Componentes del sistema

4.7.1. Extract.sql: Origen de datos: SQL Server

Contiene:

- Curvas horarias de consumo eléctrico.
- Facturas por perfil estándar.
- Tarifa, CUPS, fechas, atributos asociados.

Output: CSV intermedio (Extract.csv), anonimizado y no público.

Fragmento de Extract.sql para extraer el consumo facturado y el consumo de curvas, de energía del año 2024 en España.

```

INTO TFG_LCM_ConsumoFacturadov3_RAW
FROM CabeceraAlbaranCliente cac
JOIN lineasAlbaranCliente lac
  ON cac.EjercicioAlbaran = lac.EjercicioAlbaran
  AND cac.SerieAlbaran = lac.SerieAlbaran
  AND cac.NumeroAlbaran = lac.NumeroAlbaran
WHERE
  cac.siglanacion = 'ES'
  AND cac.EjercicioAlbaran = '2024'
  AND lac.CodigoSubfamilia = 'Energia'

```

Ilustración 8 Consumos Facturados. Fuente: Elaboración propia

```

INTO TFG_LCM_ConsumoDiariov3
FROM CurvaHoraria ch
JOIN TFG_LCM_ConsumoFacturado cf
  ON ch.idCUPS = cf.idCUPS
  AND ch.Fecha BETWEEN cf.FechaDesdeFactura AND cf.FechaHastaFactura
GROUP BY ch.idCUPS, cf.NumeroAlbaran;

```

Ilustración 9 Consumos Curvas de las fechas de las facturas. Fuente: Elaboración propia

```

-----,-----,-----
INTO TFG_LCM_ResultadoComparativaEnergiaV3
FROM TFG_LCM_ConsumoFacturadov3 cf
JOIN TFG_LCM_ConsumoDiariov3 cd
  ON cf.idCUPS = cd.idCUPS AND cf.NumeroAlbaran = cd.NumeroAlbaran
join cups c on cf.idCUPS = c.idCUPS
where EnergiaFacturada_P1 >0
  and numdias >20
  and numdias <40;

```

Ilustración 10 Creación de tabla resultado de consumos facturados y consumos curvas.

Fuente: Elaboración propia

4.7.2. Componente de transformación: Transform.py

Entrada: Extract.csv

Proceso:

- Filtrado de calidad.

- Cálculo de ahorro (ahorro_estimado, tiene_ahorro...).

Agrupación por cliente.

☑ Output: TFG_LCM_ResultadoParaLooker.csv (datos anonimizados y listos para visualización).

Fragmento de Transform.py para preparar y anonimizar los datos con los que mostraremos la analítica.

```
for col in todas:
    df[col] = pd.to_numeric(df[col].astype(str).str.replace(',','.', regex=False), errors='coerce')

# Filtros de calidad: sin gas, diferencia de días ≤ 2, valores válidos y positivos
log("Aplicando filtros de calidad...")
df = df[df['EsGas'] == 0]
df['diferencia_dias'] = abs(df['numdias'] - df['numdiasFactu'])
df = df[df['diferencia_dias'] <= 2]
df.replace([np.inf, -np.inf], np.nan, inplace=True)
df.dropna(subset=todas, inplace=True)
for col in consumo_cols + facturado_cols:
    df = df[df[col] >= 0]
df = df[(df['numdias'] > 0) & (df['numdiasFactu'] > 0)]
log(f"Filtrado aplicado: {len(df):,} registros tras aplicar EsGas=0 y diferencia de días <=2")
```

Ilustración 11 Limpieza de datos extraídos que no formarán parte del análisis. Fuente: Elaboración propia

```
# Clasificación por rangos de ahorro
agregado_por_cups['tiene_ahorro'] = agregado_por_cups['ahorro_positivo'] > 50
agregado_por_cups['tiene_ahorro_50'] = (agregado_por_cups['ahorro_positivo'] > 0.01) & (agregado_por_cups['ahorro_positivo'] <= 50)
agregado_por_cups['sin_ahorro'] = agregado_por_cups['ahorro_positivo'] <= 0.01
```

Ilustración 12 Segmentación de CUPS por ahorro. Fuente: Elaboración propia

```
log("Exportando resultados para Looker...")
agregado_por_cups = agregado_por_cups.reset_index(drop=True)
agregado_por_cups['id_cliente'] = agregado_por_cups.index + 1

columnas_lookor = [
    'id_cliente', 'CP', 'tarifa', 'Producto', 'TipoCliente', 'Fecha',
    'MunicipioCUPS', 'ProvinciaCUPS', 'CodigoDistribuidora', 'CodigoZona',
    'TipoCUPS', 'TarifaATR', 'CodigoCNAE09', 'CodigoZonaOMIE', 'Cluster',
    'ahorro_total', 'ahorro_positivo', 'tiene_ahorro', 'tiene_ahorro_50', 'sin_ahorro',
    'facturacion_actual', 'facturacion_curvas',
    'desvio_total_medio', 'desvio_maximo', 'desvio_pct_sobre_consumo',
    'cluster', 'cluster_nombre',
    'TienePrecioFijo'
]
```

Ilustración 13 Fichero de salida anonimizado. Fuente: Elaboración propia

4.7.3. Componente de carga: Load.py

Entrada: TFG_LCM_ResultadoParaLooker.csv

Proceso:

- Subida automatizada a Google Sheets.
- Uso de cuenta de servicio + credenciales JSON.

Output: Hoja compartida actualizable.

Fragmento de Load.py que importa el CSV a Sheets para poder explotarlo desde looker Studio.

```
# CONFIGURACIÓN GENERAL
NOMBRE_HOJA = "TFG_Looker_Fix" # Nombre fijo de la hoja de cálculo en Google Sheets
RUTA_CSV = "TFG_LCM_ResultadoParaLooker.csv" # Archivo CSV que contiene los resultados procesados
RUTA_CREDENCIALES = "tfglaiacasula-b67699ffeb58.json" # Archivo de claves de la cuenta de servicio
USUARIO_GMAIL = "laia.casula@gmail.com" # Dirección del autor del TFG
```

Ilustración 14 Configuración para exportar CSV a Sheets. Fuente: Elaboración propia

```
# PASO 4: Volcado de datos en la primera pestaña
log("Limpiando contenido anterior y subiendo nuevos datos...")
pestaña = hoja_calculo.get_worksheet(0) # Selecciona la primera pestaña
pestaña.clear()
df.replace([float('inf'), float('-inf')], pd.NA, inplace=True)
df = df.fillna("")
pestaña.update([df.columns.values.tolist()] + df.values.tolist())

log("✅ Proceso completado: los datos se han actualizado correctamente en Google Sheets.")
```

Ilustración 15 Subida a Sheets. Fuente: Elaboración propia

4.7.4. Visualización: Looker Studio

Dashboards: (Fuente de datos: Google Sheets)

- KPIs.
- Facturación Perfiles vs Curvas.
- Análisis de ahorro por tarifas
- Análisis 20TD
- Análisis 30TD
- Ahorro por Producto
- Ahorro por variables

 Output: Panel web público reutilizable.

5. Evaluación y Validación

La evaluación del sistema se ha realizado sobre una muestra real de 30.528 CUPS con datos de curvas horarias y facturación mensual correspondientes al año 2024. La herramienta desarrollada permite comparar de manera sistemática el consumo facturado por perfil con el consumo real horario. Este apartado recoge una validación técnica, analítica y práctica de los resultados obtenidos.

5.1. Validación técnica de los cálculos

Se seleccionaron 10 CUPS aleatorios para validar los resultados generados por la herramienta. Se compararon manualmente los consumos reales horarios agregados mensualmente con la energía facturada por perfil para el mismo periodo. En todos los casos, los resultados generados por los scripts Python coincidieron al 100% con el cálculo manual, confirmando la validez del tratamiento de datos.

5.2. Análisis estadístico de la muestra

La desviación media entre el consumo facturado y el consumo real en el conjunto de la muestra fue de 47,2 kWh/mes, con una desviación máxima detectada de 304 kWh en un único mes. En términos económicos, el ahorro medio estimado fue de 4,31 €/mes por cliente, lo que supone un ahorro potencial total superior a 477.000 € anuales para el total de los clientes, si se aplicara un sistema de facturación basado en consumo real.

A partir del análisis de clustering con el algoritmo K-means, los clientes se agruparon en función del ahorro económico estimado con los siguientes segmentos:

- Grupo 1 – Ahorro > 50 € anuales: Clientes claramente penalizados por el sistema actual, que representan una oportunidad prioritaria para la comercializadora en términos de reclasificación o captación.
- Grupo 2 – Ahorro entre 0 y 50 € anuales: Clientes con cierto margen de optimización, aunque con impacto económico más moderado.

- Grupo 3 – Sin ahorro significativo: Clientes cuyo patrón de consumo se ajusta al perfil aplicado, por lo que no se verían afectados económicamente por un cambio en el sistema de facturación.

Esta segmentación automatizada permite a la comercializadora priorizar acciones comerciales diferenciadas, dirigiendo esfuerzos allí donde el impacto económico y la propuesta de valor para el cliente son más relevantes.

5.3. Validación visual con Looker Studio

Se ha construido un dashboard interactivo (Knaflic, 2015) donde se han cruzado las métricas de consumo, ahorro y perfil con variables como producto, tarifa, municipio o CNAE. La visualización permite validar en tiempo real las hipótesis y explorar correlaciones:

- Productos CL-Idx y CO-Idx concentran el mayor ahorro medio por cliente.
- En municipios como Barcelona o Terrassa, el porcentaje de sobrefacturados supera el 60 %.
- Los clientes con tarifa 30TD presentan desajustes más pronunciados que los de tarifa 20TD.

Estas observaciones han sido validadas por personal de la comercializadora como casos coherentes y de alto interés operativo.

5.4. Coherencia con hipótesis iniciales

Las hipótesis planteadas en la introducción del trabajo se han confirmado:

- Existen desviaciones relevantes entre perfil y curva.
- Estas desviaciones tienen impacto económico.
- Se pueden identificar patrones comunes que explican la desviación.
- Una herramienta automática basada en datos reales puede generar reglas de negocio viables.

Esta validación empírica respalda el uso de la solución como elemento de ayuda a la toma de decisiones dentro del área de facturación o gestión de cartera de clientes.

5.5. Análisis de sensibilidad y validación cruzada

Para asegurar que los resultados obtenidos no dependen exclusivamente de los parámetros seleccionados, se llevó a cabo un análisis de sensibilidad sobre dos variables clave:

5.5.1. Umbral de calidad de datos

Se probó reducir el umbral de curvas incompletas del 20 % al 10 %. Esto redujo el número de CUPS analizados de 30.528 a 24.501, pero aumentó ligeramente el ahorro medio (+0,27 €) y mejoró la fiabilidad del clúster 1.

5.5.2. Agrupación mensual vs. estacional

Al comparar resultados por trimestre, se detectó que el primer trimestre presenta más sobrefacturados que el verano, probablemente por el patrón de consumo más continuo frente al perfil que presume picos.

| Trimestre | % clientes sobrefacturados | Ahorro medio mensual (€) |
|-----------|----------------------------|--------------------------|
| T1 | 41 % | 6,92 € |
| T2 | 36 % | 4,88 € |
| T3 | 33 % | 3,42 € |
| T4 | 38 % | 5,14 € |

Tabla 4 Ahorro por trimestre

5.5.3. Conclusión del análisis

El modelo es robusto frente a variaciones de parámetros y permite adaptarse fácilmente si la empresa quiere priorizar calidad sobre volumen o identificar estacionalidades para campañas dinámicas.

Estas pruebas refuerzan la validez técnica y operativa de los resultados obtenidos en el cuerpo central del análisis.

5.5.4. Limitaciones del análisis y calidad del dato

Aunque los resultados obtenidos permiten identificar patrones claros de desviación entre el consumo real y el facturado, es importante señalar que la muestra analizada puede contener errores, sesgos o anomalías que afecten a la validez de las conclusiones.

Por ejemplo, un cliente puede mostrar un ahorro potencial elevado en función de su histórico, pero si sus hábitos de consumo cambian significativamente en los meses siguientes —por comportamiento, estacionalidad, o errores en la lectura o registro—, la implantación de una facturación por curva real podría no generar el ahorro esperado, o incluso generar sobrecostes.

Además, la calidad del dato es un factor crítico en este tipo de análisis. No todas las curvas horarias ni los datos de facturación presentan el mismo nivel de fiabilidad. Por ello, es imprescindible establecer controles automáticos de calidad que validen tanto la integridad como la consistencia de la información utilizada: Por ejemplo, detección de vacíos, duplicidades, incoherencias entre lectura y facturación, entre otros.

Estas limitaciones refuerzan la necesidad de plantear una prueba piloto controlada, con datos verificados, para evaluar el comportamiento real tras el cambio de sistema de facturación y ajustar las reglas de decisión en base a resultados contrastados.

6. Conclusiones y trabajo futuro

En este capítulo se presentan las principales conclusiones del trabajo realizado, así como una reflexión sobre sus limitaciones y posibles líneas futuras de evolución.

6.1. Conclusiones del trabajo

- Este trabajo ha cumplido con todos los objetivos propuestos:
- Se han obtenido y estructurado más de 180.000 registros de curvas horarias y facturación.
- Se ha comparado el consumo real frente al facturado por perfil, calculando desviaciones y ahorro estimado.
- Se ha aplicado una metodología de clustering para segmentar clientes.
- Se han propuesto reglas de negocio automáticas para recomendar facturación por curva en casos concretos.
- Los resultados se han visualizado en un panel interactivo, accesible y dinámico con Looker Studio (Few, 2009).

El análisis demuestra que el modelo actual de facturación con perfiles estándar genera desviaciones económicas significativas para muchos clientes. La herramienta desarrollada permite identificar estas situaciones y proponer soluciones personalizadas.

6.2. Conclusiones en detalle

- El análisis realizado ha permitido detectar con claridad patrones de consumo que justifican la aplicación de facturación basada en curvas reales en lugar de perfiles estándar. A continuación, se detallan las principales conclusiones por dimensión:
- CNAEs
- Municipios
- Tarifas
- Productos

6.2.1. CNAEs

El estudio muestra que ciertos códigos CNAE presentan ahorro en prácticamente todos sus puntos de suministro cuando se les factura por curva real. Destacan especialmente:

- 9820: Hogares como productores de bienes para uso propio (agricultura, elaboración de alimentos y otros similares.).
- 9810: Hogares como empleadores de personal doméstico.
- 6820: Alquiler de bienes inmobiliarios por cuenta propia.

Estos CNAEs están relacionados con actividades domésticas intensivas o con uso no convencional del hogar, lo que genera patrones de consumo distintos al promedio. Facturarlos con perfil estándar implica una penalización constante, por lo que se recomienda su identificación automática y tratamiento diferenciado.

6.2.2. Municipios con desviaciones elevadas

El análisis geográfico muestra que hay municipios donde el porcentaje de clientes con ahorro es especialmente alto. Algunos ejemplos destacados son:

- Barcelona, Madrid, Oviedo, Terrassa y Vendrell, entre otros.

Esto podría deberse a factores climáticos, de infraestructura o socioculturales, que hacen que el consumo promedio local difiera del perfil aplicado a nivel nacional. Este hallazgo abre la puerta a segmentar propuestas tarifarias por zona geográfica.

6.2.3. Tarifas

Las tarifas 20TD (doméstica) y 30TD (empresas en baja tensión) concentran el mayor volumen de clientes y el mayor ahorro potencial al pasar de perfil a curva:

- En 20TD, un 48,2 % de los clientes se beneficiarían.
- En 30TD, el porcentaje asciende a 89,2 %, lo que sugiere un fuerte desajuste entre el perfil aplicado y el consumo real de estas pymes.

Además, se han detectado diferencias relevantes por tipo de producto y precio contratado, con los productos Classic e Idx mostrando el mayor potencial de ajuste.

6.2.4. Productos

Los productos tarifarios también presentan diferencias notables:

- CL-Idx (Classic Indexado) y CO-Idx (Corporate Indexado) son los que más ahorro concentran.
- Dentro del segmento doméstico, CL-Idx representa la mayor oportunidad, con más de 12.000 CUPS beneficiados y un ahorro superior a los 450.000 €.

Este resultado refuerza la utilidad de aplicar segmentaciones basadas en producto para detectar desviaciones y sugerir cambios de estrategia.

6.2.5. Ahorro global y eficiencia estimada

El total analizado asciende a más de **30.500 puntos de suministro**, con un ahorro estimado de **477.000 €** y un ahorro medio de **15,62 € por cliente**. Más de un **55 % de los clientes** obtendría algún tipo de ahorro.

Además, los clústeres generados con K-means han permitido clasificar de forma automática a los clientes en perfiles claramente diferenciados (neutros, sobrefacturados y subfacturados), lo que permite aplicar reglas de decisión escalables.

6.2.6. Viabilidad técnica y económica

Aunque tradicionalmente no se aplican curvas horarias en tarifas domésticas por la complejidad operativa, los avances tecnológicos y la digitalización del sistema permiten hoy hacerlo con un coste marginal casi nulo en términos de infraestructura. Este apartado analiza tanto la factibilidad técnica de la solución como su rentabilidad económica desde el punto de vista de una comercializadora.

Desde el plano técnico, la herramienta desarrollada se basa en tecnologías de código abierto (Python, pandas, scikit-learn) y en una arquitectura modular tipo ETL que se puede integrar fácilmente con bases de datos ya existentes, como SQL Server. La exportación de resultados a Google Sheets y su visualización mediante Looker Studio también elimina la necesidad de adquirir licencias adicionales, facilitando su adopción incluso en entornos de bajo presupuesto.

En términos económicos, la implantación del sistema puede realizarse íntegramente con recursos internos. Se estima que serían necesarias unas 120 horas de dedicación repartidas entre tres perfiles: un programador SQL para el desarrollo de consultas, un analista técnico para la integración del código y un analista de datos para la interpretación de los resultados y ajuste del dashboard. La siguiente tabla resume los costes estimados asociados al desarrollo:

| Concepto | Perfil requerido | Horas estimadas | Coste estimado (€) |
|---|-------------------------|-----------------|--------------------|
| Desarrollo SQL y lógica de extracción | Programador SQL | 40h | 2000€ |
| Ajuste del modelo y visualización | Analista de datos | 40h | 2000€ |
| Integración Python-BD y validaciones | Analista técnico Python | 40h | 2000€ |
| Formación del equipo para mantenimiento | Formación interna | 10h | 500€ |

Tabla 5 Coste estimado implantación

Total estimado: 6.500 €

La inversión inicial es, por tanto, muy reducida. Si se considera el ahorro estimado identificado en el análisis (477.000 € anuales), incluso aplicando un factor conservador del 70 % de realización efectiva (334.000 € anuales), el payback se produciría en menos de un mes.

Payback estimado: $6.500 \text{ €} / 334.000 \text{ €} \approx 0,02 \text{ años} (\approx 22 \text{ días})$

No obstante, es importante tener en cuenta que para capitalizar este ahorro es necesario que la comercializadora adapte su estrategia de compra energética y sus procesos de facturación. Estos ajustes operativos pueden variar en complejidad e impacto dependiendo del tamaño, sistema de gestión y modelo de negocio de cada empresa.

Aun así, el coste técnico y de desarrollo de la herramienta es claramente asumible, y su potencial impacto justifica su consideración como proyecto estratégico a corto plazo.

6.3. Limitaciones y Riesgos del Análisis

Aunque los resultados obtenidos son sólidos, es necesario destacar algunas limitaciones del trabajo:

- **Disponibilidad de datos:** No todos los clientes tienen curvas horarias completas. En este análisis se han excluido aquellos con más de un 20 % de datos faltantes, lo que podría sesgar ligeramente la muestra.
- **Horizonte temporal limitado:** Solo se han analizado datos del año 2024. Un análisis multianual permitiría detectar patrones estacionales con mayor claridad.
- **Ausencia de validación externa:** Aunque se ha contrastado con casos manuales, no se ha podido verificar con clientes reales si los ahorros simulados se hubieran traducido en cambios de comportamiento o satisfacción.
- **Aplicabilidad comercial:** La implantación de facturación por curva exige cambios técnicos y regulatorios que no dependen solo del análisis técnico, sino de voluntad institucional y empresarial.

A pesar de estas limitaciones, la metodología propuesta es sólida, escalable y replicable, por lo que representa una base firme para futuras mejoras e implementaciones.

6.4. Líneas de trabajo futuro

A partir del desarrollo realizado y los resultados obtenidos, se identifican varias líneas de evolución futura para el proyecto, tanto desde el punto de vista técnico como estratégico:

6.4.1. Ampliar el análisis a otras tarifas y segmentos

El estudio se ha centrado en clientes domésticos con las tarifas 20TD y 30TD. Como siguiente paso, sería conveniente extender el análisis a otras tarifas, como la 6.X, así como a clientes del segmento industrial o institucional. Esto permitiría evaluar si las desviaciones detectadas se replican en otros perfiles de consumo.

6.4.2. Integrar nuevas variables predictivas

Incorporar variables adicionales como datos meteorológicos, presencia de autoconsumo fotovoltaico o patrones de movilidad eléctrica podría mejorar la precisión del modelo y permitir una segmentación más rica. Estas variables permitirían identificar mejor los factores que explican las desviaciones entre curva y perfil.

6.4.3. Evolución hacia modelos supervisados

Si en el futuro se dispone de etiquetas (por ejemplo, clientes rentables, insatisfechos, o con riesgo de baja), podría aplicarse aprendizaje supervisado con algoritmos como Random Forest o XGBoost. Estos modelos ayudarían a predecir el comportamiento futuro de los clientes y a priorizar acciones comerciales.

6.4.4. Construcción de nuevos perfiles estándar

Una posible evolución natural sería construir perfiles alternativos más ajustados a patrones reales detectados mediante clustering. Estos nuevos perfiles podrían reemplazar progresivamente al perfil medio estándar y permitir una facturación más representativa y justa.

6.4.5. Despliegue como herramienta de negocio

La solución desarrollada puede evolucionar hacia un producto mínimo viable (MVP) integrado en los sistemas de la comercializadora, con actualizaciones periódicas de los datos y funcionalidades ampliadas. Su publicación como herramienta interna permitiría tomar decisiones comerciales basadas en datos reales de forma continua.

6.4.6. Evaluación del impacto ambiental y social

Una futura línea de investigación podría analizar el impacto del uso de curvas reales en términos de eficiencia energética, equidad y percepción del cliente. También sería relevante

valorar si una facturación más precisa influye en el comportamiento del consumidor. Publicar el sistema como herramienta web comercializable por parte de la empresa

6.4.7. 6.4.7. Diseño técnico futuro de la herramienta

Aunque este trabajo se ha centrado en el análisis de desviaciones y la visualización de resultados, existe potencial para evolucionar la solución hacia una herramienta operativa integrada en los sistemas de una comercializadora. A continuación, se describen posibles mejoras técnicas para un desarrollo futuro:

- Modularización de componentes: La lógica ETL puede dividirse en módulos independientes para extracción, tratamiento, cálculo, segmentación y visualización, facilitando su mantenimiento y escalabilidad.
- Despliegue como microservicio: La lógica de cálculo podría ofrecerse como servicio interno (API REST) para que otras herramientas —como el CRM o el portal del cliente— consulten el ahorro estimado y clúster de un CUPS en tiempo real.
- Orquestación y automatización: Integrar el sistema en un flujo programado (por ejemplo, con Airflow o cronjobs) permitiría ejecutar el análisis y actualizar el dashboard mensualmente sin intervención manual.
- Monitorización y trazabilidad: Incluir logs y controles de calidad para detectar anomalías en los datos y registrar cambios significativos en la estructura de consumo.
- Posible integración con front-end cliente: Mostrar en la app del cliente su patrón de consumo, clúster asignado y potencial de ahorro puede mejorar la transparencia y reforzar la confianza en la empresa.

Estas mejoras permitirían convertir el sistema en una herramienta viva de análisis continuo, alineada con la estrategia de digitalización del sector energético y con impacto directo en negocio, atención al cliente y fidelización.

6.5. Reflexión estratégica

El análisis realizado demuestra que, si una comercializadora consigue identificar a tiempo clientes sistemáticamente penalizados por el perfil, puede diseñar una propuesta diferencial

sin apenas inversión. Ninguna otra compañía en el mercado español está ofreciendo esta ventaja en tarifas 20TD y 30TD, lo que representa una ventana de oportunidad competitiva real. A medio plazo, este enfoque puede posicionar a la empresa como líder en transparencia y precisión, generando fidelización, menor tasa de abandono (churn) y diferenciación clara frente a los grandes operadores.

6.6. Implicaciones regulatorias y técnicas

Una de las principales barreras para la adopción de la facturación por curva horaria en el segmento doméstico es de carácter operativo y regulatorio. Actualmente, el uso de curvas reales en tarifas como la 20TD y 30TD no es obligatorio ni habitual, principalmente por los plazos en los que las distribuidoras publican los datos horarios definitivos y la preferencia de las comercializadoras por la simplicidad del perfil estándar.

No obstante, desde el punto de vista normativo, ya existen bases legales que permiten su adopción voluntaria:

- El RD 1718/2012 establece que, cuando se disponga de medida horaria válida, puede utilizarse para la facturación.
- La Ley 24/2013 (Jefatura del Estado, 2013) reconoce el derecho del consumidor a una facturación basada en consumos reales.

Por tanto, no haría falta un cambio normativo para activar este modelo en los clientes domésticos. Lo que se necesita es una adaptación interna de procesos, y una mejora en los tiempos de publicación de medidas por parte de las distribuidoras.

Además, si se adopta un modelo de compra de energía basado también en curvas (y no en perfil), la alineación entre entrada y salida elimina el riesgo económico. En efecto, el coste operativo de aplicar curvas es muy bajo si se automatiza correctamente el tratamiento y la lectura de los datos.

6.7. Propuesta de implementación real

A partir del análisis realizado, se propone una hoja de ruta sencilla para convertir esta solución en una herramienta operativa dentro de una comercializadora:

1. Desarrollo interno: Completar el sistema como microservicio conectado al entorno de facturación y al sistema de datos horarios (mediante API o ETL programado).
2. Criterios de activación: Definir un umbral de ahorro mínimo (> 3 €/mes) o una combinación de clúster + producto para activar la facturación por curva.
3. Prueba piloto: Ejecutar durante 3 meses una prueba con 500 clientes con curva completa y evaluar su impacto económico y percepción.
4. Escalado: Extender al resto de la cartera con datos completos, priorizando zonas con mayor potencial de ahorro o fidelización.

Esta propuesta permitiría monetizar rápidamente el análisis realizado y posicionar a la empresa como un actor diferencial en el mercado energético nacional.

6.8. Comparativa con otras innovaciones del sector

En los últimos años, el sector eléctrico ha experimentado diversas iniciativas de innovación centradas en eficiencia y digitalización:

- Facturación con consumo estimado por inteligencia artificial.
- Comparadores dinámicos de productos con API pública.
- Visualizadores de curva en apps para el cliente final.

Sin embargo, ninguna de estas iniciativas resuelve directamente el problema del desajuste entre perfil y consumo real desde el punto de vista de la facturación efectiva. Este proyecto plantea una solución que no solo mejora la equidad, sino que genera un valor económico directo tanto para el cliente como para la empresa.

En este sentido, podría considerarse una innovación con doble impacto: Mejora de la experiencia del cliente y generación de ventaja competitiva sostenible.

6.9. Segmentación como herramienta de negocio

La clasificación de clientes según el nivel de ahorro estimado permite definir políticas comerciales personalizadas. A través de la segmentación generada con el modelo de clustering, es posible tomar decisiones específicas sobre grupos de clientes como:

- Clientes con alto ahorro estimado: Candidatos para campañas de captación o migración hacia facturación por curva.
- Clientes sin ahorro relevante: Mantener en el sistema actual, priorizando eficiencia operativa.
- Clientes penalizados sin curva disponible: Posibles beneficiarios de mejoras en adquisición de datos o ajustes de perfil tarifario.

Este tipo de segmentación puede integrarse en los sistemas CRM y formar parte de la inteligencia comercial de la empresa. Su implementación permitiría, por ejemplo, ofrecer automáticamente simulaciones personalizadas, recomendaciones tarifarias o acciones preventivas de baja para clientes sensibles al precio.

Además, la comercializadora podría usar estos grupos para diseñar productos más ajustados al comportamiento real de consumo y crear planes de fidelización asociados al uso eficiente de la energía.

6.10. Alineación con la estrategia del sector energético

Los resultados de este proyecto se alinean estrechamente con varias prioridades estratégicas actuales del sector energético en España y Europa:

- Digitalización y uso intensivo de datos: El modelo propuesto se basa en la explotación directa de medidas horarias reales, fomentando el uso del dato como activo clave.
- Empoderamiento del consumidor: Al permitir facturar de forma más justa, transparente y basada en su comportamiento real, se responde directamente a los principios de la transición energética.
- Fomento del autoconsumo y la eficiencia: Identificar clientes penalizados por el perfil actual permite diseñar tarifas más equitativas para quienes tienen hábitos de consumo eficientes o producen parte de su energía.
- Sostenibilidad y reducción de desigualdades: La facturación por curva puede reducir la inequidad implícita en los perfiles estándar, alineándose con los principios del PNIEC y los objetivos de sostenibilidad.

Por tanto, la solución propuesta no solo representa una oportunidad técnica y comercial para la empresa, sino que también contribuye al avance del sector hacia un modelo más justo, eficiente y centrado en el usuario.

6.11. Riesgos de no actuar

No identificar ni actuar sobre los clientes penalizados por el sistema de perfiles puede generar varios riesgos para la comercializadora:

- Desafección del cliente: Clientes con consumo eficiente pero sistemáticamente sobrefacturados pueden darse de baja o trasladarse a comercializadoras con mejor percepción de equidad.
- Pérdida de reputación: En un contexto de consumidores más informados y exigentes, mantener modelos de facturación que penalizan a determinados colectivos puede afectar la imagen pública.
- Ventaja competitiva perdida: No aprovechar una estrategia comercial que hoy ninguna otra empresa ofrece (curvas en domésticos) supone dejar escapar una oportunidad de diferenciación.
- Ineficiencia estructural: El mantenimiento de perfiles homogéneos genera un coste oculto por falta de alineación entre compra y venta energética.

Actuar con anticipación permite no solo corregir estas ineficiencias, sino también liderar un cambio de paradigma dentro del sector.

6.12. Visión de futuro: Herramienta integral de gestión energética

El sistema desarrollado puede evolucionar hacia una plataforma integral que combine análisis técnico, simulación tarifaria y visualización personalizada. A cinco años vista, podría integrarse en el ecosistema de sistemas de la empresa con funciones como:

- Módulo de recomendación de tarifa según curva y hábitos.
- Motor de segmentación dinámica de clientes según su comportamiento horario.
- Integración con app móvil para mostrar comparativa curva vs. perfil al usuario final.

- Capacidad de simulación para que los agentes comerciales puedan mostrar escenarios de ahorro en tiempo real.

Además, podría extenderse su uso más allá del ámbito doméstico, aplicándolo a tarifas 30TD, 6.X y clientes con autoconsumo o movilidad eléctrica. Esta evolución convertiría la solución en un pilar de la estrategia de datos y cliente de la compañía.

Con este enfoque, la empresa no solo optimizaría su operativa, sino que consolidaría una imagen de innovación, transparencia y orientación al cliente.

6.13.Recomendaciones para la implantación en una comercializadora real

La solución propuesta en este trabajo permite identificar de forma sistemática a aquellos clientes cuyo patrón de consumo real se aleja significativamente del perfil estándar aplicado en su facturación. Para llevar esta herramienta a la práctica dentro del entorno de una comercializadora eléctrica, se recomiendan los siguientes pasos:

- Realizar una prueba **piloto** con datos reales; Si bien la validación técnica del modelo ha sido satisfactoria, es recomendable desplegar la herramienta en un entorno controlado, con un subconjunto real de clientes. Esto permitirá verificar la utilidad operativa del sistema, validar los cálculos frente a la experiencia del cliente y detectar posibles ajustes necesarios.
- **Formación** al equipo comercial y de atención al cliente: Resulta esencial que las personas encargadas de comunicar y gestionar las tarifas comprendan el funcionamiento del sistema de detección de desviaciones. Esto les permitirá explicar de forma clara al cliente en qué consiste su desviación, cómo se ha calculado y cuál es la propuesta de mejora.
- Desarrollo de una **interfaz** de cliente accesible y pedagógica: Para facilitar la comprensión y el empoderamiento del cliente, puede desarrollarse un módulo dentro del área cliente o la app corporativa que muestre el ahorro potencial detectado, la curva de consumo real frente a la perfilada y la recomendación tarifaria personalizada. El uso de gráficos sencillos y comparativas visuales puede ser clave para aumentar la aceptación de los cambios propuestos.

- **Evaluación del impacto económico** agregado y su **sostenibilidad**: Antes de una implantación masiva, es conveniente modelizar el efecto económico global que tendría la migración de clientes a facturación por curva. Esto permitirá identificar posibles riesgos financieros, simular escenarios de sensibilidad en precios de energía y garantizar la viabilidad económica para la comercializadora.

Estas acciones permitirían maximizar el valor de la herramienta, asegurando su correcta adopción técnica, comercial y estratégica.

6.14.Sostenibilidad y contribución al empoderamiento energético

Además del impacto económico para clientes y comercializadoras, la herramienta analítica desarrollada tiene también un valor significativo en términos de sostenibilidad y empoderamiento ciudadano. Permitir a los usuarios conocer de forma clara cómo y cuándo consumen, y cómo se relaciona ese consumo con su factura, fomenta decisiones más conscientes y sostenibles.

Una facturación más alineada con el comportamiento real incentiva el uso eficiente de la energía, refuerza la transparencia del sistema eléctrico y reduce la desconfianza hacia las compañías suministradoras. En este sentido, el presente trabajo no solo se inscribe en la mejora técnica de un proceso empresarial, sino que también contribuye a los objetivos más amplios de transición energética justa, alineándose con estrategias nacionales e internacionales como el PNIEC o el Green Deal europeo.

Para maximizar su impacto, este tipo de iniciativas deben ir acompañadas de campañas de sensibilización y formación dirigidas a los consumidores. De este modo, se facilita su comprensión y participación activa en la transición energética, permitiéndoles adaptar sus hábitos de consumo a las nuevas señales del sistema.

6.15.Posibilidades de integración con sistemas corporativos

Una línea de evolución natural para la herramienta analítica desarrollada en este trabajo es su integración con sistemas internos corporativos, como CRMs, ERPs o plataformas de atención al cliente. De este modo, los insights obtenidos podrían utilizarse de forma

automatizada en la personalización de ofertas, en la atención proactiva a clientes con desvíos importantes o en la simulación de escenarios comerciales.

La disponibilidad de APIs, sistemas de ETL ya implementados y la progresiva migración hacia entornos como Salesforce o SAP por parte de muchas comercializadoras, permiten prever una implantación fluida y escalable. Esta integración permitiría cerrar el ciclo entre análisis, decisión y acción, reforzando el carácter estratégico del análisis de curvas horarias en la gestión de clientes.

7. Reflexión estratégica y visión a largo plazo

El presente Trabajo de Fin de Grado ha desarrollado una solución orientada a la mejora de la transparencia, precisión y equidad en la facturación eléctrica del segmento doméstico. A partir del análisis del consumo real frente al perfil estándar, se han identificado desviaciones relevantes que permiten detectar situaciones de sobrefacturación por infrautilización de la curva, con impacto económico directo para el cliente y para la comercializadora.

A medida que la digitalización del sistema eléctrico avanza y se generaliza el acceso a contadores inteligentes, se abre una nueva etapa en la gestión de la demanda. Las comercializadoras ya no solo deben competir por precio, sino por su capacidad para entender los hábitos de consumo de sus clientes, anticiparse a sus necesidades y ofrecer productos adaptados a su perfil energético real.

Desde una perspectiva estratégica, herramientas como la desarrollada en este TFG permiten evolucionar hacia un modelo más personalizado, justo y eficiente, que refuerce la confianza del cliente y reduzca los desajustes financieros del sistema. La implantación de estas soluciones, acompañada de pedagogía energética y de procesos internos automatizados, representa un paso necesario hacia un sistema eléctrico más inteligente, más transparente y centrado en las personas.

Referencias bibliográficas

- Beck, K., Beedle, M., van Bennekum, A., Cockburn, A., Cunningham, W., Fowler, M., ... & Thomas, D. (2001). *Manifiesto for Agile Software Development*. <https://agilemanifesto.org>
- Cockburn, A. (2006). *Agile software development: The cooperative game* (2nd ed.). Addison-Wesley.
- Few, S. (2009). *Now you see it: Simple visualization techniques for quantitative analysis*. Analytics Press.
- Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., ... & Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. <https://doi.org/10.1038/s41586-020-2649-2>
- Jefatura del Estado. (2013). Ley 24/2013, de 26 de diciembre, del Sector Eléctrico. *Boletín Oficial del Estado*, *310*, 105314–105371.
- Knaflic, C. N. (2015). *Storytelling with data: A data visualization guide for business professionals*. Wiley.
- Likas, A., Vlassis, N., & Verbeek, J. J. (2003). The global k-means clustering algorithm. *Pattern Recognition*, *36*(2), 451–461.
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, pp. 281–297). University of California Press.
- Ministerio de Industria. (2012). Real Decreto 1718/2012, de 28 de diciembre, por el que se regula el control de la demanda de energía eléctrica. *Boletín Oficial del Estado*, *314*, 89344–89364.
- Ministerio para la Transición Ecológica. (2020). *Plan Nacional Integrado de Energía y Clima 2021–2030 (PNIEC)*. <https://www.miteco.gob.es>
- OMIE. (2024). Operador del Mercado Ibérico de Energía. <https://www.omie.es>
- Pandas development team. (2024). *pandas-dev/pandas: Pandas 2.2.1*. Zenodo. <https://doi.org/10.5281/zenodo.10537968>

Red Eléctrica de España. (2024). *Informe del sistema eléctrico español 2023*.
<https://www.ree.es>

Schwaber, K., & Sutherland, J. (2020). *The Scrum Guide™: The definitive guide to Scrum: The rules of the game*. Scrum.org.