



UNIVERSIDAD POLITÉCNICA SALESIANA
SEDE CUENCA
CARRERA DE INGENIERÍA ELÉCTRICA

“CARACTERIZACIÓN DE CURVAS DE CARGA EN CONSUMIDORES
DE LA EMPRESA ELÉCTRICA REGIONAL CENTRO SUR C.A.
MEDIANTE TÉCNICAS DE AGRUPAMIENTO DE DATOS”

Trabajo de titulación previo a la obtención
del título de Ingeniero Eléctrico

AUTORES: MARIA FERNANDA CABRERA ARTEAGA
WILLIAMS ANDRES MENDOZA VITONERA

TUTOR: ING. JOHNNY XAVIER SERRANO GUERRERO, PH.D.
CO-TUTOR: ING. CESAR ANDRES PATIÑO CHITACA
ING. SERGIO PATRICIO ZAMBRANO ASANZA

CERTIFICADO DE RESPONSABILIDAD Y AUTORÍA DEL TRABAJO DE TITULACIÓN

Nosotros, Maria Fernanda Cabrera Arteaga con documento de identificación N° 1105795007 y Williams Andres Mendoza Vitonera con documento de identificación N° 0706283595; manifestamos que:

Somos los autores y responsables del presente trabajo; y, autorizamos a que sin fines de lucro la Universidad Politécnica Salesiana pueda usar, difundir, reproducir o publicar de manera total o parcial el presente trabajo de titulación.

Cuenca, 04 de abril del 2023

Atentamente,



Maria Fernanda Cabrera Arteaga
1105795007



Williams Andres Mendoza Vitonera
0706283595

**CERTIFICADO DE CESIÓN DE DERECHOS DE AUTOR DEL TRABAJO DE
TITULACIÓN A LA UNIVERSIDAD POLITÉCNICA SALESIANA**

Nosotros, Maria Fernanda Cabrera Arteaga con documento de identificación N° 1105795007 y Williams Andres Mendoza Vitonera con documento de identificación N° 0706283595, expresamos nuestra voluntad y por medio del presente documento cedemos a la Universidad Politécnica Salesiana la titularidad sobre los derechos patrimoniales en virtud de que somos autores del proyecto técnico: “Caracterización de curvas de carga en consumidores de la Empresa Eléctrica Regional Centro Sur C.A. mediante técnicas de agrupamiento de datos”, el cual ha sido desarrollado para optar por el título de: Ingeniero Eléctrico, en la Universidad Politécnica Salesiana, quedando la Universidad facultada para ejercer plenamente los derechos cedidos anteriormente.

En concordancia con lo manifestado, suscribimos este documento en el momento que hacemos la entrega del trabajo final en formato digital a la Biblioteca de la Universidad Politécnica Salesiana.

Cuenca, 04 de abril del 2023

Atentamente,



Maria Fernanda Cabrera Arteaga

1105795007



Williams Andres Mendoza Vitonera

0706283595

CERTIFICADO DE DIRECCIÓN DEL TRABAJO DE TITULACIÓN

Yo, Johnny Xavier Serrano Guerrero con documento de identificación N° 0104983382, docente de la Universidad Politécnica Salesiana, declaro que bajo mi tutoría fue desarrollado el trabajo de titulación: "CARACTERIZACIÓN DE CURVAS DE CARGA EN CONSUMIDORES DE LA EMPRESA ELÉCTRICA REGIONAL CENTRO SUR C.A. MEDIANTE TÉCNICAS DE AGRUPAMIENTO DE DATOS", realizado por Maria Fernanda Cabrera con documento de identificación N° 1105795007 y por Williams Andres Mendoza Vitonera con documento de identificación N° 0706283595, obteniendo como resultado final el trabajo de titulación bajo la opción Proyecto técnico que cumple con todos los requisitos determinados por la Universidad Politécnica Salesiana.

Cuenca, 4 de abril del 2023

Atentamente,



Firmado electrónicamente por:
**JOHNNY XAVIER
SERRANO GUERRERO**

Ing. Johnny Xavier Serrano Guerrero, PhD
0104983382

DEDICATORIA

Dedico mi trabajo a Dios por sus bendiciones y sabiduría, A mis padres Rosario y Manuel quienes han sido mi mayor inspiración. A ustedes, que siempre creyeron en mí y me enseñaron a nunca rendirme, les dedico este logro. A mis hermanos Diego, Yessenia, Manolo y Elías por sus palabras de aliento fueron mi motor en los momentos difíciles, y su amor incondicional me dio la fuerza para seguir adelante. También, quiero dedicar esta tesis a Juan Daniel, por su cariño, motivación, por su apoyo inagotable y por su paciencia infinita en este largo camino de investigación. Este logro es también para ustedes familia y amigos, porque cada uno ha dejado una huella en mi corazón.

Con cariño y gratitud,

Maria Fernanda Cabrera Arteaga

DEDICATORIA

En primer lugar, dedico este trabajo a Dios por darme la bendición de tener una familia que se apoya mutuamente y se esfuerza para salir a delante. A mi madre Maryuri Vitonera quién es un ejemplo de una mujer y madre que quiere y se esfuerza para darle lo mejor a sus seres queridos. A mis abuelos Anibal y Petronila, que han sido de gran apoyo para mi madre y para mí. A mi padre Williams Mendoza, que ha sido de gran ayuda para mi madre y para mí en los momentos donde se necesitaba de su ayuda. A mis tíos que siempre me han aconsejado para que no cometa tantos errores y me han apoyado en lo que pueden. Dedico este trabajo a todos los que han puesto su confianza en mí y me han apoyado a lo largo de esta etapa universitaria.

Williams Andres Mendoza Vitonera

AGRADECIMIENTOS

Primeramente, agradecer a Dios por permitirnos tener salud para poder culminar este proyecto.

A mis asesores académicos, Ing. Xavier Serrano por su guía experta, paciencia y apoyo constante a lo largo de este proceso. Sus valiosas sugerencias y comentarios fueron fundamentales para mejorar este trabajo.

Agradezco profundamente a mis padres y familiares por su amor incondicional, su constante estímulo y el apoyo financiero que me permitió enfocarme en mis estudios y llevar a cabo esta investigación.

También, quiero reconocer el apoyo de la a la Empresa Eléctrica Regional Centro Sur C.A. por brindar sus instalaciones, al Ing. Andrés Patiño y al Ing. Sergio Zambrano por su conocimiento y colaboración para realizar este proyecto de titulación.

Este logro no habría sido posible sin el apoyo incondicional de cada uno de ustedes. Gracias por ser parte de este importante capítulo de mi vida académica.

Maria Fernanda Cabrera Arteaga

AGRADECIMIENTOS

Para empezar agradecer a Dios por permitir culminar este trabajo con normalidad. Agradezco a nuestro asesor en este trabajo, el Ing. Xavier Serrano, por ayudarnos a poder desarrollar el trabajo de la manera más adecuada. Al Ing. Andres Patiño y al Ing. Sergio Zambrano por su asesoría y predisposición para ayudarnos a desarrollar el trabajo. A la Empresa Eléctrica Regional Centro Sur C.A. por permitirnos realizar el trabajo en su institución y darnos acceso al material necesario para su realización.

Agradezco al sacrificio y esfuerzo que ha hecho mi madre Maryuri Vitonera para permitir que pueda llevar a cabo toda esta etapa universitaria. A mi padre por darme su apoyo financiero cuando se necesitaba. A mis abuelos por ser un apoyo para mi madre y para mí cuando más lo necesitábamos. A mis tíos que me han ayudado mucho y apoyado en lo que pueden. Agradezco a todas las personas que conocí en esta etapa y en algún punto han sido personas importantes para mí, en especial algunos amigos que he llegado a considerar parte de mi familia, ya que he vivido muchas situaciones tanto buenas como malas y ellos siempre estuvieron para dar una mano amiga. También quiero reconocer a los ingenieros de la universidad que me han sabido ayudar cuando lo necesitaba.

Gracias por ser parte de esta etapa de mi vida.

Williams Andres Mendoza Vitonera

RESUMEN

La clasificación y caracterización de los datos de medición de campaña es primordial para poder identificar sus perfiles de carga. Sin embargo, analizar este tipo de datos sin un análisis del método de agrupamiento es complicado. Por lo que se propone una metodología para realizar el estudio de investigación y caracterización de la carga mediante la aplicación de los diferentes métodos de agrupamiento (clustering) aplicado a información de los perfiles de carga de clientes de media y baja tensión y transformadores de distribución obtenida de la Empresa Eléctrica Regional Centro Sur C.A. (Centrosur).

Hoy en día el requerimiento de procesar una gran cantidad de datos es cada vez mas frecuente. Investigaciones recientes proponen los métodos de agrupación de datos, los cuales consisten en una técnica que permite dividir una gran cantidad de datos en subgrupos similares entre si pero diferente de otros.

Se realizó una investigación de diversos métodos de agrupación y se determino cuál de los métodos tenía mejores resultados para este caso en particular. Esta propuesta presenta como resultado los perfiles de carga similares en comportamiento, pero con diferentes resultados en los datos de cantidad de observaciones, energía proyectada, factor de carga, potencia maxima del estrato, la potencia maxima, minima y promedio del perfil de carga obtenido los cuales se comparan con el trabajo previamente realizado por la Centrosur denominado “Investigación y Caracterización de la carga - Muestreo Aleatorio por Estratos” [1] (EERCSCA).

Se determinó que el método es eficiente, y sus resultados pueden ser aplicados para la planificación del sistema de distribución, ya sea en el dimensionamiento de estaciones de transformación, en estudios de pérdida o gestión de la carga.

ABSTRACT

The classification and characterization of campaign measurement data is essential to be able to identify their load profiles. However, analyzing this type of data without a clustering method analysis is difficult. Therefore, a methodology is proposed to carry out the research study and characterization of the load through the application of the different grouping methods (clustering) applied to information on the load profiles of medium and low voltage customers and distribution transformers obtained of the Empresa Eléctrica Regional Centro Sur C.A. (Centrosur).

Today the requirement to process a large amount of data is becoming more frequent. Recent research proposes data grouping methods, which consist of a technique that allows dividing a large amount of data into subgroups that are similar to each other but different from others.

An investigation of various grouping methods was carried out and it was determined which of the methods had the best results for this particular case. This proposal presents as a result load profiles similar in behavior, but with different results in the data of number of observations, projected energy, load factor, maximum power of the stratum, maximum, minimum and average power of the load profile obtained by the which are compared with the work previously carried out by the Centrosur called “Investigation and Characterization of the load - Random Sampling by Strata” [1] (EERCSCA). It was determined that the method is efficient, and its results can be applied for the planning of the distribution system, either in the sizing of transformation stations, in loss studies or load management.

Glosario

CEE Consumo de Energía Eléctrica.

Centrosur Empresa Eléctrica Regional Centro Sur C.A.

clustering Referencia para los métodos de agrupación.

clúster Grupo de datos.

DBSCAN Density-Based Spatial Clustering of Applications with Noise o
Agrupamiento Espacial Basado en Densidad de Aplicaciones con Ruido.

EERCSCA Estudio de investigación denominado “Investigación y
caracterización de la Carga - Muestreo Aleatorio por Estratos” [1]
realizado por la Centrosur.

GMM Método de Mezcla Gaussiana.

kVA_r Kilovoltamperio reactivo.

kW Kilovatio.

RD1 Estrato residencial de 0 a 60 kWh.

Índice General

1. CAPÍTULO 1: ESTUDIO DEL ESTADO DEL ARTE	4
1.1. FUNDAMENTOS TEÓRICOS	5
1.1.1. Energía eléctrica	5
1.1.2. Factor de carga	6
1.1.3. Factor de potencia	6
1.1.4. Tarifa Eléctrica	6
1.1.5. Estratos	8
1.2. Herramientas computacionales	8
1.2.1. Lenguaje de programación Python	8
1.2.1.1. Librería de Python: SCIKIT-LEARN	9
1.2.2. Big Data	9
1.3. Machine Learning (Aprendizaje automático)	9
1.3.1. Aprendizaje supervisado	10
1.3.1.1. Clasificación	10
1.3.1.2. Regresión	10
1.3.2. Aprendizaje no supervisado	10
1.3.2.1. Análisis clúster	11
1.3.3. Aprendizaje por refuerzo	12
1.3.4. Aprendizaje semi-supervisado	12
1.4. Gestor de base de datos MySQL	13
2. CAPÍTULO 2: MÉTODOS DE AGRUPACIÓN	14
2.1. Algoritmo K-means	14
2.2. Mezcla Gaussiana	15
2.3. Algoritmo DBSCAN	15
2.4. Método del diagrama del codo	17
3. CAPÍTULO 3 : METODOLOGÍA PROPUESTA	18
3.1. Etapa 1: Adquisición de datos y proceso	19
3.1.1. Adquisición de datos de clientes de Media Tensión y Baja Tensión	19
3.1.2. Validación de datos	20
3.1.3. Cálculos necesarios	20
3.1.4. Filtrado de información	20
3.2. Etapa 2: Exportación al gestor de datos y evaluación de los métodos de agrupación	20

3.2.1.	Exportación de datos a MySQL	20
3.2.2.	Evaluación de técnicas de agrupación	21
3.2.2.1.	Método K-means	21
3.2.2.2.	Método DBSCAN	21
3.2.2.3.	Método Mezcla Gaussiana	22
3.2.3.	Selección de los métodos a aplicar	22
3.3.	Etapa 3: Clasificación, caracterización y análisis de los resultados	23
3.3.1.	Clasificación de los clientes y caracterización	23
3.3.2.	Gráficos y análisis	23
3.3.3.	Análisis de los resultados en el dimensionamiento de estaciones de transformación	23
4.	CAPÍTULO 4: RESULTADOS Y CONCLUSIONES	24
4.1.	Resultados	24
4.1.1.	Métodos de Prueba	24
4.1.2.	Comparación de los resultados obtenidos por los métodos aplicados y los de la Centrosur	26
4.1.3.	Análisis de los resultados de Mezcla Gaussiana	31
4.2.	Valoración del método en el dimensionamiento de estaciones de transformación	33
4.3.	Discusión	35
4.4.	Conclusiones	36
4.5.	Recomendaciones	37

Índice de figuras

1.	Clasificación.	10
2.	Aprendizaje.	11
3.	Tipos de clustering.	11
4.	Agrupación Jerárquica.	12
5.	Agrupación por particiones.	12
6.	Agrupamiento DBSCAN [2].	16
7.	Método del codo [3].	17
8.	Diagrama de bloques	19
9.	Ejemplo de curva de selección de cantidad de clústers.	21
10.	Ejemplo de curva de selección de épsilon.	22
11.	Perfiles de carga de los diversos clústers generados por los métodos en el estrato RD1.	24
12.	Perfiles de carga del estrato RD1 obtenidos de la aplicación de los 3 métodos.	26
13.	Perfiles de carga de los estratos residenciales.	29
14.	Perfiles de carga de los distintos grupos de consumo.	30
15.	Perfiles de carga residencial de la región Sierra y Oriente	32
16.	Perfiles de carga de los grupos de consumo en diferentes tipos de día	32
17.	Perfiles de carga de los diferentes tipos de clientes del transformador 31240.	34
18.	Perfil de carga resultante de los clientes y Perfil de carga del transformador 31240.	34

Índice de tablas

1.1. Pliego tarifario de la Empresa Eléctrica Regional Centro Sur [4].	7
1.2. Estratos por grupo de consumo [1]	8
4.1. Tabla de comparación de los métodos aplicados	25
4.2. Resumen de resultados	27
4.3. Comparación de Estratos Residenciales	28
4.4. Datos de los estratos de la región Sierra en tipo de día laborable	31
4.5. Tabla de datos del transformador 31240 y sus tipos de clientes	33

INTRODUCCIÓN

El consumo de la energía eléctrica (CEE) ha aumentado continuamente, convirtiéndose en un recurso esencial y primordial para el desarrollo de actividades industriales, comerciales y económicas. El avance tecnológico en las últimas décadas ha provocado el aumento de dispositivos conectados a la red, incrementando su datos de medición [5].

Las empresas eléctricas cuentan actualmente con una amplia base de datos de mediciones de perfiles de carga obtenidos en campañas de medición. El análisis de estos datos permite la caracterización del consumidor; para describir el consumo de energía del mismo, sin embargo, esta tarea se vuelve compleja cuando no se dispone de las herramientas adecuadas, por lo que se analiza los métodos de agrupación [6].

El agrupamiento de datos es un conjunto de metodologías que se utiliza para la partición automática de una gran cantidad de datos en grupos definidos, utilizando medidas de asociación, en los que cada grupo consta de objetos que son similares entre sí y diferentes a los objetos de otros grupos [7].

Este documento se estructura de la siguiente manera, en el primer capítulo, se realiza una introducción a los conceptos básicos para describir técnicamente, o estimar, el comportamiento y características de la demanda eléctrica, en forma colectiva, según el contexto de las mediciones de calidad. En el segundo capítulo se realiza un estudio de los métodos de clustering para realizar la clasificación y caracterización de clientes. En el tercer capítulo se estudia y se muestra la metodología detallada para la elaboración del proyecto, incluyendo el manejo y procesamiento de datos, diseño y desarrollo del programa y la metodología para el análisis comparativo de resultados. En el cuarto capítulo se resalta las comparaciones y resultado de los métodos de agrupación.

JUSTIFICACIÓN

Ante la aparición de nuevas técnicas de agrupamiento de datos y software para ejecutar los diferentes métodos de clustering o para alojar una gran cantidad de datos, surge la necesidad de actualizar el estudio de investigación de la carga de clientes. En la Universidad Politécnica Salesiana ya se realizó previamente un trabajo de caracterización con clustering o agrupamiento de datos [6], tomando como partida ese trabajo se busca realizar la clasificación y caracterización de los distintos clientes de la Centrosur., mediante técnicas de clustering, con el objetivo de verificar su funcionalidad para la representación de los diferentes estratos a los que pertenecen los clientes, proporcionando de esta manera un método que sea de utilidad para futuros análisis.

La caracterización de los perfiles de carga en consumidores mediante técnicas de agrupamiento es un desafío complejo que requiere una cuidadosa preparación y análisis de los datos, elección adecuada de técnicas y una interpretación adecuada de los resultados para obtener información valiosa y práctica para la CentroSur.

OBJETIVOS

Objetivo General

- Caracterizar los perfiles de carga de los consumidores de la Empresa Eléctrica Regional Centro Sur C.A. utilizando técnicas de agrupamiento.

Objetivos Específicos

- Realizar una investigación bibliográfica de métodos de agrupamiento, Big Data y Machine Learning.
- Procesar y validar datos obtenidos en campañas de medición.
- Evaluar técnicas de agrupamiento para la caracterización de la demanda de los usuarios.
- Caracterizar y clasificar los perfiles de carga de los usuarios analizados.
- Valorar la aplicación de técnicas de agrupamiento de perfiles de carga en el dimensionamiento de estaciones de transformación.

1. CAPÍTULO 1: ESTUDIO DEL ESTADO DEL ARTE

En este apartado, se presenta una revisión literaria relacionada a la aplicación de la caracterización de las curvas de los usuarios mediante técnicas de agrupación. La información recaudada proviene de bases de datos como: Scopus, IEEE Explore, Science Direct, entre otras. En general los artículos se componen de palabras relacionadas con la caracterización de curvas, técnicas de agrupación y otros criterios.

La mayor parte de los algoritmos que existen tienden a investigar información del conjunto a nivel de objeto, pero carecen de una capacidad de explorar a niveles más altos es decir con una gran cantidad de datos. Para poder abordar uno de los problemas esta investigación presenta un enfoque de agrupamiento de conjuntos basado en una propagación de semejanzas a través de caminos aleatorios [8]. La técnica de aproximación de agrupamiento de datos nos muestra resultados con calidad, para poder obtener el equilibrio entre la eficiencia del agrupamiento junto con la estimación de precisión [9].

Una forma para poder describir el consumo de energía de los clientes es mediante el procesamiento de datos, de esta forma se pueden definir las características del usuario mediante un método de clustering [6], [10].

La clasificación de los usuarios en los distintos estratos se realiza con base en las características de estos, encasillándolos en diversos grupos con propiedades exclusivas de cada uno. Para los diferentes agrupamientos se toman como referencia características o dimensiones de los datos. Algunas de las técnicas de agrupamiento que se pueden aplicar son K-means, Mezcla Gaussiana y DBSCAN [11], [12].

La técnica K-means consta de una cantidad k de clústeres para inicializarse; la asignación de objetos (elementos a examinar) hacia el centroide (datos centrales en común entre un grupo de elementos) más cercano, una actualización constante de los centroides de cada clúster, índices de validez externos que realiza una comparación de los miembros del clúster, índices de validez internos que evalúan la confiabilidad de la estructura del clúster.

Esta técnica posee deficiencias, por lo que han surgido variantes con modificaciones o ajustes para el mismo, como un método no supervisado, para liberarlo de parámetros de selección e inicialización o una variante de

análisis de solo puntos extremos del clúster [13], [14], [15].

El trabajo de Wang, Wang, Jia, Chen propone una investigación de agrupamiento de carga basada en la descomposición de valores singulares y el algoritmo de agrupamiento de K-means, donde se extraen ocho índices característicos de carga, y el valor singular se usa para descomponer las características de carga del lado del usuario [16].

Tambunan, Barus, Hartono, Alam presentan en su trabajo un método de análisis para obtener información mediante la agrupación, donde estudian el comportamiento de la carga y proporcionan información de cómo opera la planta de energía. Además, analiza los datos mediante K-means, los agrupa de manera óptima en 3 grupos los cuales se evalúan mediante puntajes de siluetas interpretado por un nivel de carga alto, intermedio y bajo. El coeficiente de silueta (SC) evalúa el grupo en función de la medida de la distancia promedio entre un punto de datos y otros puntos de datos de los mismos grupos [17].

Marrero y otros utilizan un algoritmo que permite realizar una clasificación de los perfiles de clientes que se basa en el consumo eléctrico, el cual es el método jerárquico K means que particiona los clientes en un número específico de grupos, basándose en las características que poseen [18].

El trabajo de Ullah y otros, investiga el análisis de consumo de energía basado en agrupaciones para poder clasificar el uso de electricidad de los consumidores en diferentes niveles, entrena un codificador automático profundo que transfiere datos de consumo de energía de baja dimensión a una representación de alto nivel y luego utiliza algoritmos para autoorganizar un mapa adaptativo de agrupamiento [19].

1.1. FUNDAMENTOS TEÓRICOS

La sección de fundamentos teóricos reúne los conceptos aplicados en el estudio de la caracterización de las curvas.

1.1.1. Energía eléctrica

Se refiere a la cantidad de energía utilizada durante un período de tiempo específico, que se calcula entre la demanda de energía eléctrica y la duración del intervalo[20]. Se evalúa que en la ciudad de Cuenca cada persona consume alrededor de 1.600 kWh eléctricos al año. Este valor se encuentra por encima

del promedio nacional, que es alrededor de 1.300kWh por persona en el año [21].

1.1.2. Factor de carga

Es la relación entre la carga real y la carga máxima que podría soportar un periodo determinado, indica el grado de los niveles de la carga máxima estimados durante un periodo de estudio que suele ser diarios, mensuales o anuales, si el factor de carga es cercano a 1, el consumo esta correctamente distribuido [22], [23].

1.1.3. Factor de potencia

La vinculación entre la potencia activa y la potencia aparente se refiere comúnmete a la cantidad de energía eléctrica que se a convertido en trabajo [24].

1.1.4. Tarifa Eléctrica

Al contextualizar la situación actual de las tarifas eléctricas en Ecuador, se hace alusión al pliego tarifario del país emitido por el ARCERNNR en el año 2022 [25].

Tabla 1.1: Pliego tarifario de la Empresa Eléctrica Regional Centro Sur [4].

Categoría	Tarifa	Categoría	Tarifa
-General en baja tensión sin demanda.	-RD (residencial) -TE (tercera edad) -AS (asistencia social) -BP (beneficio público) -CO (comercial) -ES (escenario deportivo) -IA (industrial artesanal) -OF (entidad oficiales) -BJ (bombeo de agua) -AB (Asistencia social en baja tensión con demanda) -BB (Beneficio público en baja tensión con demanda) -CB (Comercial baja tensión con demanda) -B (Industrial baja tensión con demanda) -MB (Entidad Municipal Baja Tensión con demanda)	-General con demanda horaria	-AH (Asistencia social demanda horaria) -BH (Beneficio Público demanda horaria) -CH (Comercial demanda horaria) -EH (Escenario deportivo demanda horaria) -IH (Industrial demanda horaria) -JH (Industrial con medición horaria con incentivos MT) -MH (Entidad Municipal demanda horaria) -DH (Entidades Oficiales demanda horaria) -UH (culto religioso demanda horaria) -W3 (Bombeo de Agua con demanda horaria)
-General en baja tensión con registro de demanda horaria	-A3 (Asistencia social BT con demanda horaria) -B3 (Beneficio público BT con demanda horaria) -C3 (Comercial BT con demanda horaria) -E3 (Escenario Deportivo BT con demanda horaria) -BT (Industrial Artesanal con demanda Horaria)	-General con demanda	-BA (Bombeo de Agua) -BD (Beneficio Público) -CD (Comercial) -CK (Culto Religioso) -ED (Escenario Deportivo) -ID (Industrial) -MD (Entidad Municipal) -DD (Entidades Oficiales) -DC (Servicio eventual comercial) -DI (Servicio eventual industrial) -WD (Bombeo de Agua con demanda)
-Alumbrado público	-AP (Alumbrado Público)	-Consumidores de servicios en alta tensión	-KH (Industriales con medición horaria con incentivo AT)

1.1.5. Estratos

Para establecer el estrato en general, se considera el factor del consumo de energía y el tipo de uso que se da a la energía que puede ser: residencial, comercial, industrial y otros. Según la aplicación de la metodología de Dalenius-Hodges se calcula la frecuencia, con la amplitud baja determinando el grupo de consumo [1]. En la Tabla 1.2 se puede observar los estratos por grupo de consumo.

Tabla 1.2: Estratos por grupo de consumo [1]

Grupo de consumo	Estrato kWh>	Estrato kWh>=	Estrato
Residencial	0	60	1
Residencial	60	110	2
Residencial	110	180	3
Residencial	180	310	4
Residencial	310	Superior	5
Comercial	0	290	1
Comercial	290	1235	2
Comercial	1235	Superior	3
Industrial	0	410	1
Industrial	410	2520	2
Industrial	2520	Superior	3
Otros	0	405	1
Otros	405	1820	2
Otros	1820	Superior	3

1.2. Herramientas computacionales

1.2.1. Lenguaje de programación Python

Python es un lenguaje de programación que ofrece una variedad de estructuras de datos eficientes y de alto nivel con un enfoque simple para la programación orientada a objetos [26]. Python se enfoca en un lenguaje ideal para secuencias de comandos y desarrollo rápido de aplicaciones en muchas áreas y en la mayoría de las plataformas [27].

Además, Python está considerado como uno de los lenguajes de programación más versátiles para procesar grandes cantidades de datos, por las librerías que permiten procesar información en poco tiempo. [5].

1.2.1.1. Librería de Python: SCIKIT-LEARN La biblioteca SCIKIT LEARN inició como un proyecto de Google Summer en el 2007 por David Cournapeau e inmediatamente se unió la comunidad científica de Python. En el 2009 se lanzó una primera versión de Scikit-Learn, hasta llegar a obtener 40 versiones [28].

Es una librería eficiente de aprendizaje automático orientada a la predicción y análisis de datos. Cuenta la funcionalidad principal para el análisis, es la capacidad de diferenciar la clasificación y regresión [29].

1.2.2. Big Data

Con el continuo crecimiento de datos que cada día genera las empresas han avanzado con el uso de BIG DATA, el cual, surge como una nueva opción [30]. En la actualidad es posible almacenar grandes cantidades de datos que con el tiempo se incrementan rápidamente. Por tanto, es necesario que el analista se mantenga actualizado en nuevas formas de procesamiento [31].

Se caracteriza por la rapidez en el procesamiento de datos, la diversidad de información y la cantidad de datos manejados, considerando el volumen de información [5].

1.3. Machine Learning (Aprendizaje automático)

Se considera una rama de la inteligencia artificial que es responsable de crear algoritmos que contienen la capacidad de aprender a realizar tareas sin una programación específica. En sí, no se programa para que responda de una determinada forma según los datos de entrada, sino extra patrones de comportamiento a partir de las entradas y con base en lo anterior realizar evaluación de nuevas entradas [32].

Con el continuo avance de la tecnología y los algoritmos buscan mejorar la eficiencia de sus procesos, el aprendizaje automático se clasifica en aprendizaje supervisado, no supervisado y por refuerzo como se observa en la Figura 1 [5].

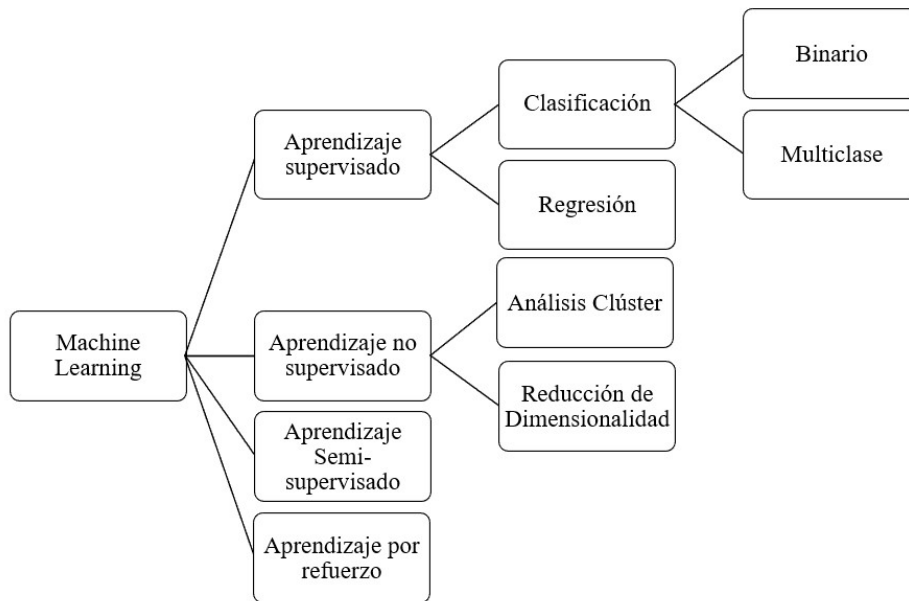


Figura 1: Clasificación.

1.3.1. Aprendizaje supervisado

Es un campo de estudio que da a los ordenadores la habilidad de aprender algo [33]. El algoritmo es instruido sobre cómo desempeñar su función mediante un conjunto de datos previamente clasificados bajo una cierta perspectiva o concepto, con el propósito de identificar patrones que puedan aplicarse en un análisis y generar una salida que ya se encuentra conocida [34].

1.3.1.1. Clasificación La clasificación tiene una perspectiva supervisada en que la máquina estudia los datos que se le otorgan y luego usa del mismo para clasificar una nueva observación.

1.3.1.2. Regresión La regresión se utiliza para asignar categorías sin la necesidad de etiquetar los datos.

1.3.2. Aprendizaje no supervisado

Se refiere a un modelo predictivo entrenado de manera similar al aprendizaje supervisado, pero con la distinción de que su comprensión se da en datos no clasificados o etiquetados, permitiendo el análisis de nuevos patrones en grupos de datos tal como se observa en la Figura 2[35].

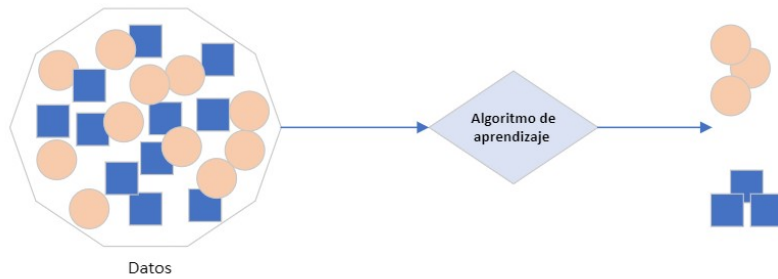


Figura 2: Aprendizaje.

1.3.2.1. Análisis clúster Es una técnica de aprendizaje no supervisado conocido como análisis conglomerados, su principal objetivo es agrupar conjuntos en subgrupos que son diferentes de los otros, y que sus miembros sean similares entre sí [36].

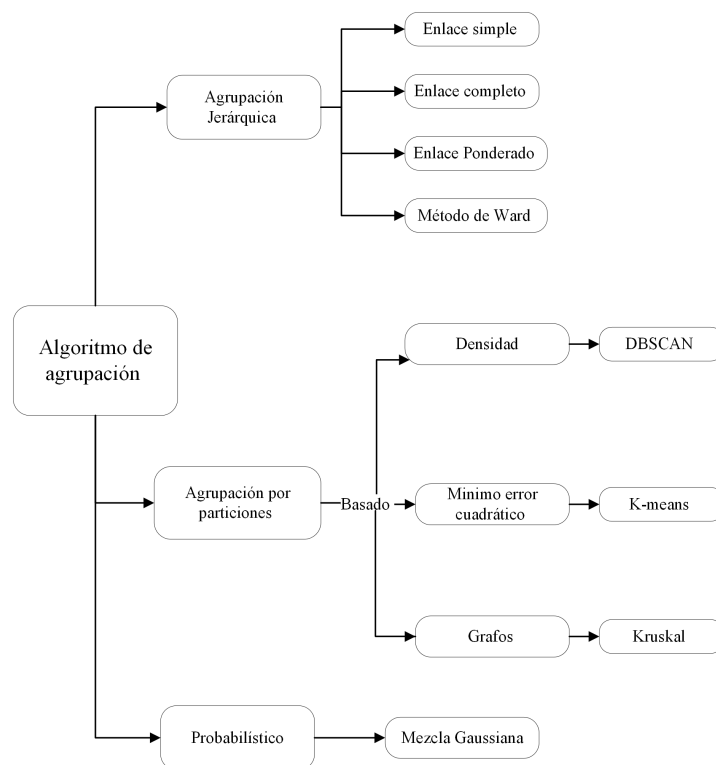


Figura 3: Tipos de clustering.

Agrupación Jerárquica Esta agrupación se caracteriza por ser más apropiada en situaciones donde se manejan muestras de tamaño reducido, dado que los grupos se forman mediante la agrupación secuencial de

individuos [37].

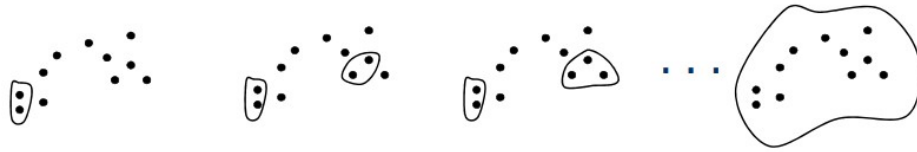


Figura 4: Agrupación Jerárquica.

Agrupación por particiones El objetivo principal de la Agrupación particional es generar una partición de los objetos, agrupándolos en conjuntos o clusters de manera que cada objeto pertenezca a uno de los K clusters disponibles [38].



Figura 5: Agrupación por particiones.

Agrupación probabilístico Esta técnica emplea la distribución de probabilidad como base para la formación de clusters. El algoritmo utilizado en esta técnica se conoce como el Modelo de agrupamiento Gaussiano [39].

1.3.3. Aprendizaje por refuerzo

Se trata de un tipo de aprendizaje automático en el cual no se realiza ningún entrenamiento previo con datos clasificados o no clasificados. En este enfoque, el sistema aprende en un entorno donde no se dispone de información sobre la posible salida, sino que adquiere conocimiento a través de sus acciones y los resultados obtenidos en el proceso [35].

1.3.4. Aprendizaje semi-supervisado

Este tipo de aprendizaje utiliza del aprendizaje supervisado y no supervisado, se utiliza cuando existe gran cantidad de datos de entrada y datos no etiquetados. Los desafíos que surgen en situaciones del mundo real se alinean con esta área debido a que asignar etiquetas a una gran cantidad de datos resulta en un proceso que consume mucho tiempo, mientras que

recopilar y almacenar datos no etiquetados resulta más sencillo [35]. El modelo más antiguo es una distribución de mezcla confiable, como ejemplo, la mezcla gaussiana [35]. Los modelos de mezcla más comunes son:

- Mezcla de distribuciones gaussianas.
- Mezcla de distribuciones multinomiales.
- Modelo oculto de Márkov.

1.4. Gestor de base de datos MySQL

MySQL se basa en un modelo cliente-servidor, maneja todas las instrucciones de la base. Permite seleccionar y manejar datos de una gran cantidad de tipos de tablas, con seguridad y no desperdicia los recursos del servidor [40].

2. CAPÍTULO 2: MÉTODOS DE AGRUPACIÓN

La definición de “agrupación”, es un método de Análisis Exploratorio de Datos que analiza problemas de clasificación [41], varía según la aplicación que se le asigne. Sin embargo, el agrupamiento como un método puede ser interpretado de diversas maneras de acuerdo con el caso que se aplique, en nuestro caso el proceso es agrupar los datos en grupos homogéneos, con la finalidad de descubrir la estructura de los datos e identificar los individuos que están dentro de cada grupo [42].

2.1. Algoritmo K-means

El algoritmo K-means es uno de los métodos de agrupamiento más conocido, que se utiliza en la minería de datos, en una forma general para dividir un conjunto de datos en k grupos y cada clúster agrupado con un centroide. El principal objetivo es disminuir la suma de los errores cuadráticos utilizando la distancia euclidiana medida [43], [44]. La ventaja principal del método es la rapidez en presentar la robustez a dimensiones no tan elevadas. La forma más clara de determinar la proximidad de una observación al centroide se logra mediante el cálculo de la distancia euclidiana [45]. Por otro lado, la principal desventaja es que depende de la cantidad de k clústeres que defina el usuario. Del mismo modo, otro problema es la influencia directa es que k-means es muy sensible al ruido [46], [47].

El proceso para agrupar se divide en 4 etapas:

1. Inicialización: se establece la cantidad de grupos que determinan los centroides iniciales.
2. Clasificación: para cada grupo se calcula la distancia a cada centroide y se determina el más cercano.
3. Determinación de centroides: para cada grupo inicializado en el paso anterior se vuelve a calcular el centroide.
4. Condición de convergencia: La convergencia ocurre en situaciones donde no hay un intercambio de objetos entre los grupos y cuando la distancia entre los centroides de dos iteraciones sucesivas es menor que un valor umbral predefinido [48].

2.2. Mezcla Gaussiana

El modelo de mezcla gaussiana es un algoritmo de aprendizaje no supervisado que se emplea en una función de densidad de probabilidad, representada como una suma ponderada de densidades, utilizada como modelo paramétrico para la distribución de probabilidad de las medidas [49]. Puede ser comprendido como una extensión de K-means en la cual, en lugar de asignar cada observación a un único clúster, se obtiene una distribución de probabilidad que indica la pertenencia a cada uno de los clústeres [50].

El objetivo es encontrar una aproximación o estimación a partir de los datos encontrando un acomodamiento de los datos que contienen los componentes [51].

La principal ventaja es aplicar una distribución normalizada que asegura que las observaciones asociadas a la agrupación no se repitan o compartan características de otros centroides [52].

1. Inicialización: se selecciona un número K de distribuciones gaussianas que se utilizan para modelar datos y se inician los parámetros de las distribuciones como la media y varianza.
2. Expectación: para cada punto se calcula la probabilidad a la que pertenece cada una de las K distribuciones gaussianas.
3. Maximización: se actualizan los parámetros de cada una de las K distribuciones para maximizar la probabilidad de los datos observados.
4. Evaluación: se estima la convergencia del algoritmo comparando la probabilidad logarítmica de los datos observados antes y después de los datos actualizados.

2.3. Algoritmo DBSCAN

DBSCAN es un algoritmo de agrupación espacial basado en la densidad. Es una técnica de agrupación de regiones con alta densidad en los mismos clústeres, capaz de manejar con eficacia el ruido y los valores atípicos. A diferencia de k-means este algoritmo puede identificar clústeres de manera análoga, de tal forma que el usuario puede identificar y seleccionar los clústeres [53], [2], [54]. La idea es encontrar áreas densas y expandirlas recursivamente para encontrar grupos [55].

Para la representación central de DBSCAN para cada objeto de un clúster se necesita de:

1. La vecindad de un radio determinado **Eps (epsilon)** debe contener al menos un número mínimo (**min_samples**) de objetos.
2. Recopilación de los datos de densidad alcanzable directamente de estos datos centrales, lo que puede implicar un nuevo grupo de densidad alcanzable.
3. El proceso termina cuando no se puede agregar ningún objeto nuevo a ningún grupo. Los puntos que no están asignados a los grupos formados son denominados ruido, mientras que aquellos que no son ni ruido ni puntos centrales reciben el nombre de puntos borde [2], [56].

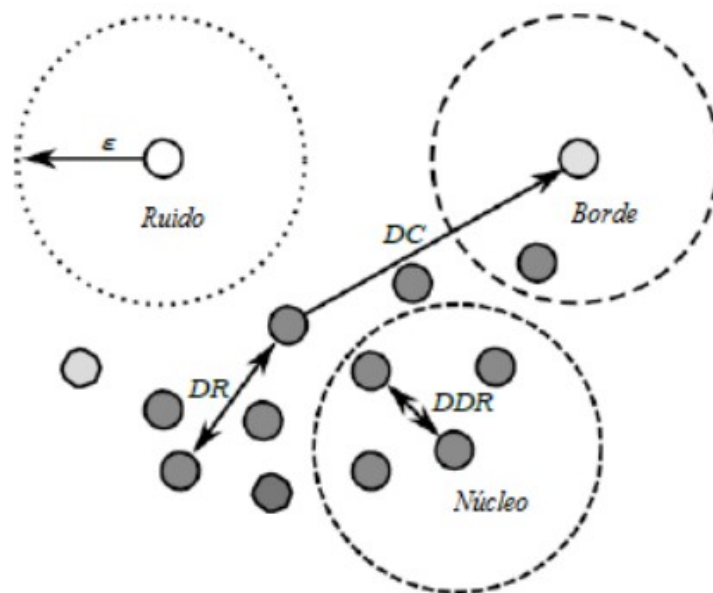


Figura 6: Agrupamiento DBSCAN [2].

El algoritmo permite realizar clústeres sin importar la estructura de los datos, a diferencia del algoritmo k-means no es necesario los parámetros iniciales del número de clúster para realizar sus iteraciones [52]. Existen dos principales desventajas, la primera es la elección de los parámetros, los mismo que deben ser ingresadas por el usuario, dado que estos afectan de manera directa los resultados de los clústeres. La segunda desventaja se relaciona con el tiempo de ejecución computacional, el cual aumenta a medida que se incrementa la dimensión de la base de datos ingresada [52].

2.4. Método del diagrama del codo

El método del codo permite la medición de la media de los grupos de tal manera que determina el número de n grupos eficientes para poder realizar la técnica de agrupamiento [52].

Este método utiliza valores de inercia que se obtienen al aplicar el método de agrupamiento, siendo la inercia la suma de las distancias al cuadrado de cada clúster al centroide. La ecuación que representa la inercia de los objetos esta dada por la ecuación [3].

$$Inercia = \sum_{i=0}^k \|x_i - \mu\|^2 \quad (2.1)$$

Esta representación puede ser graficada, mostrando el número de clústeres que deben ser utilizados como dato inicial [52]. La Figura 5 muestra en su eje horizontal el número de clústers y en el eje vertical la inercia de las distancias de los agrupamientos de cada clúster. El punto en dónde se produce un cambio brusco es el indicador del número de clústeres óptimo [3].

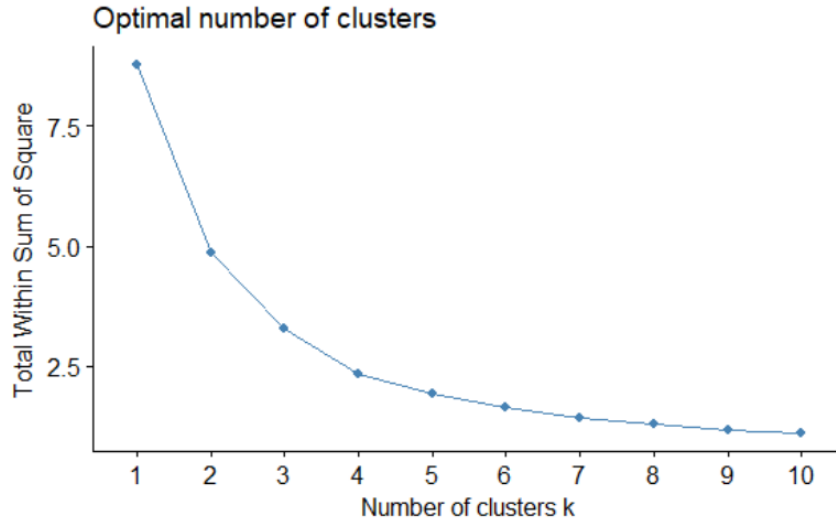


Figura 7: Método del codo [3].

3. CAPÍTULO 3 : METODOLOGÍA PROPUESTA

Para el desarrollo del trabajo se plantea una metodología aplicada, a través de un proceso sistemático de indagación, análisis e interpretación de información, que consta de tres etapas principales para lograr resultados confiables.

En la Figura 8 se observa el resumen de la metodología. La primera etapa aborda la adquisición y tratamiento de los datos, en la segunda etapa se exportan los datos al gestor y se analizan los métodos de agrupación. Por último en la tercera etapa se realiza la clasificación, caracterización y el análisis de los resultados.

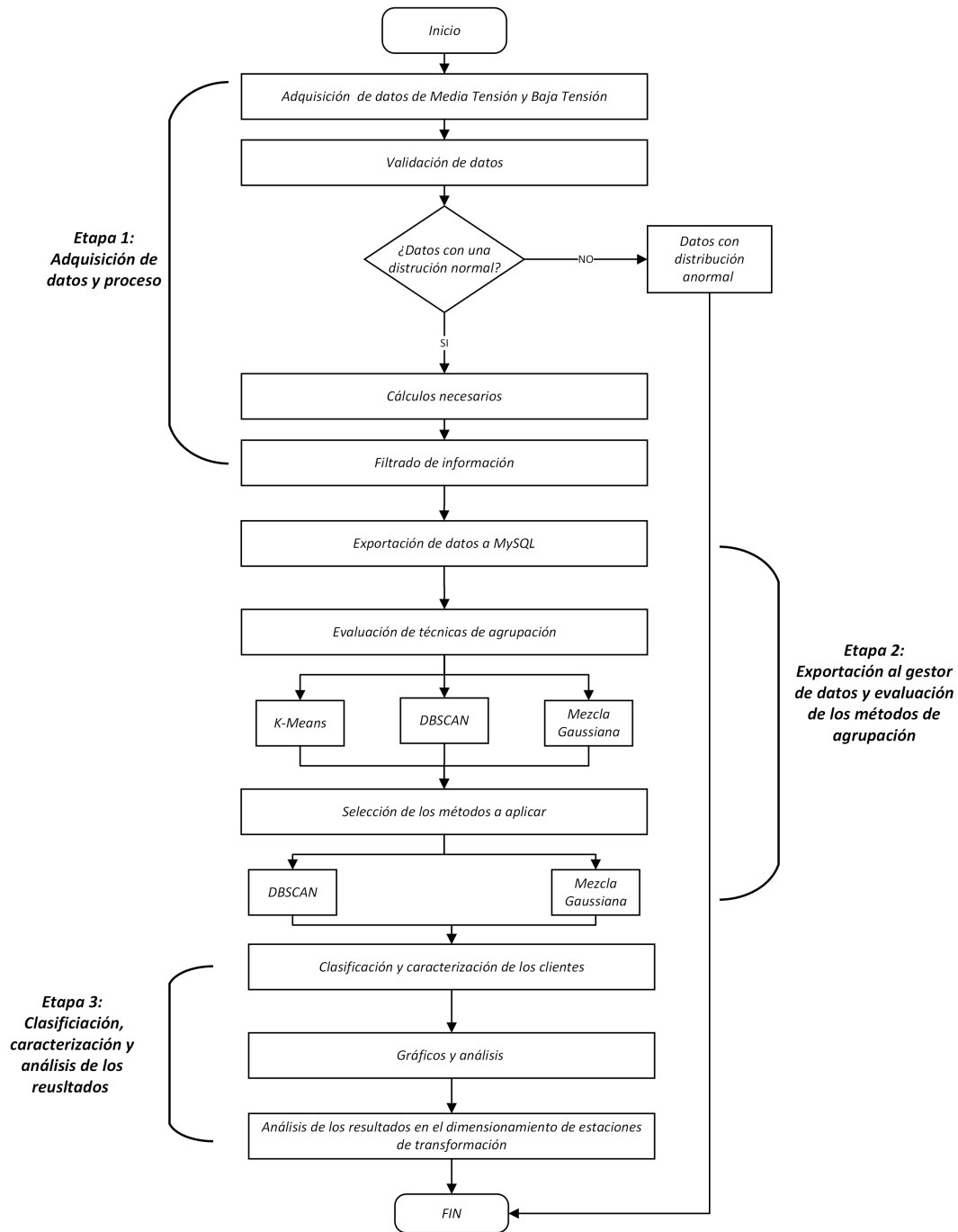


Figura 8: Diagrama de bloques

3.1. Etapa 1: Adquisición de datos y proceso

3.1.1. Adquisición de datos de clientes de Media Tensión y Baja Tensión

Los datos adquiridos para el análisis y caracterización de perfiles corresponden a las mediciones de calidad de baja tensión y media tensión de la Empresa Eléctrica Regional Centro Sur C.A, datos de medición de clientes

residenciales, comerciales e industriales durante 7 días de la semana cada 10 minutos. Se cuenta con información disponible para el período comprendido entre marzo del 2015 y mayo del 2023.

3.1.2. Validación de datos

Si los documentos receiptados por la empresa contienen mediciones negativas en la potencia activa el documento se descarta. Esta fase permite trabajar con los archivos de los clientes completos y evita problemas a futuro.

3.1.3. Cálculos necesarios

Se obtiene la potencia activa **P[kW]** y reactiva **Q[kVAr]** mediante la suma de **P promedio [W]** y **Q promedio [VAr]** de cada fase del cliente.

$$\mathbf{P[kW]} = P_{prom.A} + P_{prom.B} + P_{prom.C} \quad (3.1)$$

$$\mathbf{Q[kVAr]} = Q_{prom.A} + Q_{prom.B} + Q_{prom.C} \quad (3.2)$$

Una vez realizada la suma de las potencias, se calcula el promedio dividiendo las sumas por el número total de clientes. Al igual que el máximo y los valores por unidad.

3.1.4. Filtrado de información

Se filtran los datos de los usuarios según lo requerido en el programa, en el caso de estudio se filtra por carga, región, día laborable, feriados (se toman los días festivos como tipo de día domingo) y grupo de consumo.

3.2. Etapa 2: Exportación al gestor de datos y evaluación de los métodos de agrupación

3.2.1. Exportación de datos a MySQL

Para establecer la comunicación entre Python y MySQL, se procede a realizar la instalación de los paquetes necesarios mediante el comando “pip install mysql-connector”. A continuación, se importa la librería requerida utilizando “import mysql.connector”. Para lograr una conexión correcta, es necesario proporcionar los detalles pertinentes, tales como el nombre del host, el usuario, la contraseña y el nombre de la base de datos.

3.2.2. Evaluación de técnicas de agrupación

Con el estudio previo de los métodos de agrupación K-means, Mezcla Gaussiana y DBSCAN se evaluó cuál de ellos proporciona la agrupación y caracterización de los datos de los clientes de manera más precisa.

Los métodos son aplicados en el estrato RD1, con el fin de obtener los distintos tipos de perfiles que hay en él estrato y la distribución de las observaciones (datos de consumo captados en cada cliente) en cada clúster según cada método, con esta información ya se puede realizar una comparación y evaluar cual posee un mejor rendimiento.

3.2.2.1. Método K-means El parámetro importante es el número de clústers, se determina mediante el criterio de “diagrama del codo”, en este caso la Figura 9 representa la distancia media entre las observaciones a su centroide vs el número de clústers.

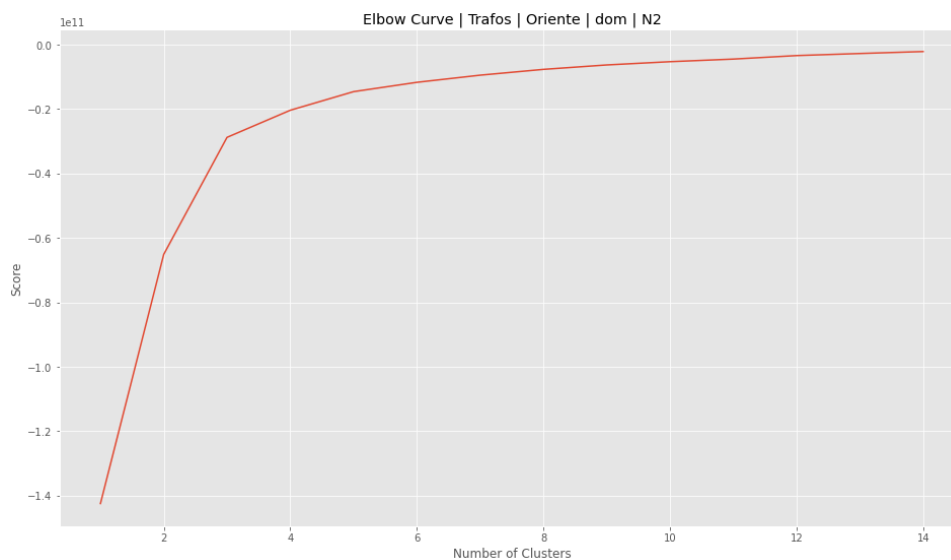


Figura 9: Ejemplo de curva de selección de cantidad de clústers.

3.2.2.2. Método DBSCAN Uno de los parámetros importantes de este método es ϵ (distancia máxima a la cual se debe encontrar otra observación) para determinar este valor nuevamente se recurre al criterio de “diagrama del codo”, pero esta vez se determina mediante la distancia entre observaciones vs la cantidad de observaciones como se muestra en la Figura 10. Otro parámetro importante es la cantidad mínima de observaciones para formar un clúster, el cual se establece en el valor de “3” para todos los casos,

ya que no existe un cambio significativo al momento de realizar pruebas con diferentes valores alrededor del mismo.

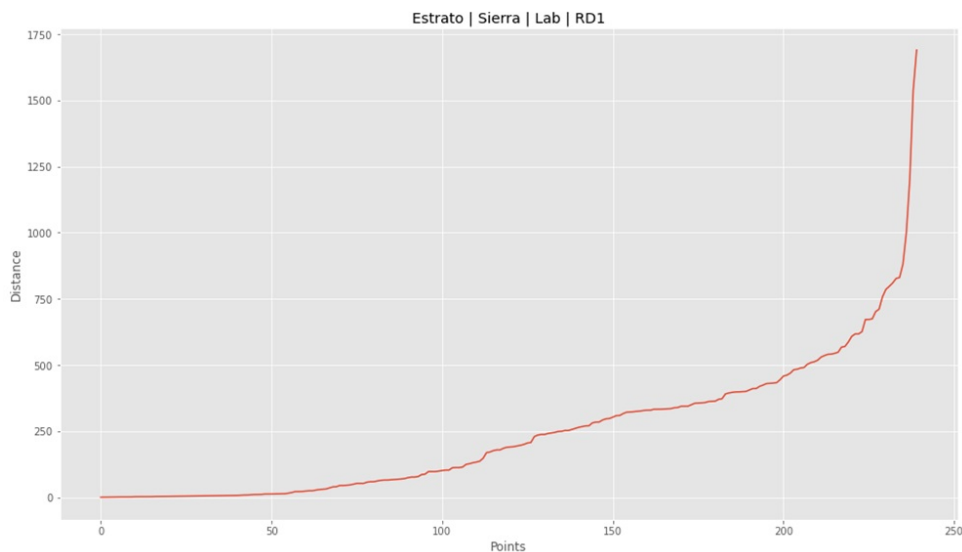


Figura 10: Ejemplo de curva de selección de ϵ .

3.2.2.3. Método Mezcla Gaussiana Este guarda muchas similitudes con el algoritmo K-means, ya que es una variante del mismo. Una de sus diferencias es la existencia de un filtro previo basado en la desviación estándar (indica cuánto se alejan los datos de la media). Otra diferencia es su agrupamiento, el cuál se basa en una matriz de probabilidades.

3.2.3. Selección de los métodos a aplicar

Mediante un análisis de los métodos K-means, DBSCAN y Mezcla Gaussiana, se seleccionó cuál de estos es el más adecuado para realizar la clasificación y caracterización de clientes de forma satisfactoria. Considerando las ventajas y desventajas que ofrecen los distintos métodos en relación al contexto requerido en el trabajo, donde se determina que Mezcla Gaussiana y DBSCAN ofrecen buenos resultados:

- Mezcla Gaussiana: es el método que se adapta a diferentes formas de distribución y proporciona información sobre la probabilidad de pertenecer a cada grupo.
- DBSCAN : es útil con datos con distribuciones de densidad no uniformes y se desea identificar grupos de diferentes formas y tamaño. Así mismo, es capaz de detectar puntos ruidosos y no necesita el número de clusters.

3.3. Etapa 3: Clasificación, caracterización y análisis de los resultados

3.3.1. Clasificación de los clientes y caracterización

El método DBSCAN se implementa tras llevar a cabo una clasificación previa basada en los estratos establecidos en la Tabla 1.2, ya que este método es más efectivo para identificar el comportamiento predominante en el clúster. Por otro lado, el GMM se utiliza con una categorización más simple, debido a que no necesita un enfoque tan específico como el basado en los estratos, en su lugar se dividen los usuarios basándose en el grupo de consumo al que pertenece (adicionando la categoría de residencial con cocinas de inducción), debido a su mejor diferenciación entre los distintos tipos de comportamiento dentro del grupo.

3.3.2. Gráficos y análisis

Con la visualización de los perfiles de carga de los estratos obtenidos por cada método y datos puntales, como el factor de carga, la energía consumida u observaciones que posee cada uno de los estratos, con estos datos se puede comparar la efectividad de estos en la obtención de conclusiones apreciables en la definición del método óptimo entre estos, finalizando con la realización de una clasificación y caracterización de los clientes satisfactoria.

3.3.3. Análisis de los resultados en el dimensionamiento de estaciones de transformación

El método que mejores resultados consigue es el GMM, para valorar su utilidad en el dimensionamiento de estaciones de transformación se lo aplica tanto en un transformador como en los clientes del mismo y se comparan sus perfiles de carga.

4. CAPÍTULO 4: RESULTADOS Y CONCLUSIONES

4.1. Resultados

4.1.1. Métodos de Prueba

Se aplican los métodos K-means, Mezcla Gaussiana y DBSCAN al estrato RD1, como se observa en la Figura 11, estos métodos ofrecen resultados diversos puntualizados en la Tabla 4.1, como la cantidad de clústers o la energía de cada uno.

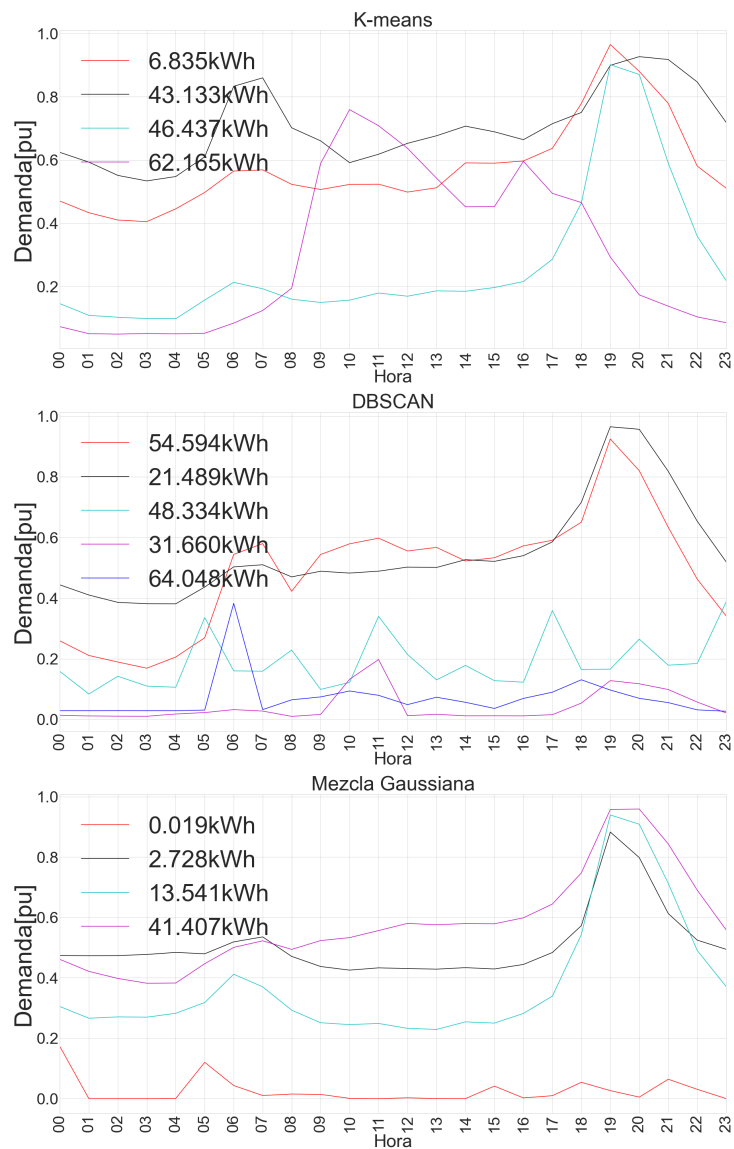


Figura 11: Perfiles de carga de los diversos clústers generados por los métodos en el estrato RD1.

En la Figura 12 se muestra una comparación del perfil de carga del clúster más representativo del estrato RD1 generado por cada método. Estos clústers son seleccionados mediante la identificación del comportamiento del CEE predominante del estrato, respaldándose en la información presentada en la Figura 11. Además, se utiliza la cantidad de muestras y la energía del perfil de carga de cada clúster, los cuales se presentan en la Tabla 4.1.

Para analizar cuál de los métodos mencionados en esta sección es el más idóneo, se utiliza el perfil de carga del clúster más representativo del estrato de cada método, como se muestra en la Figura 12, y los datos expuestos en la Tabla 4.1, como el porcentaje de representación del estrato. Este porcentaje indica la proporción de observaciones pertenecientes al clúster más representativo del estrato con respecto al total de observaciones analizadas por el método.

Tabla 4.1: Tabla de comparación de los métodos aplicados

Método	Clústers	E [kWh/mes]	Observaciones	Representación del estrato [%]
K-means	4	6.835	1647	31.559
		43.133	1008	
		46.437	407	
		62.165	132	
DBSCAN	5	54.594	390	92.375
		21.489	2793	
		46.334	5	
		31.660	3	
Mezcla Gaussiana	4	0.019	436	46.024
		2.728	482	
		13.541	616	
		41.407	1308	

Como se aprecia en la Tabla 4.1 los métodos K-means y Mezcla Gaussiana generan la misma cantidad de clústers, pero no contienen la misma distribución en las observaciones. De los 3 métodos el que mejor porcentaje de representación tiene del estrato es el método DBSCAN, pero la energía proyectada es la menor de todos, ya que concentra su distribución

en un clúster principal, utilizando a los otros como una forma de aislar observaciones ajenas al comportamiento predominante del estrato, por esta razón hay clústers que contienen tan pocas observaciones. K-means crea los clústers de más alto valor de energía, pero al compararlo con GMM la distribución de las observaciones son más diversas.

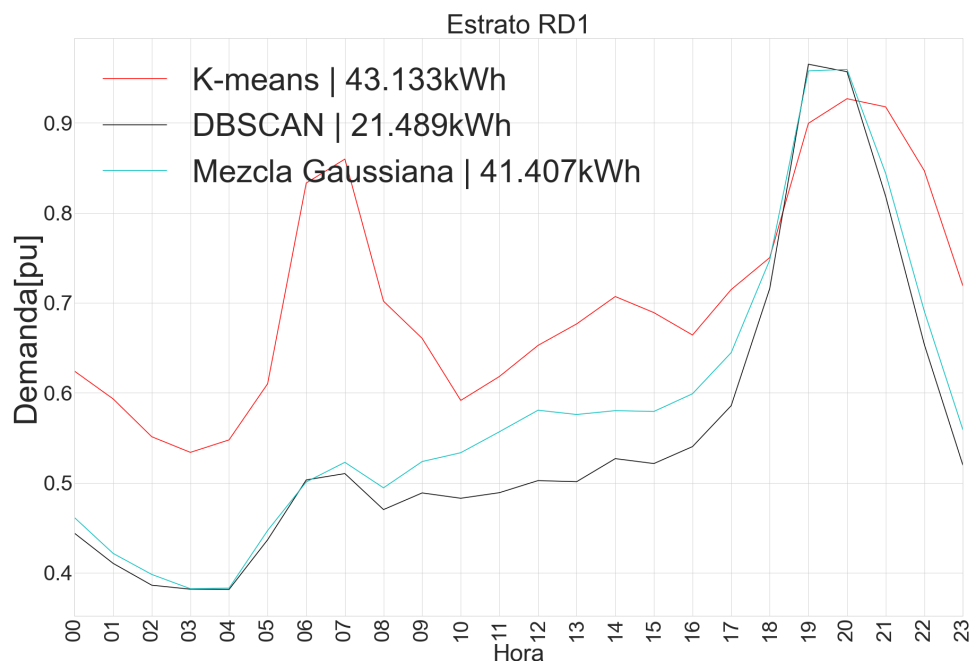


Figura 12: Perfiles de carga del estrato RD1 obtenidos de la aplicación de los 3 métodos.

Se concluye que el método DBSCAN es considerado el mejor para la representación del perfil de carga de un estrato, pero no realiza una clasificación adecuada del mismo. Por otro lado, GMM logra una adecuada clasificación del estrato, sin embargo, no es eficiente al momento de proporcionar una representación única del mismo. Por tanto, se aplican ambos métodos de forma diferenciada con el fin de evaluar cuál de ellos realiza una mejor clasificación y caracterización de los clientes.

4.1.2. Comparación de los resultados obtenidos por los métodos aplicados y los de la Centrosur

Mediante un análisis comparativo de los métodos utilizados y los datos obtenidos en el estudio de investigación denominado “Investigación y caracterización de la Carga - Muestreo Aleatorio por Estratos” [1] realizado por la Centrosur (EERCSCA), se identificará cuál de estas técnicas genera una mejor clasificación y caracterización. Se realizará la creación de tablas

puntualizando datos de los métodos empleados, como se muestra en la Tabla 4.2, con el objetivo de obtener conclusiones cuantificables.

Tabla 4.2: Resumen de resultados

Método	Observaciones	Estratos	Cantidad
DBSCAN	7480	Residencial	8
		Comercial	5
		Industrial	2
		Otros	1
Centrosur	1214	Residencial	5
		Comercial	5
		Industrial	3
		Otros	3
Mezcla Gaussiana	8487	Residencial	7
		Comercial	4
		Industrial	3
		Otros	3

En la Tabla 4.2 se observa la totalidad de 16 estratos formados para DBSCAN y Centrosur, mientras que se generan 17 para el GMM. El método DBSCAN produce una distribución de estratos desequilibrada, ya que concentra gran parte de su clasificación en solo 2 grupos de consumo. Por otro lado, los otros 2 métodos se distribuyen de manera más uniforme, siendo el estudio de la Centrosur el mejor en términos de equilibrio de estratos. Sin embargo, vale mencionar que este estudio contiene la menor cantidad de observaciones. Por otro lado, el GMM posee la mayor cantidad de observaciones y también logra una buena clasificación de sus clientes. Cabe destacar que en el grupo de consumo residencial existe una categoría adicional en comparación con la Centrosur, que corresponde a residencial con cocinas de inducción.

Los perfiles de carga de los estratos generados por DBSCAN y GMM están expuestos en el Anexo I, específicamente en Anexo I.I se muestran los resultantes con DBSCAN y en el Anexo I.II con GMM.

Los datos de comparación principales son la energía proyectada y el factor de carga, los cuales se calculan de la siguiente forma:

$$E = \frac{\Sigma P * \left(\frac{\text{Intervalo de minutos}}{60 \text{min}} \right)}{\text{Días} * \left(\frac{365 \text{ días}}{12 \text{ meses}} \right)}, Fc = \left(\frac{P_{pro}}{P_{max}} \right) * 100 \quad (4.1)$$

Mientras que la reducción de observaciones es dada por el porcentaje de las muestras que le pertenecen al estrato eliminado.

Tabla 4.3: Comparación de Estratos Residenciales

Métodos	Estratos	Observaciones	E [kWh/mes]	Fcarga [%]	Reducción de Observaciones [%]
DBSCAN	RD1	2793	21.5	55.0	33.0
	RD2	1991	80.0	57.2	
	RD3	1662	135.8	62.1	
	RD4	1355	228.3	69.6	
	RD5	662	481.1	74.8	
Centrosur	Residencial 1	161	60	49.3	24.366
	Residencial 2	212	110	54.2	
	Residencial 3	173	180	58.2	
	Residencial 4	116	310	65.6	
	Residencial 5	48	500	75.2	
Mezcla Gaussiana	Residencial 1	1455	4.5	42.1	16.6
	Residencial 2	1919	44.0	56.7	
	Residencial 3	2544	92.4	56.7	
	Residencial 4	2052	170.9	65.2	
	Residencial 5	775	346.5	73.4	

Mediante el análisis de la Tabla 4.3 y la Figura 13, se examinan los estratos formados en el grupo de consumo residencial por los métodos DBSCAN, Mezcla Gaussiana y el estudio de la Centrosur. Se elimina un estrato de cada método para realizar la comparación.

En el caso del método DBSCAN, se generan 5 estratos antes de la eliminación. Los estratos RD2 y RD3 no muestran diferencias apreciables entre sí, y al compararlos con el GMM, se determina que las observaciones contenidas en estos estratos se distribuyen en otros estratos.

En cuanto al GMM, también se obtiene la misma cantidad de estratos antes de la eliminación. Sin embargo, en el estrato Residencial 1, la energía

proyectada es menor a 5 kWh, por lo que se descarta y se reorganizan los demás estratos formados.

Por otro lado, los datos de Centrosur presentan los mismos inconvenientes que el método DBSCAN, como se muestra en la Tabla 4.3.

De los 3 métodos aplicados el que mejor resultado de clasificación muestra es Mezcla Gaussiana, se establece que 4 estratos son la distribución correcta del grupo de consumo, como se observa en la Tabla 4.3 el GMM tiene la mejor distribución de sus observaciones, parece representar mejor su diferencia entre estratos mediante la energía, evitando generar estratos relativamente cercanos.

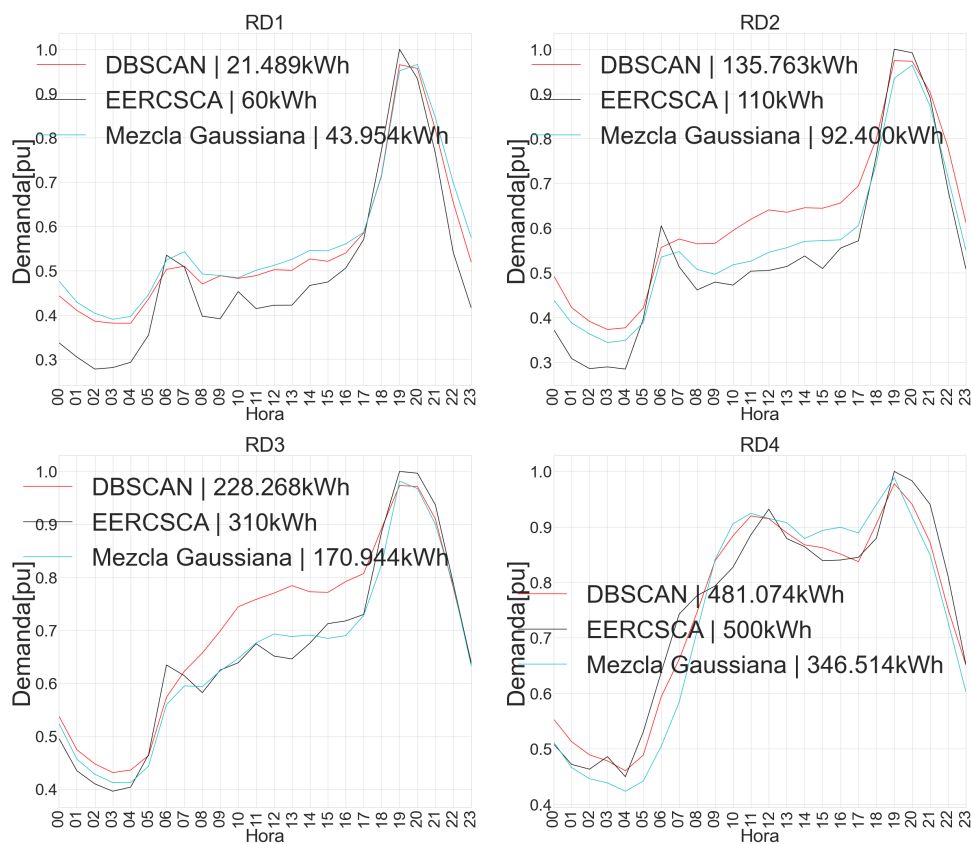


Figura 13: Perfiles de carga de los estratos residenciales.

En la Figura 13, se muestran los centroides o media de cada clúster representando a los perfiles de carga, los cuales corresponden a los estratos finales del grupo de consumo residencial. Es evidente que tanto los métodos DBSCAN como Mezcla Gaussiana generan una buena estabilidad en las formas de onda. Esto confirma la eficiencia de estos métodos de agrupación, ya que muestran una similitud cercana a los resultados de Centrosur.

Cuando se comparan únicamente los perfiles de consumo generados

por los métodos Mezcla Gaussiana y DBSCAN, se observa que tienen comportamientos similares. Sin embargo, en términos de energía proyectada, el GMM muestra una mejor respuesta en todos sus estratos. Por otro lado, el método DBSCAN no logra obtener buenos resultados para los estratos de nivel bajo.

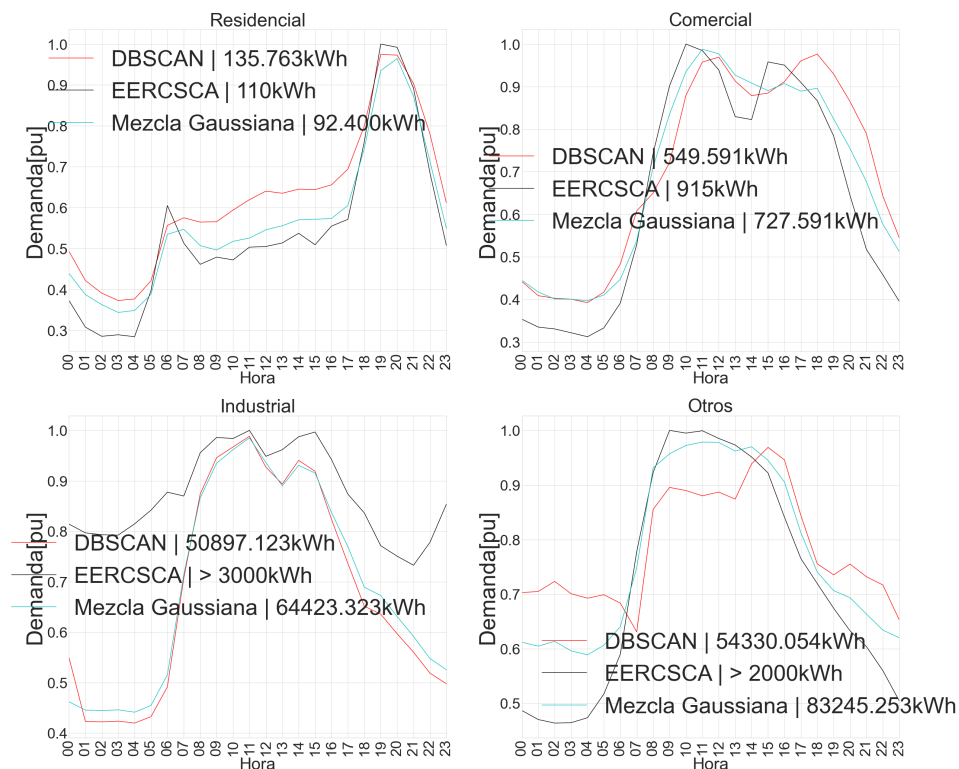


Figura 14: Perfiles de carga de los distintos grupos de consumo.

En la Figura 14 se visualiza una comparación entre los perfiles de carga generados por los 3 métodos en algún estrato de cada grupo de consumo, en el comportamiento mantiene una cierta semejanza entre sí en cada uno de los grupos de consumo, pero la energía sí difiere entre cada método, siendo los estratos más altos en los que mayor discrepancia existe. Como se evidencia en la Tabla 4.3 Centrosur y GMM cuentan con una mayor diversidad en cuanto a cantidad de estratos industrial y otros, pudiéndose notar un claro cambio en los niveles de energía de cada estrato entre sí.

Con respecto a todo lo analizado anteriormente se concluye que la Mezcla Gaussiana es el método óptimo para la realización de la clasificación y caracterización de los clientes, ya que su clasificación es la que más se acerca a lo expuesto en el trabajo de investigación de la Centrosur y evidencia el cambio que existe dentro de los grupos de consumo y estratos dentro de los mismos.

4.1.3. Análisis de los resultados de Mezcla Gaussiana

Al definir Mezcla Gaussiana como el método con mejor clasificación y caracterización, se procede a realizar un análisis de sus estratos y los perfiles de carga en la región Sierra, Oriente y en los diferentes tipos de días.

Tabla 4.4: Datos de los estratos de la región Sierra en tipo de día laborable

Estratos	Obser	EPmax [kW]	PPmax [kW]	PPmin [kW]	Ppro [kW]	E [kWh/mes]	Fc [%]
Residencial 1	1919	1.16	0.11	0.04	0.06	44	56.7
Residencial 2	2544	2.72	0.22	0.08	0.13	92.4	56.7
Residencial 3	2052	2.88	0.36	0.14	0.23	170.9	65.2
Residencial 4	775	2.91	0.65	0.27	0.48	346.5	73.4
Resi Coc 1	118	3.79	0.13	0.05	0.08	56.8	61.4
Resi Coc 2	134	4.38	0.50	0.12	0.29	212.9	58.9
Resi Coc 3	36	4.63	0.99	0.35	0.58	421.8	58.6
Comercial 1	250	7.57	1.44	0.56	1.00	727.6	69.4
Comercial 2	67	218.88	45.50	17.76	31.46	22965.9	69.2
Comercial 3	10	241.50	210.94	53.95	127.39	92992.2	60.4
Comercial 4	4	316.43	300.84	66.29	174.37	127292.7	58
Industrial 1	152	1096.59	21.94	1.439	6.69	4886.1	30.5
Industrial 2	201	626.77	127.51	55.55	88.25	64423.3	69.2
Industrial 3	55	7561.84	2582.18	2131.64	2404.89	1755569.1	93.1
Otros 1	85	12.47	2.00	0.66	1.19	869.4	59.6
Otros 2	52	74.86	23.10	12.14	17.86	13036.7	77.3
Otros 3	33	287.41	148.078	86.40	114.04	83245.3	77

La Tabla 4.4 contiene datos relevantes de los estratos, como la energía proyectada y el factor de carga. Estos datos en conjunto con los perfiles de carga de cada estrato, resulta en una descripción de la clasificación de los clientes, donde se denota que diferencias existen entre grupos de consumo y los estratos dentro de los mismos.

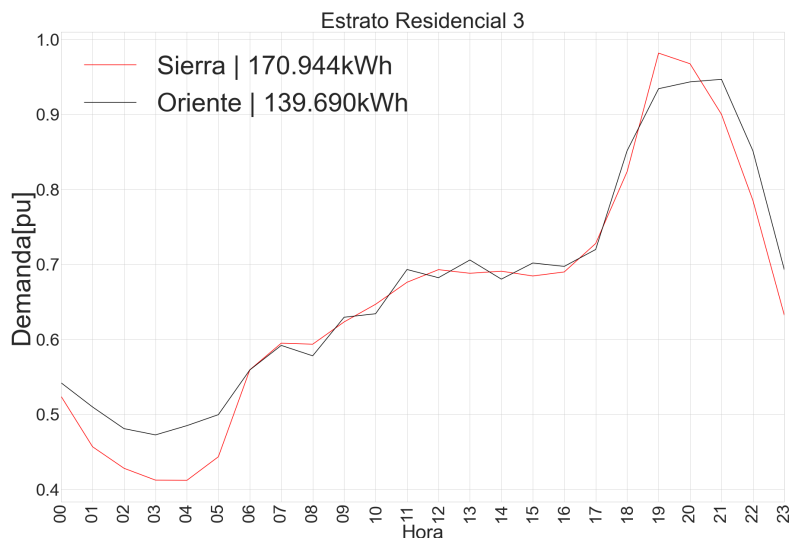


Figura 15: Perfiles de carga residencial de la región Sierra y Oriente

La Figura 15 muestra como la región Sierra posee un comportamiento más estable y por ende un mejor factor de carga, pero la región Oriente cuenta con un menor número de observaciones, por lo tanto, también tiene menos estratos y solo cuenta con 2 grupos de consumo, el residencial y otros.

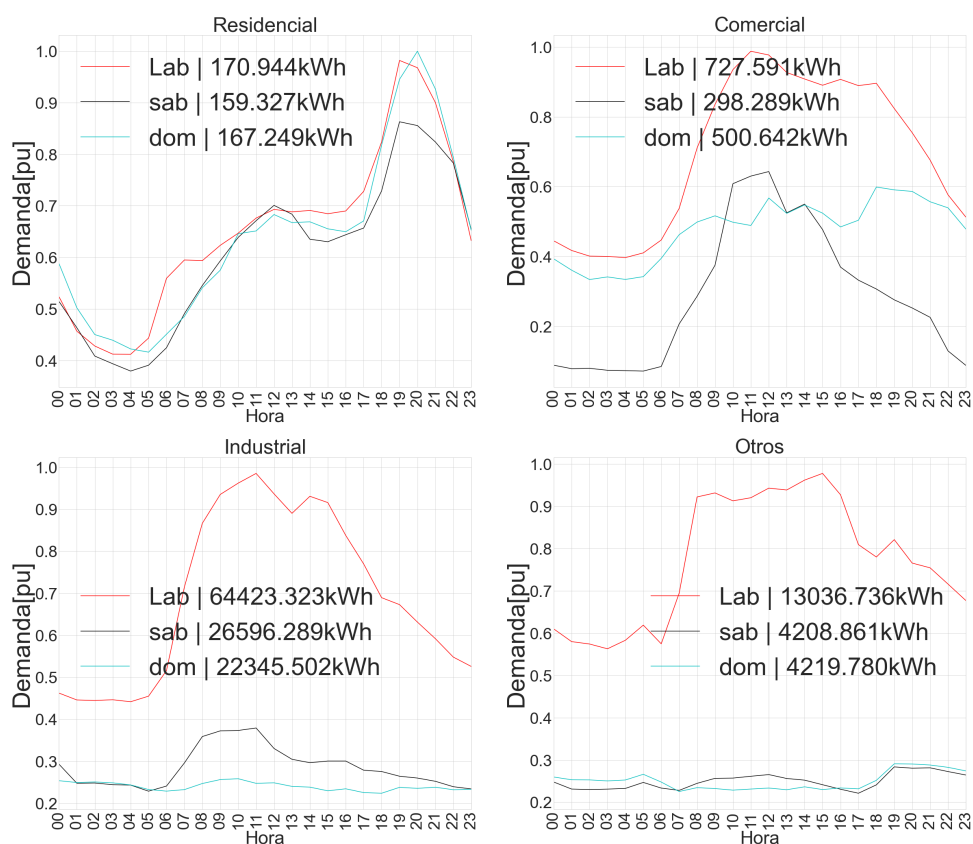


Figura 16: Perfiles de carga de los grupos de consumo en diferentes tipos de día

En la Figura 16 se muestran los comportamientos de un estrato de cada grupo de consumo en diferentes tipos de días, donde se observa que en los perfiles residenciales tienen comportamientos similares. Sin embargo, se distingue que los días no laborables (tipos de día sábado y domingo) tienen un menor consumo de energía, siendo sábado donde se produce la mayor caída de energía, este comportamiento se replica en todos los grupos de consumo, con excepción del perfil industrial siendo el único que presenta su mayor caída de consumo en los días domingos. Solo residencial mantiene su forma de onda, los otros grupos de consumo sufren cambios apreciables con el paso de los tipos de días.

4.2. Valoración del método en el dimensionamiento de estaciones de transformación

La Tabla 4.5 expone los datos de energía proyectada, cantidad de clientes y factor de carga del transformador Nro. 31240 de 37.5 kVA y de los diferentes estratos generados por el CEE de cada cliente del mismo. Se observa que la suma de la cantidad de clientes en cada estrato compone el total del transformador, por lo cual se tiene una información más detallada del comportamiento del CEE de los clientes en el transformador.

Tabla 4.5: Tabla de datos del transformador 31240 y sus tipos de clientes

Elemento	E [kWh/mes]	Clientes	Fc [%]
Trafo 31240	9443.85	71	67.948
Residencial 1	43.95	40	56.7
Residencial 3	170.94	23	65.2
Residencial 4	346.51	5	73.4
Comercial 1	727.59	3	61.4

En la Figura 17 se exponen los perfiles de carga obtenidos por la aplicación del GMM en sus clientes. Se puede apreciar cómo se formaron cuatro grupos diferentes de clientes cada uno perteneciente a un estrato con su propio perfil de carga, los cuales al sumarse deberían resultar en el perfil de carga del transformador.

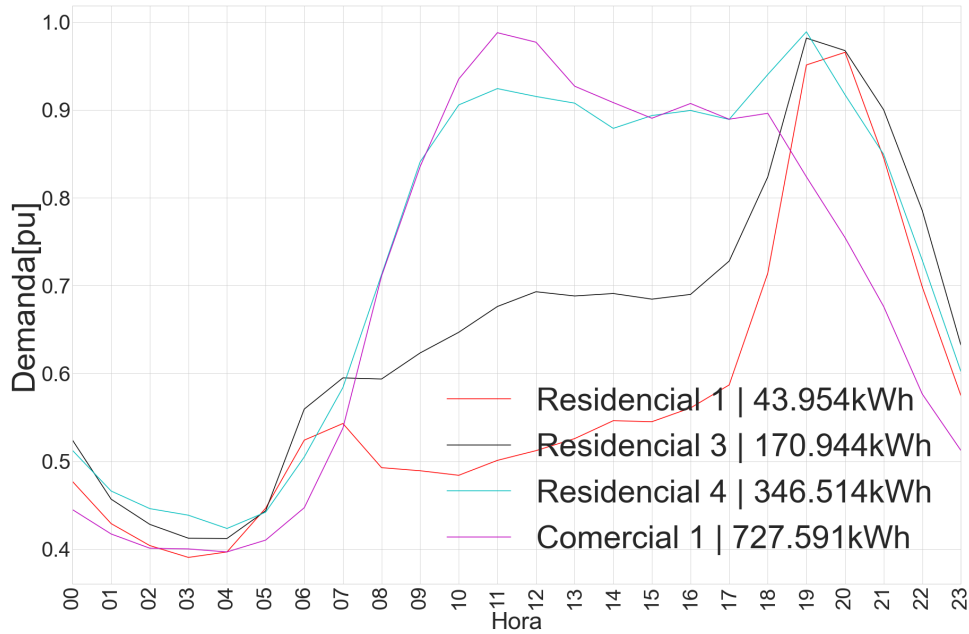


Figura 17: Perfiles de carga de los diferentes tipos de clientes del transformador 31240.

En la Figura 18 se muestra los perfiles de carga de la suma resultante de todos los clientes y el del transformador 31240, donde se aprecia que tienen un consumo de energía relativamente cercano, sus perfiles de carga son similares.

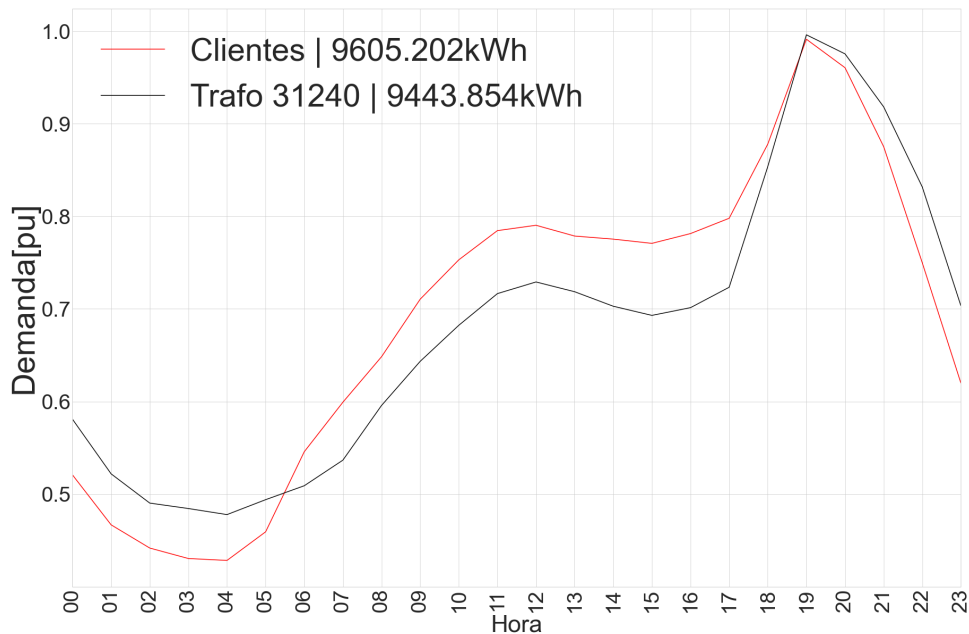


Figura 18: Perfil de carga resultante de los clientes y Perfil de carga del transformador 31240.

El GMM resulta eficiente para realizar un análisis más preciso sobre el comportamiento de los clientes dentro del transformador. Al desglosar el

perfil del transformador en grupos, se pueden tomar mejores decisiones al momento de dimensionarlo e incluso redistribuir los clientes a diferentes transformadores si es necesario.

4.3. Discusión

El GMM propuesto en este trabajo obtiene mejores resultados en la comparación con el estudio existente en la Centrosur, “Investigación y Caracterización de la carga - Muestreo Aleatorio por Estratos” [1]. Los mejores resultados se deben a su capacidad de capturar y modelar patrones irregulares o comportamientos atípicos que pueden existir en los datos. Al identificar y separar estos perfiles de carga, se logra una clasificación mejorada de los clientes en grupos de consumo específicos. Si bien se consiguen factores y perfiles de carga similares estos tienen una forma más estable [1].

El método utilizado en el trabajo “Characterization and classification of daily electricity consumption profiles: shape factors and K-means clustering technique” [6] no resulta tan efectivo para este caso, debido al enfoque del mismo, ya que se limita a la caracterización de un solo cliente con un gran número de datos de análisis. Por tanto, las observaciones que analiza tienen comportamientos muy similares, pero en el caso de estudio se tienen muchos clientes con observaciones con comportamientos muy diferentes entre sí; sin embargo, poseen ciertas tendencias o niveles que los encajan dentro de una determinada clasificación, además de existir observaciones con comportamientos atípicos.

La aplicación de este método en este caso resulta ineficiente, ya que genera clústers innecesarios debido a que no posee un filtro contra observaciones atípicas; no obstante, este método sería muy efectivo para un análisis desarrollado de cada cliente de forma individual, el GMM en cambio, es efectivo para el contexto del caso de estudio, ya que aísla observaciones atípicas y agrupa de forma adecuada los comportamientos dentro de los grupos de consumo, pero no es efectivo para reconocer los diferentes tipos de días y regiones por sí solo [6].

Por otro lado, en el trabajo “Household energy consumption segmentation using hourly data” [57] se propone un método para hacer un análisis de diferentes clientes basado en la implementación de K-means con un umbral de comprobación de clúster. A primera vista sería muy eficiente, sin embargo, desglosa en muchas partes el estrato a causa de la naturaleza de los perfiles

irregulares del mismo, debido a esto genera un análisis deficiente. El GMM aísla muchas de estas observaciones irregulares y su asignación probabilística provoca que las observaciones obtengan una mejor distribución, por lo cual hay una clasificación adecuada y sin clústers innecesarios [57].

4.4. Conclusiones

Al analizar los métodos de agrupación, se llega a la conclusión que la utilización del método de Mezcla Gaussiana resulta adecuado debido a la presencia de clientes con comportamientos irregulares en cada grupo de consumo.

El método de GMM aprovecha la desviación estándar como filtro para eliminar clientes con consumos muy variables y realiza una asignación probabilística, permitiendo que cada dato pueda pertenecer a cualquiera de los clústers.

Se logra una mayor precisión en la segmentación y clasificación de los datos, lo que resulta fundamental para comprender y gestionar adecuadamente los diferentes perfiles de consumo.

El GMM tiene habilidades destacables en términos de clasificación de clientes, pero por sí solo no logra un desempeño óptimo, ya que tiene una falta de distinción entre las observaciones de diferentes regiones, lo que impide identificar comportamientos internos específicos de cada grupo. Además, existen relaciones entre clientes residenciales y comerciales que deben tenerse en cuenta.

El GMM presenta ventajas en cuanto a perfiles de carga y clasificación de clientes en comparación con los demás métodos utilizados.

Con el estudio se puede identificar el CEE y se puede ayudar a la Centrosur a promover prácticas de eficiencia energética.

4.5. Recomendaciones

Se sugiere realizar un análisis individual de cada cliente, basado en al menos un año de datos, para lograr una clasificación más precisa y depurada. Aunque los resultados generales pueden no variar significativamente, se considera valioso comprobar esta idea.

Se recomienda obtener más datos de otros clientes para ampliar la muestra de cada estrato y así realizar un análisis más exhaustivo de aquellos estratos con un número limitado de observaciones.

Idear un proceso para actualizar los datos de los clientes con las nuevas campañas de mediciones que se realicen.

Referencias

- [1] S. Zambrano and M. Molina, “Investigación y caracterización de la carga - muestreo aleatorio por estratos,” in *EMPRESA ELÉCTRICA REGIONAL CENTRO SUR C.A.* DIPLA, 2016, pp. 1–45.
- [2] D. Pascual, F. Pla, and S. Sánchez, “Algoritmos de agrupamiento,” *Método Informáticos Avanzados*, pp. 164–174, 2007.
- [3] A. Sanz Suárez, “Evaluación de una nueva metodología para la estimación del transvase de votos entre elecciones,” *Archivo digital UPM Universidad Politécnica de Madrid*, 2018.
- [4] (2023) Categorías tarifarias. Categorías Tarifarias. [Online]. Available: <https://www.centrosur.gob.ec/>
- [5] G. A. Guamán Figueroa, “Mejora del pronóstico de carga a corto plazo en el sistema de distribución de la empresa eléctrica ambato regional centro norte sa (eensa) basada en criterios de big data,” B.S. thesis, Quito, 2019., 2019.
- [6] M.-A. Milton, C.-O. Pedro, S.-G. Xavier, and E.-E. Guillermo, “Characterization and classification of daily electricity consumption profiles: shape factors and k-means clustering technique,” in *E3S Web of Conferences*, vol. 64. EDP Sciences, 2018, p. 08004.
- [7] A. Jurio, H. Bustince, and V. Torra, “Clustering difuso con pertenencias intervalares,” in *XVIII Conferencia de la Asociación Española para la Inteligencia Artificial (CAEPIA 2018): avances en Inteligencia Artificial. 23-26 de octubre de 2018 Granada, España.* Asociación Española para la Inteligencia Artificial (AEPIA), 2018, pp. 300–305.
- [8] D. Huang, C.-D. Wang, H. Peng, J. Lai, and C.-K. Kwoh, “Enhanced ensemble clustering via fast propagation of cluster-wise similarities,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 1, pp. 508–520, 2018.
- [9] J. Li, D. Li, and Y. Zhang, “Efficient distributed data clustering on spark,” in *2015 IEEE International Conference on Cluster Computing.* IEEE, 2015, pp. 504–505.

- [10] F. Aldhubaib, “Generic aircraft model recognition by two shape factors: in the resonance region,” *IET Radar, Sonar & Navigation*, vol. 14, no. 1, pp. 81–88, 2020.
- [11] M. B. Roberts, N. Haghdadi, A. Bruce, and I. MacGill, “Characterisation of australian apartment electricity demand and its implications for low-carbon cities,” *Energy*, vol. 180, pp. 242–257, 2019.
- [12] S. Yilmaz, J. Chambers, and M. K. Patel, “Comparison of clustering approaches for domestic electricity load profile characterisation-implications for demand side management,” *Energy*, vol. 180, pp. 665–677, 2019.
- [13] A. Janßen and P. Wan, “K-means clustering of extremes,” 2020.
- [14] K. P. Sinaga and M.-S. Yang, “Unsupervised k-means clustering algorithm,” *IEEE access*, vol. 8, pp. 80 716–80 727, 2020.
- [15] J. Hämmäläinen, T. Kärkkäinen, and T. Rossi, “Improving scalable k-means++,” *Algorithms*, vol. 14, no. 1, p. 6, 2020.
- [16] J. Wang, K. Wang, R. Jia, and X. Chen, “Research on load clustering based on singular value decomposition and k-means clustering algorithm,” in *2020 Asia Energy and Electrical Engineering Symposium (AEEES)*. IEEE, 2020, pp. 831–835.
- [17] H. B. Tambunan, D. H. Barus, J. Hartono, A. S. Alam, D. A. Nugraha, and H. H. H. Usman, “Electrical peak load clustering analysis using k-means algorithm and silhouette coefficient,” in *2020 International Conference on Technology and Policy in Energy and Electric Power (ICT-PEP)*. IEEE, 2020, pp. 258–262.
- [18] L. Marrero, D. Carrizo, L. García-Santander, and F. Ulloa-Vásquez, “Uso de algoritmo k-means para clasificar perfiles de clientes con datos de medidores inteligentes de consumo eléctrico: Un caso de estudio,” *Ingeniare. Revista chilena de ingeniería*, vol. 29, no. 4, pp. 778–787, 2021.
- [19] A. Ullah, K. Haydarov, I. Ul Haq, K. Muhammad, S. Rho, M. Lee, and S. W. Baik, “Deep learning assisted buildings energy consumption profiling using smart meter data,” *Sensors*, vol. 20, no. 3, p. 873, 2020.

- [20] L. F. Siavichay Cumbe, “Aplicación de métodos estadísticos para el análisis y predicción de perfiles de consumo de energía eléctrica,” B.S. thesis, 2018.
- [21] A. de regulación y control de energía y recursos naturales no renovables, *Estadística Anual y Multianual del Sector Eléctrico Ecuatoriano*, 2022.
- [22] S. J. Vest, “Estimating kw demand for future loads on rural distribution systems,” *Transactions of the American Institute of Electrical Engineers. Part III: Power Apparatus and Systems*, vol. 76, no. 3, pp. 652–655, 1957.
- [23] L. Bayas, M. Jaramillo, E. Betancourt, and J. P. Reinoso, “La tarifa horaria en el ecuador como incentivo de eficiencia energética,” 2009.
- [24] A. de Regulación y Control de Energía y Recursos Naturales no Renovables, “Pliego tarifario del servicio público de energía eléctrica,” in *Dirección de regulación económica y tarifas del sector eléctrico*, vol. 3, 2022, p. 39.
- [25] E. E. Betancourt Vasco, “Estudio y planteamiento para establecer una tarifa horaria en el pico del sistema eléctrico en el ecuador como incentivo de eficiencia energética,” B.S. thesis, QUITO/EPN/2012, 2012.
- [26] M. R. Valarezo Pardo, J. A. Honores Tapia, A. S. Gómez Moreno, and L. F. Vences Sánchez, “Comparación de tendencias tecnológicas en aplicaciones web,” *3c Tecnología: glosas de innovación aplicadas a la pyme*, vol. 7, no. 3, pp. 28–49, 2018.
- [27] (2023) Microsoft excel. Microsoft Excel. [Online]. Available: <https://www.microsoft.com/es/microsoft-365/excel>
- [28] (2023) learn. Learn. [Online]. Available: <https://datascientest.com/es/scikit-learn-decubre-la-biblioteca-python>
- [29] M. A. Arango González, “scikit-forecasts: Una librería en python para el pronóstico de series de tiempo no lineales.”
- [30] L. J. Aguilar, *Big Data, Análisis de grandes volúmenes de datos en organizaciones*. Alfaomega Grupo Editor, 2016.
- [31] M. Pérez, *BIG DATA-Técnicas, herramientas y aplicaciones*. Alfaomega Grupo Editor, 2015.

- [32] R. Kaur and D. Gabrijelčič, “Behavior segmentation of electricity consumption patterns: A cluster analytical approach,” *Knowledge-Based Systems*, vol. 251, p. 109236, 2022.
- [33] M. G. Gramajo, L. Ballejos, and M. Ale, “Software requirements engineering through machine learning techniques: A literature review,” in *2018 IEEE Biennial Congress of Argentina (ARGENCON)*. IEEE, 2018, pp. 1–7.
- [34] E. M. Rojas, “Machine learning: análisis de lenguajes de programación y herramientas para desarrollo,” *Revista Ibérica de Sistemas e Tecnologias de Informação*, no. E28, pp. 586–599, 2020.
- [35] C. A. Brusil Cruz, “Análisis comparativo entre aprendizaje supervisado y aprendizaje semi-supervisado para la clasificación de señales sísmicas vulcanológicas del volcán cotopaxi,” B.S. thesis, Quito, 2020., 2020.
- [36] J. A. Ibáñez Perea, “Análisis clúster aplicado al ámbito agrónomo para estudiar el comportamiento de nuevos tipos de compost.”
- [37] R. Vilà Baños, M. J. Rubio Hurtado, V. Berlanga Silvente, M. Torrado Fonseca *et al.*, “Cómo aplicar un cluster jerárquico en spss,” *REIRE: revista d’innovació i recerca en educació*, 2014.
- [38] P. Larranaga, I. Inza, and A. Moujahid, “Tema 14. clustering,” *País Vasco: Universidad del País Vasco*, 2012.
- [39] Á. S. Choez Franco, “Análisis de las características de los tipos de algoritmos de clustering en el aprendizaje no supervisado.” B.S. thesis, Babahoyo: UTB-FAFI. 2022, 2022.
- [40] J. M. Lozano Banqueri, “Creación y gestión de una base de datos con mysql y phpmyadmin,” 2018.
- [41] J. L. V. Villardón, “Introducción al análisis de clúster,” *Departamento de Estadística, Universidad de Salamanca. 22p*, 2007.
- [42] M. S. Amoroso Ordóñez and H. S. Ávila Flores, “Aplicación de las técnicas de agrupamiento para la distribución cuasi-óptima de una red híbrida wdm-tdm/pon en cascada multinivel que da soporte a una smart grid o smart city,” B.S. thesis, 2015.

- [43] Y. Li, Y. Zhang, Q. Tang, W. Huang, Y. Jiang, and S.-T. Xia, “tk-means: A robust and stable k-means variant,” in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 3120–3124.
- [44] A. Likas, N. Vlassis, and J. J. Verbeek, “The global k-means clustering algorithm,” *Pattern recognition*, vol. 36, no. 2, pp. 451–461, 2003.
- [45] J. Amat, “Clustering y heatmaps: aprendizaje no supervisado,” *Recuperado de https://rpubs.com/Joaquin_AR/310338*, 2017.
- [46] N. Sapkota, A. Alsadoon, P. Prasad, A. Elchouemi, and A. K. Singh, “Data summarization using clustering and classification: Spectral clustering combined with k-means using nfph,” in *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*. IEEE, 2019, pp. 146–151.
- [47] A. Sabzi, Y. Farjani, and M. ZiHayat, “An improved fuzzy k-medoids clustering algorithm with optimized number of clusters,” in *2011 11th International Conference on Hybrid Intelligent Systems (HIS)*. IEEE, 2011, pp. 206–210.
- [48] J. Pérez, M. Henriques, R. Pazos, L. Cruz, G. Reyes, J. Salinas, and A. Mexicano, “Mejora al algoritmo de agrupamiento k-means mediante un nuevo criterio de convergencia y su aplicación a bases de datos poblacionales de cáncer,” *Liver-2do Taller Latino Iberoamericano de Investigacion de Operaciones “la IO aplicada a la solución de problemas regionales*, pp. 1–7, 2007.
- [49] D. A. Reynolds *et al.*, “Gaussian mixture models.” *Encyclopedia of biometrics*, vol. 741, no. 659-663, 2009.
- [50] J. A. Rodrigo. (2020) Detección de anomalías con gaussian mixture model (gmm) y python», diciembre de 2020. Microsoft Excel. [Online]. Available: <https://www.cienciadedatos.net/documentos/py23-deteccion-anomalias-gmm-python.html>
- [51] A. García Herrero *et al.*, “Algoritmos para la estimación de modelos de mezclas gaussianas,” 2015.
- [52] S. I. Segovia Ortega, “Evaluación de métodos de agrupamiento de perfiles de carga para gestión de demanda en clientes de media tensión.” 2021.

- [53] N. Ohadi, A. Kamandi, M. Shabankhah, S. M. Fatemi, S. M. Hosseini, and A. Mahmoudi, “Sw-dbscan: A grid-based dbscan algorithm for large datasets,” in *2020 6th International Conference on Web Research (ICWR)*. IEEE, 2020, pp. 139–145.
- [54] S. Babichev, V. Lytvynenko, and V. Osypenko, “Implementation of the objective clustering inductive technology based on dbscan clustering algorithm,” in *2017 12th international scientific and technical conference on computer sciences and information technologies (csit)*, vol. 1. IEEE, 2017, pp. 479–484.
- [55] S. Jebari, A. Smiti, and A. Louati, “Af-dbscan: An unsupervised automatic fuzzy clustering method based on dbscan approach,” in *2019 IEEE International Work Conference on Bioinspired Intelligence (IWOBI)*. IEEE, 2019, pp. 000 001–000 006.
- [56] K. Khan, S. U. Rehman, K. Aziz, S. Fong, and S. Sarasvady, “Dbscan: Past, present and future,” in *The fifth international conference on the applications of digital information and web technologies (ICADIWT 2014)*. IEEE, 2014, pp. 232–238.
- [57] J Kwac, J Flora, and R Rajagopal, “Household energy consumption segmentation using hourly data,” *IEEE Trans Smart Grid*, vol. 5, no. 1, pp. 420–430, 2014.

Anexos

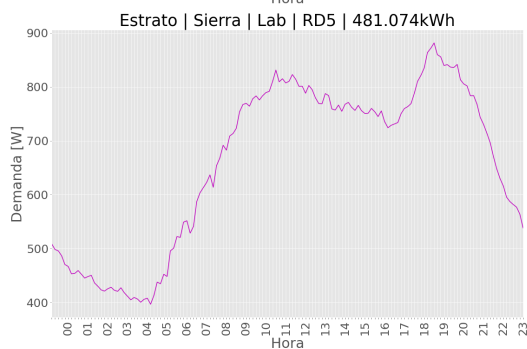
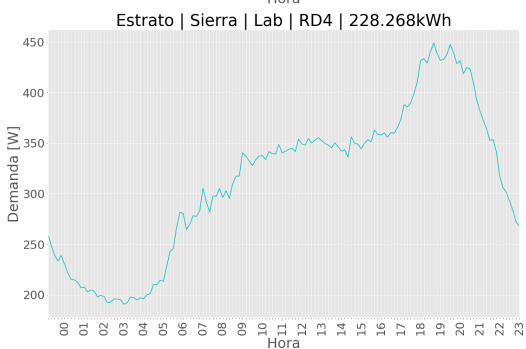
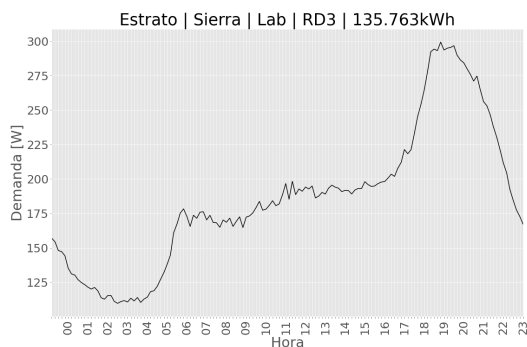
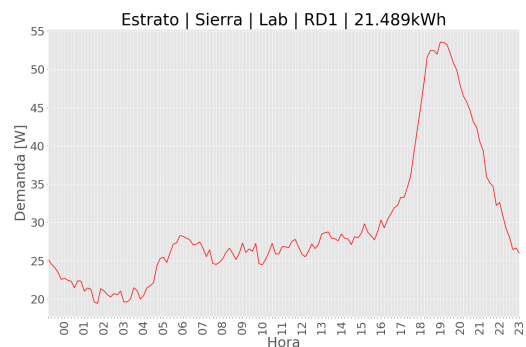
Anexo I: Resultados del programa.

Anexo I.I: Método DBSCAN.

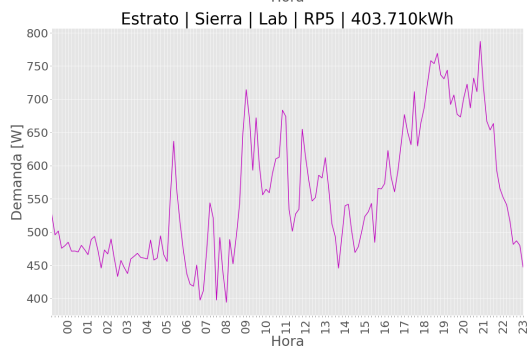
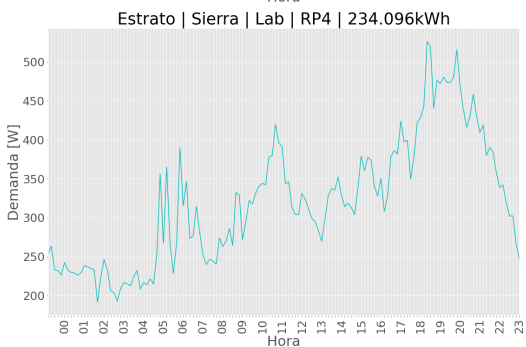
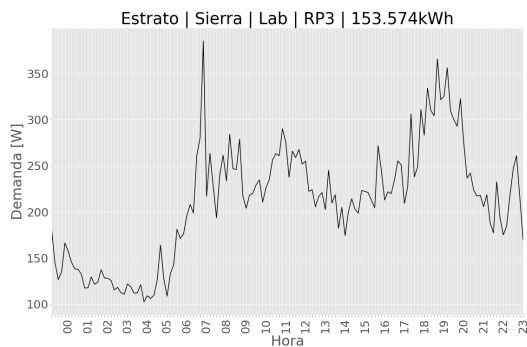
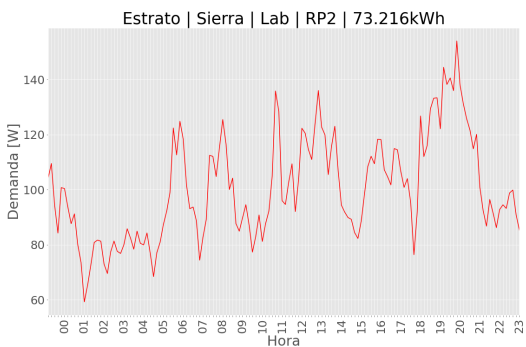
Datos de la Región Sierra en tipo de día Lab

Estratos	Obser	EPmax [kW]	PPmax [kW]	PPmin [kW]	Ppro [kW]	E [kWh/mes]	Fc [%]
RD1	2793	1.16	0.05	0.02	0.03	21.5	55
RD2	1662	4.36	0.30	0.11	0.19	135.8	62.1
RD3	1355	5.30	0.45	0.19	0.31	228.3	69.6
RD4	662	6.77	0.88	0.40	0.66	481.1	74.8
RP2	57	1.21	0.15	0.06	0.1	73.2	65.1
RP3	53	3.44	0.39	0.10	0.21	153.6	54.6
RP4	53	3.04	0.53	0.19	0.32	234.1	60.9
RP5	33	3.97	0.79	0.39	0.55	403.7	70.3
CB1	123	2.17	0.24	0.05	0.14	99.5	57.6
CB2	14	3.11	2.23	0.55	1.09	798.6	49.1
CO1	216	2.42	0.21	0.08	0.14	101.2	65.1
CO2	122	5.66	1.06	0.40	0.75	549.6	71
CH3	5	586.62	562.26	131.90	334.94	244502.3	59.6
IH3	268	1096.59	102.36	42.44	69.72	50897.1	68.1
IA1	35	3.90	0.53	0.13	0.27	196.6	50.5
OH3	29	353.60	94.68	57.13	74.43	54330.1	78.6

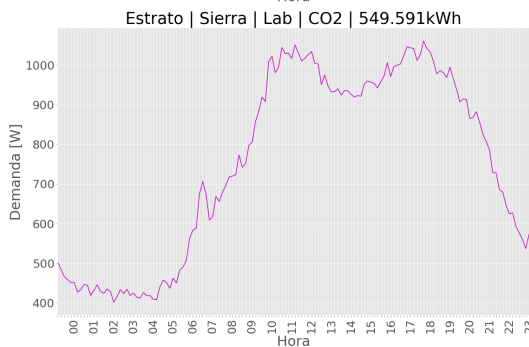
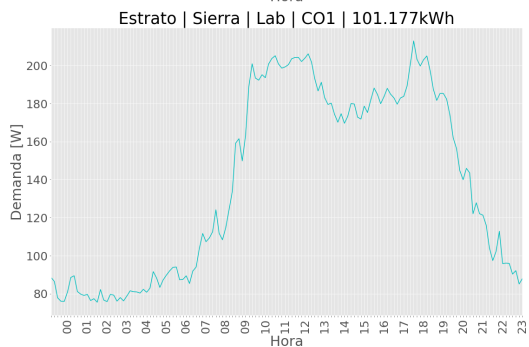
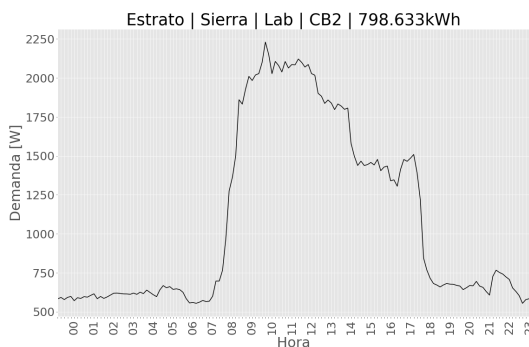
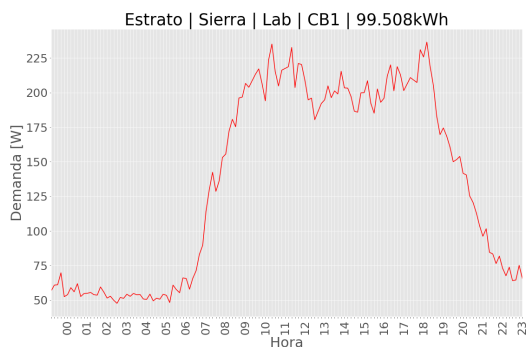
Perfiles de carga de cada Estrato de DBSCAN



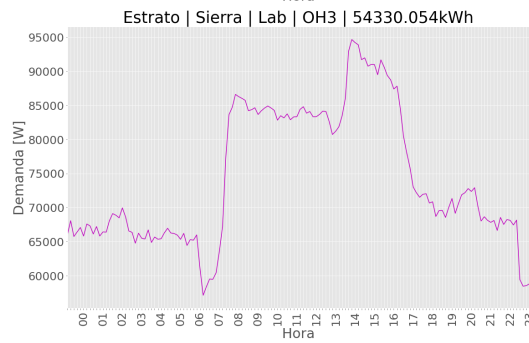
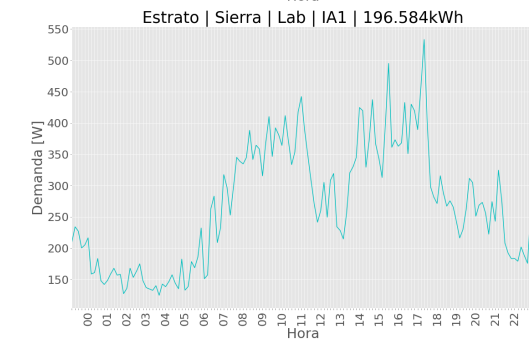
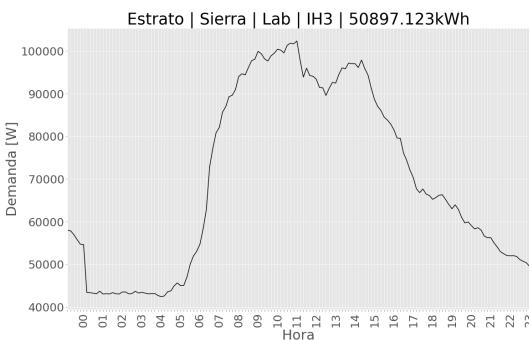
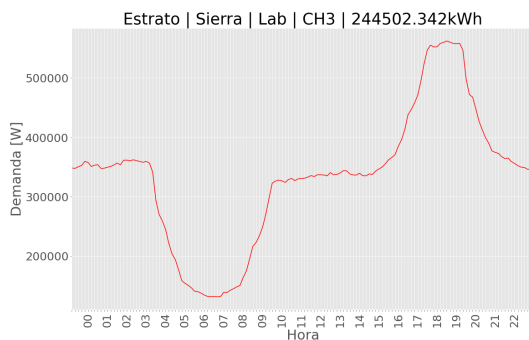
Perfiles de carga de cada Estrato de DBSCAN



Perfiles de carga de cada Estrato de DBSCAN



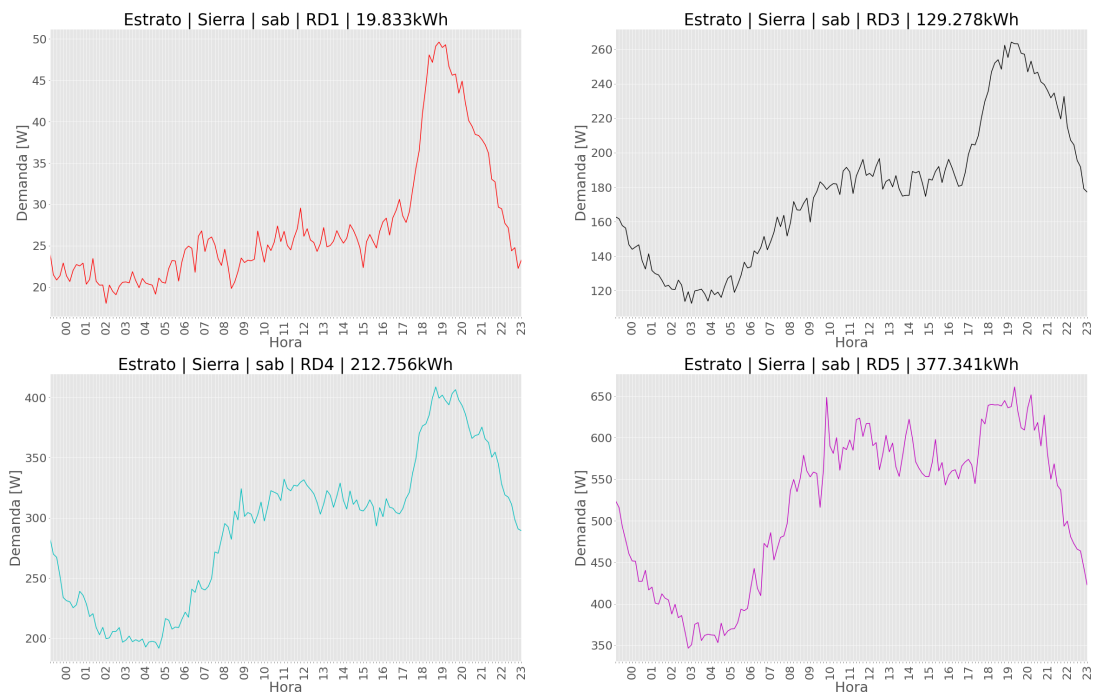
Perfiles de carga de cada Estrato de DBSCAN



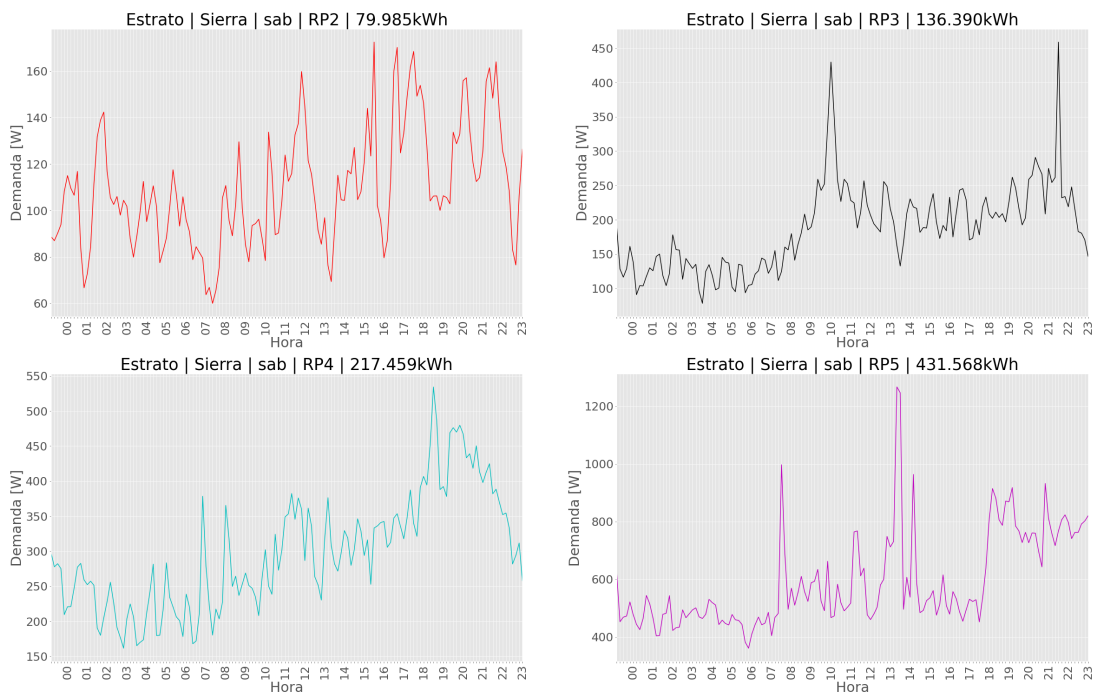
Datos de la Región Sierra en tipo de día sab

Estratos	Obser	EPmax [kW]	PPmax [kW]	PPmin [kW]	Ppro [kW]	E [kWh/mes]	Fc [%]
RD1	566	0.77	0.05	0.02	0.03	19.8	54.7
RD2	279	2.09	0.30	0.11	0.18	129.3	67
RD3	238	3.94	0.41	0.19	0.30	212.8	71.3
RD4	92	4.36	0.66	0.35	0.52	377.3	78.2
RP2	10	0.81	0.17	0.06	0.11	80	63.5
RP3	8	1.64	0.46	0.08	0.19	136.4	40.7
RP4	11	2.36	0.53	0.16	0.30	217.5	55.8
RP5	5	4.19	1.27	0.36	0.60	431.6	46.7
CB1	22	1.57	0.18	0.02	0.06	43.5	34.1
CB2	7	2.08	0.80	0.43	0.56	406.2	69.8
CO1	49	1.67	0.19	0.08	0.13	95.1	67.3
CO2	12	2.56	0.73	0.37	0.52	380.6	72
CH3	12	139.96	43.23	20.76	30.72	22425.5	71.1
IH3	57	544.33	74.25	47.08	57.77	42172.7	77.8
IA1	5	1.54	0.40	0.06	0.14	101	34.4
OH3	6	109.71	63.36	52.61	57.82	42208.2	91.3

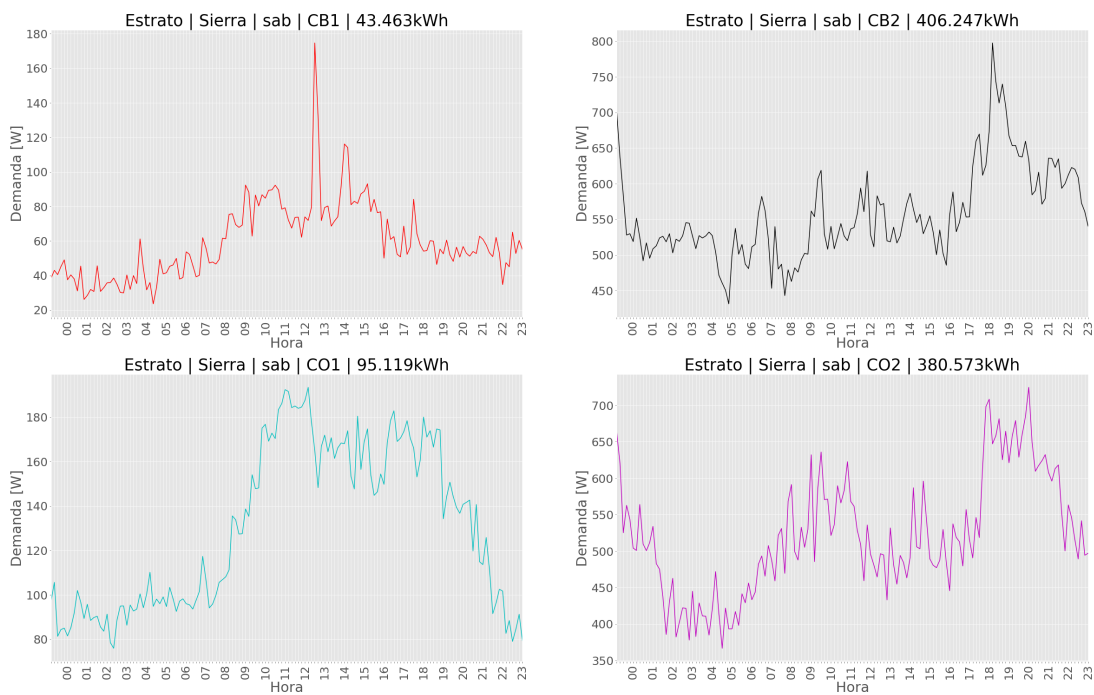
Perfiles de carga de cada Estrato de DBSCAN



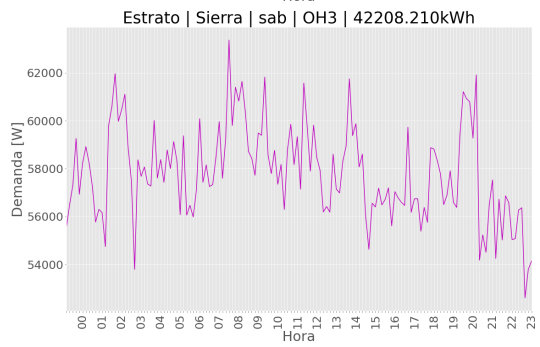
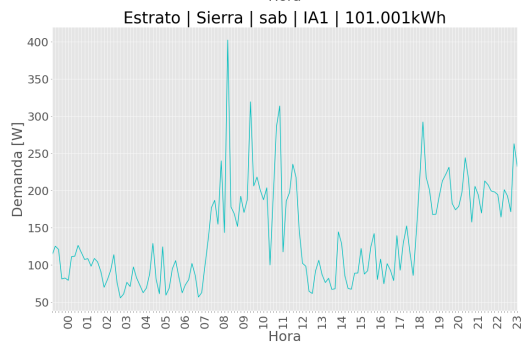
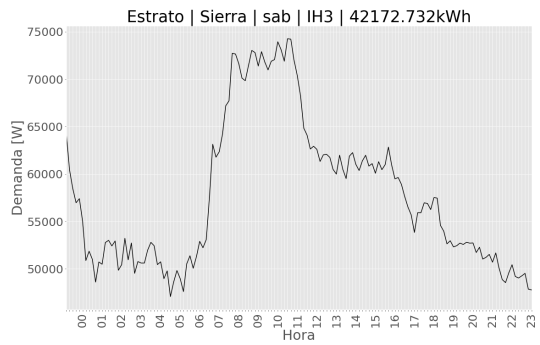
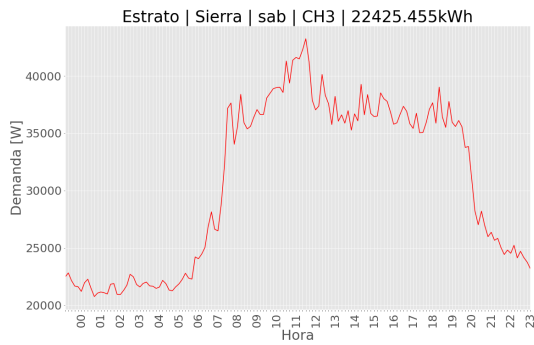
Perfiles de carga de cada Estrato de DBSCAN



Perfiles de carga de cada Estrato de DBSCAN



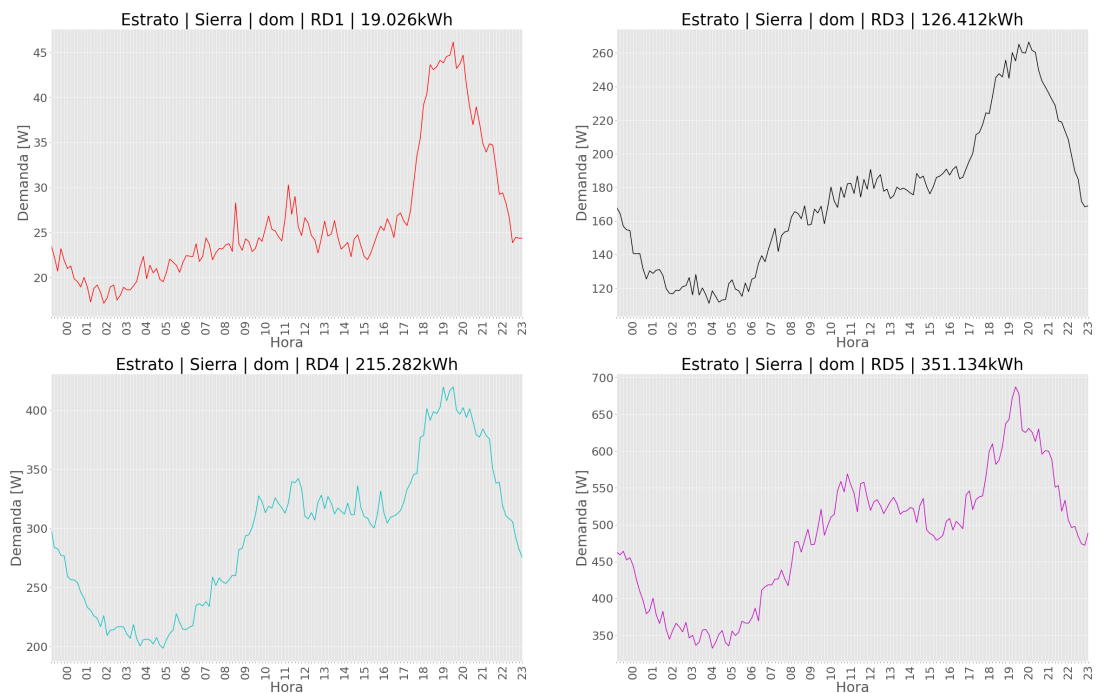
Perfiles de carga de cada Estrato de DBSCAN



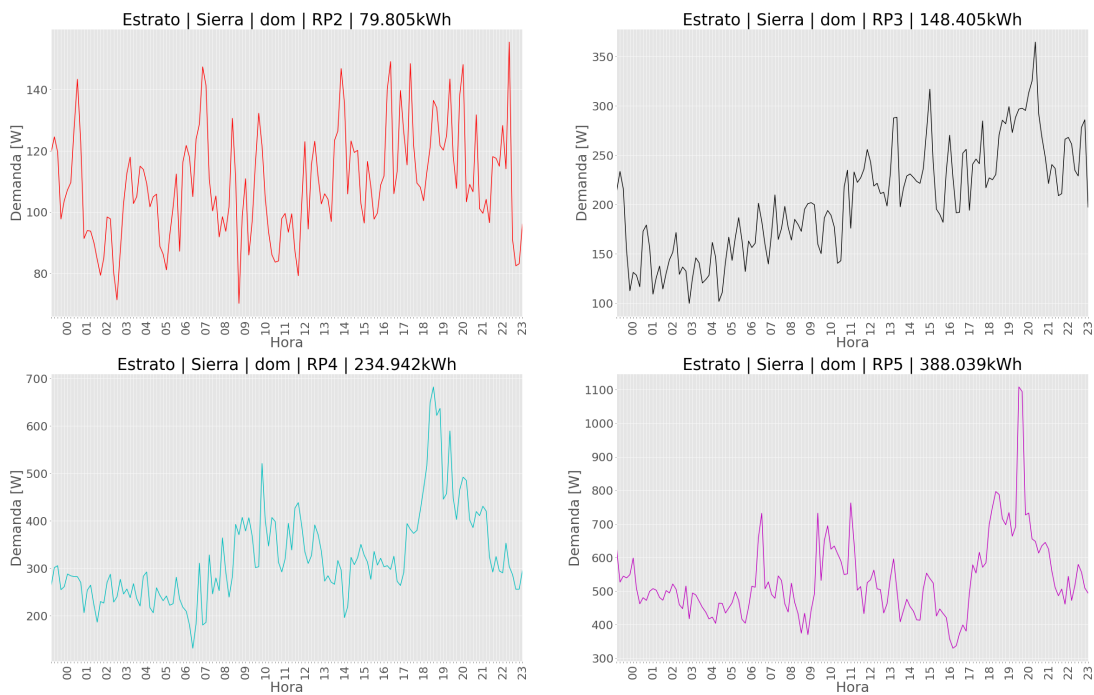
Datos de la Región Sierra en tipo de día dom

Estratos	Obser	EPmax [kW]	PPmax [kW]	PPmin [kW]	Ppro [kW]	E [kWh/mes]	Fc [%]
RD1	699	1.22	0.05	0.02	0.03	19	56.5
RD2	337	2.06	0.27	0.11	0.17	126.4	65
RD3	285	4.58	0.42	0.20	0.30	215.3	70.3
RD4	114	4.10	0.69	0.33	0.48	351.1	70
RP2	13	0.83	0.16	0.07	0.11	79.8	70.3
RP3	9	0.95	0.37	0.10	0.21	148.4	55.8
RP4	8	1.64	0.68	0.13	0.32	234.9	55.8
RP5	9	4.78	1.11	0.33	0.53	388	48
CB1	35	0.58	0.08	0.02	0.04	31.5	55.3
CB2	15	3.22	0.82	0.44	0.59	430.4	72.1
CO1	57	1.89	0.18	0.07	0.11	81.9	62.7
CO2	21	3.93	0.68	0.30	0.45	325.9	65.3
CH3	14	140.08	37.49	23.05	29.79	21747.7	79.5
IH3	63	245.73	45.56	37.82	41.59	30363	91.3
IA1	6	1.98	0.46	0.04	0.14	98.5	29.2
OH3	108.43	61.19	953.65	53.65	56.93	41561.8	93.1

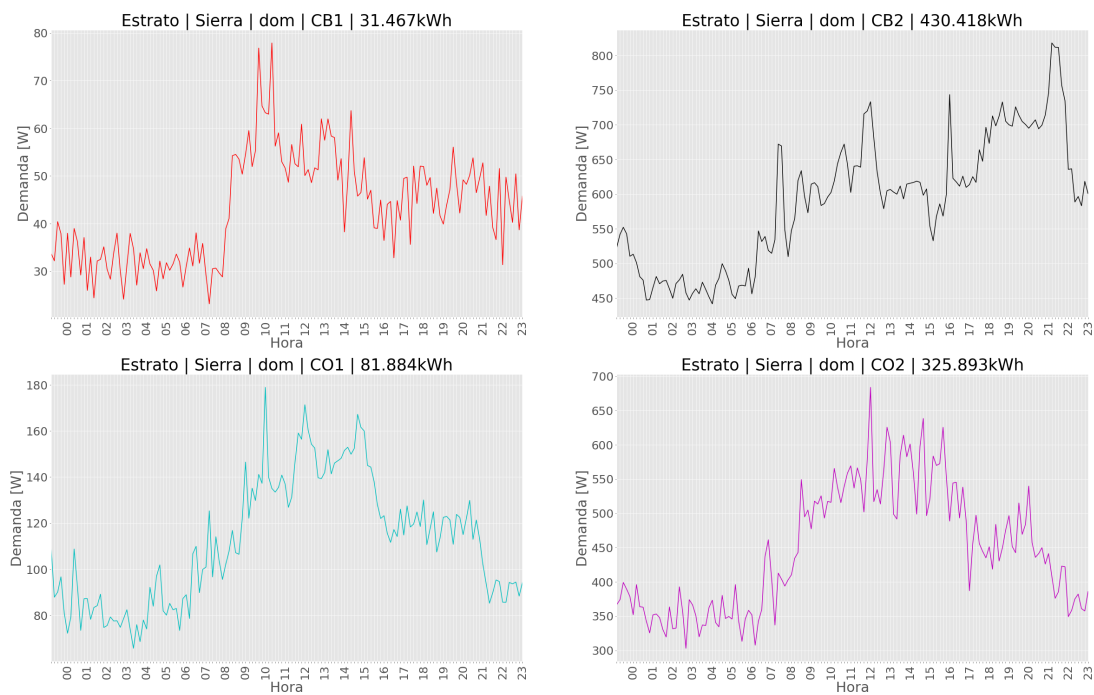
Perfiles de carga de cada Estrato de DBSCAN



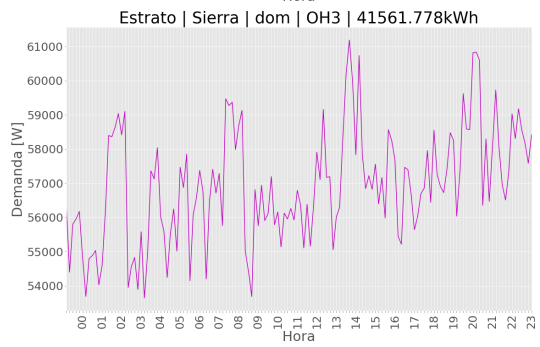
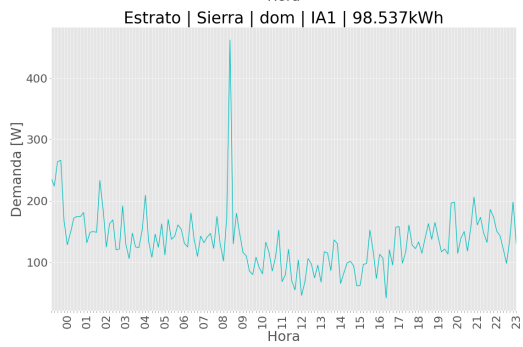
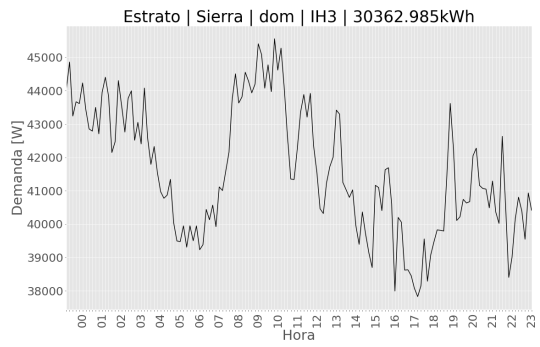
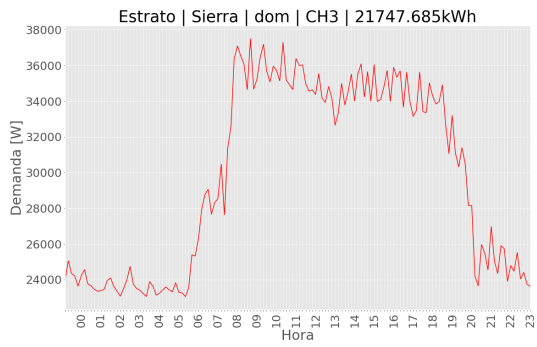
Perfiles de carga de cada Estrato de DBSCAN



Perfiles de carga de cada Estrato de DBSCAN



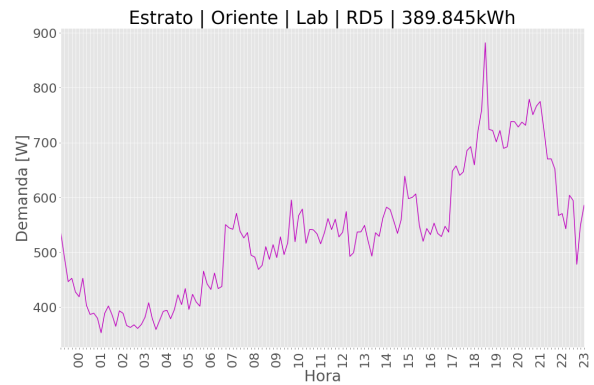
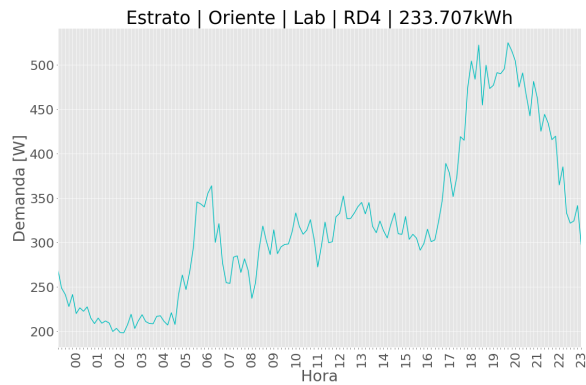
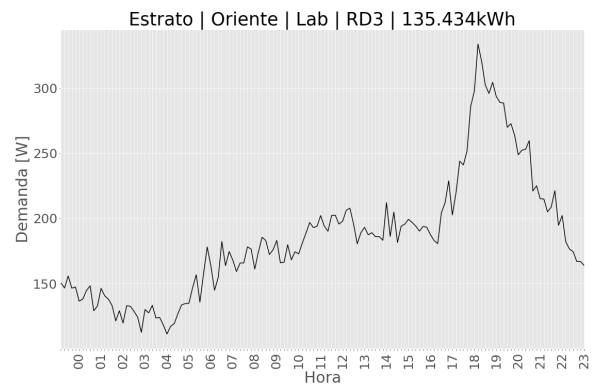
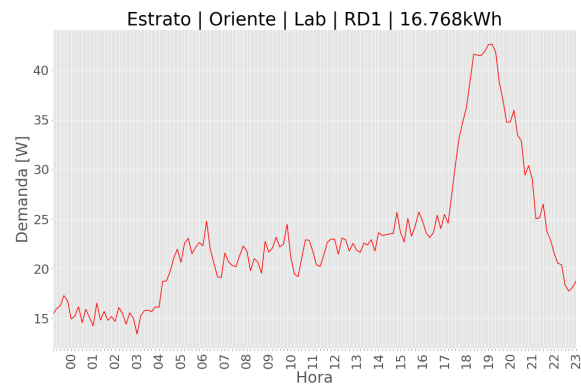
Perfiles de carga de cada Estrato de DBSCAN



Datos de la Región Oriente en tipo de día Lab

Estratos	Obser	EPmax [kW]	PPmax [kW]	PPmin [kW]	Ppro [kW]	E [kWh/mes]	Fc [%]
RD1	460	0.55	0.04	0.01	0.02	16.8	53.9
RD2	193	3.49	0.33	0.11	0.19	135.4	55.6
RD3	96	4.21	0.53	0.20	0.32	233.7	61
RD4	70	6.17	0.88	0.35	0.53	389.8	60.6

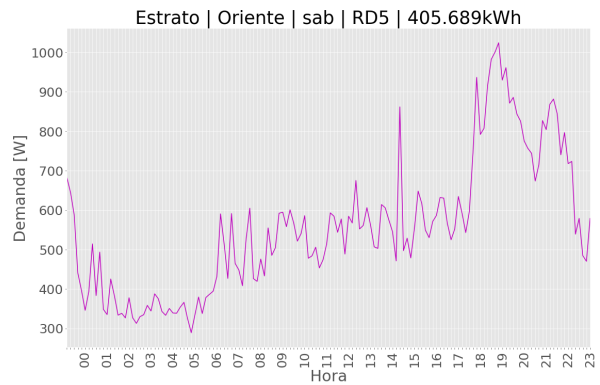
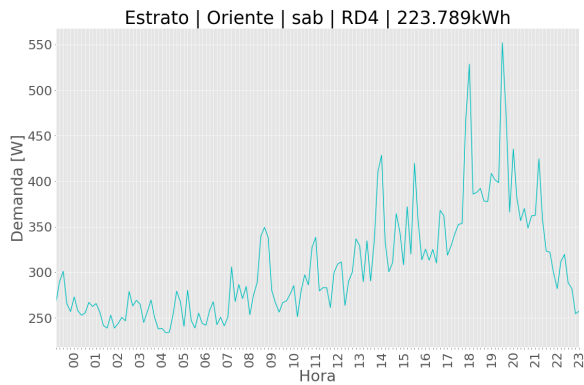
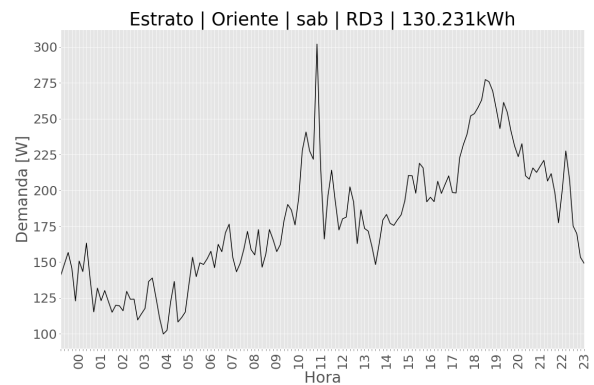
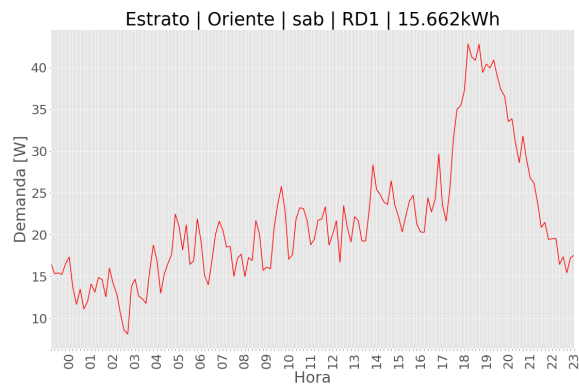
Perfiles de carga de cada Estrato de DBSCAN



Datos de la Región Oriente en tipo de día sab

Estratos	Obser	EPmax [kW]	PPmax [kW]	PPmin [kW]	Ppro [kW]	E [kWh/mes]	Fc [%]
RD1	90	0.29	0.04	0.01	0.02	15.7	50.1
RD2	32	2.79	0.30	0.10	0.18	130.2	59.1
RD3	17	2.39	0.55	0.23	0.31	223.8	55.6
RD4	15	4.49	1.03	0.29	0.56	405.7	54.2

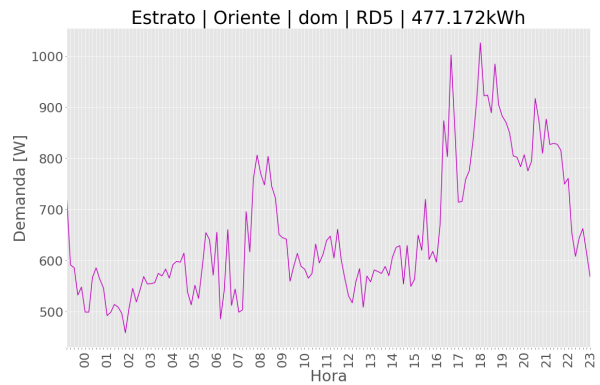
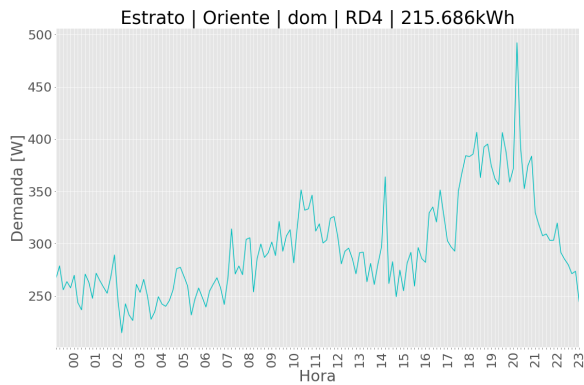
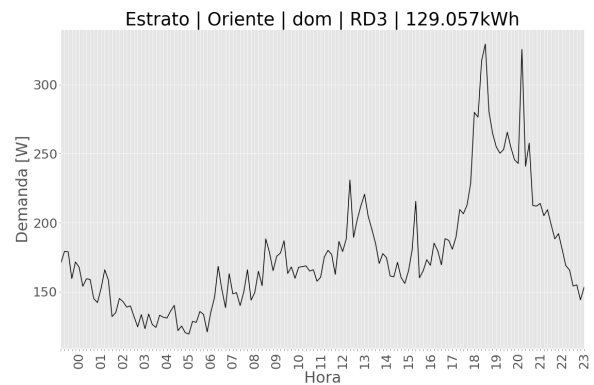
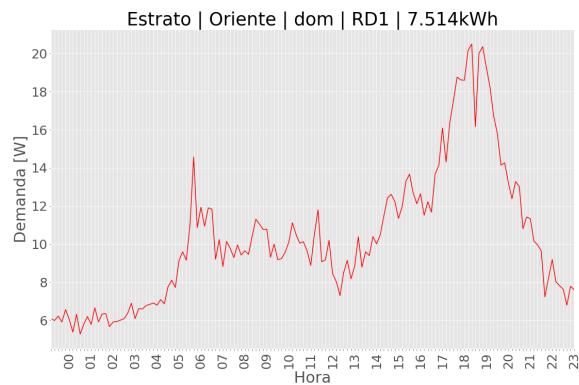
Perfiles de carga de cada Estrato de DBSCAN



Datos de la Región Oriente en tipo de día dom

Estratos	Obser	EPmax [kW]	PPmax [kW]	PPmin [kW]	Ppro [kW]	E [kWh/mes]	Fc [%]
RD1	79	0.35	0.02	0.01	0.01	7.5	50.2
RD2	50	2.39	0.33	0.12	0.18	129.1	53.7
RD3	18	1.90	0.49	0.22	0.30	215.7	60
RD4	19	5.33	1.03	0.46	0.65	477.2	63.7

Perfiles de carga de cada Estrato de DBSCAN

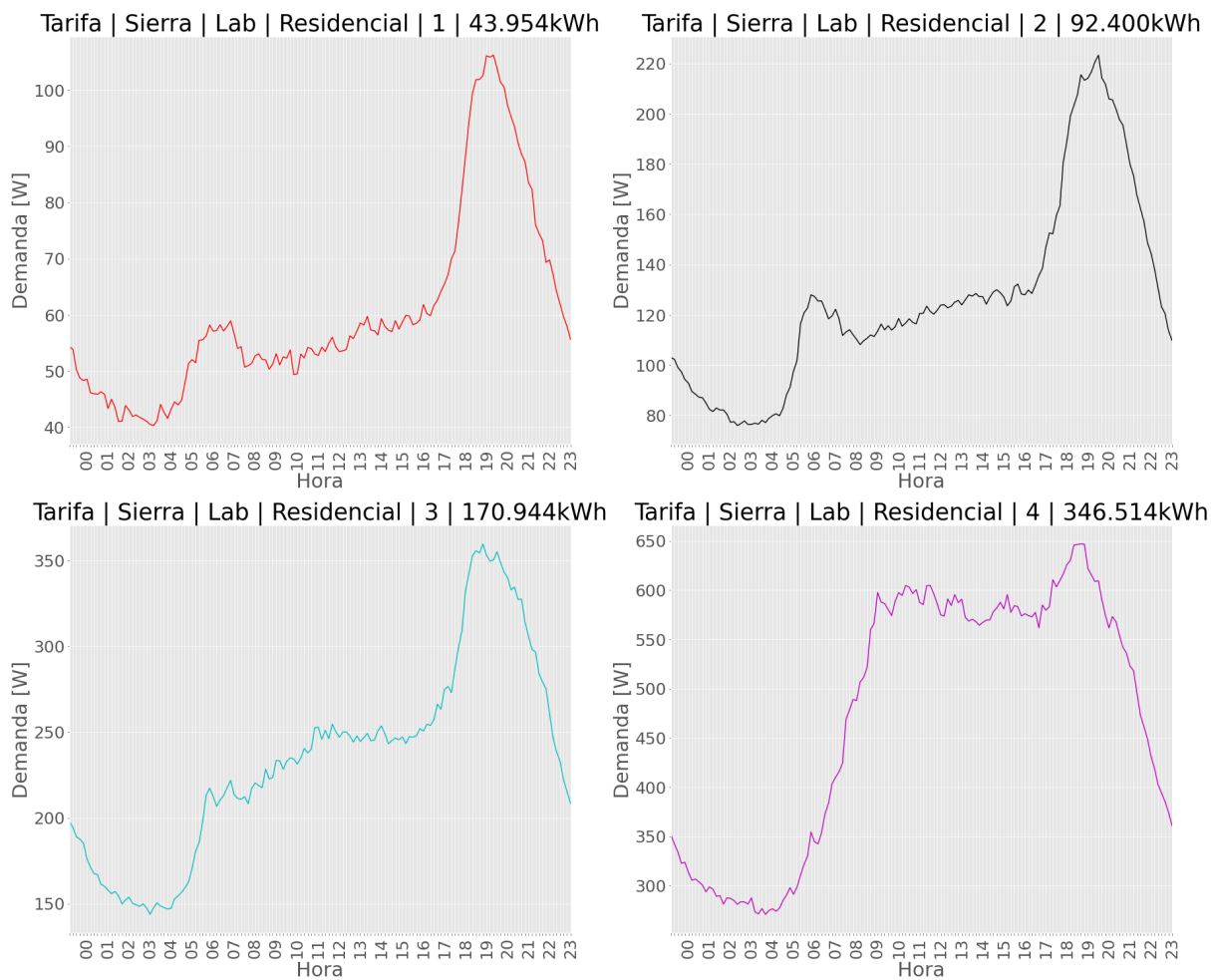


Anexo I.II: Método Mezcla Gaussiana.

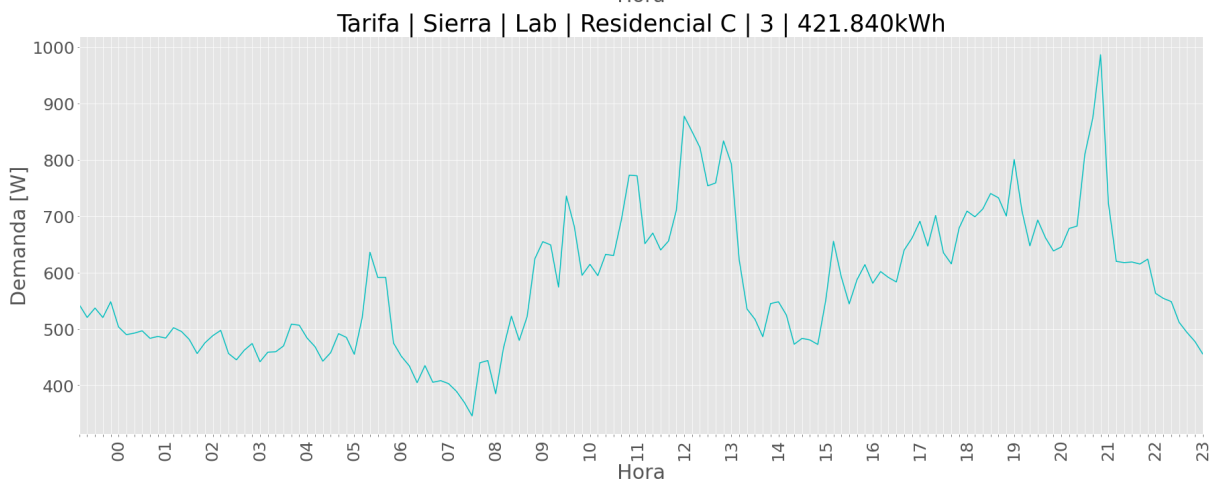
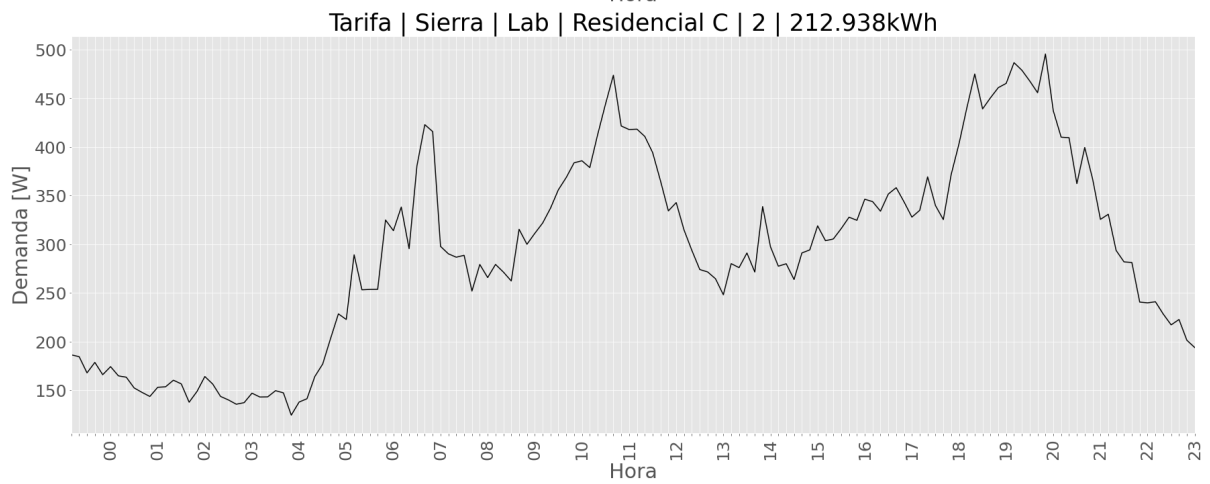
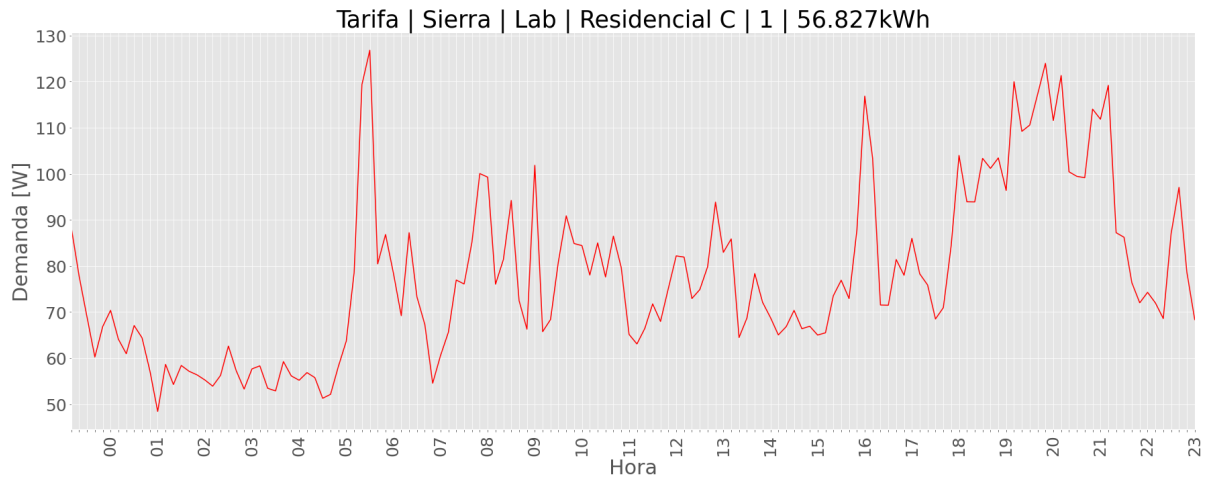
Datos de la Región Sierra en tipo de día Lab

Estratos	Obser	EPmax [kW]	PPmax [kW]	PPmin [kW]	Ppro [kW]	E [kWh/mes]	Fc [%]
Residencial 1	1919	1.16	0.11	0.04	0.06	44	56.7
Residencial 2	2544	2.72	0.22	0.08	0.13	92.4	56.7
Residencial 3	2052	2.88	0.36	0.14	0.23	170.9	65.2
Residencial 4	775	2.91	0.65	0.27	0.48	346.5	73.4
Resi Coc 1	118	3.79	0.13	0.05	0.08	56.8	61.4
Resi Coc 2	134	4.38	0.50	0.12	0.29	212.9	58.9
Resi Coc 3	36	4.63	0.99	0.35	0.58	421.8	58.6
Comercial 1	250	7.57	1.44	0.56	1.00	727.6	69.4
Comercial 2	67	218.88	45.50	17.76	31.46	22965.9	69.2
Comercial 3	10	241.50	210.94	53.95	127.39	92992.2	60.4
Comercial 4	4	316.43	300.84	66.29	174.37	127292.7	58
Industrial 1	152	1096.59	21.94	1.439	6.69	4886.1	30.5
Industrial 2	201	626.77	127.51	55.55	88.25	64423.3	69.2
Industrial 3	55	7561.84	2582.18	2131.64	2404.89	1755569.1	93.1
Otros 1	85	12.47	2.00	0.66	1.19	869.4	59.6
Otros 2	52	74.86	23.10	12.14	17.86	13036.7	77.3
Otros 3	33	287.41	148.078	86.40	114.04	83245.3	77

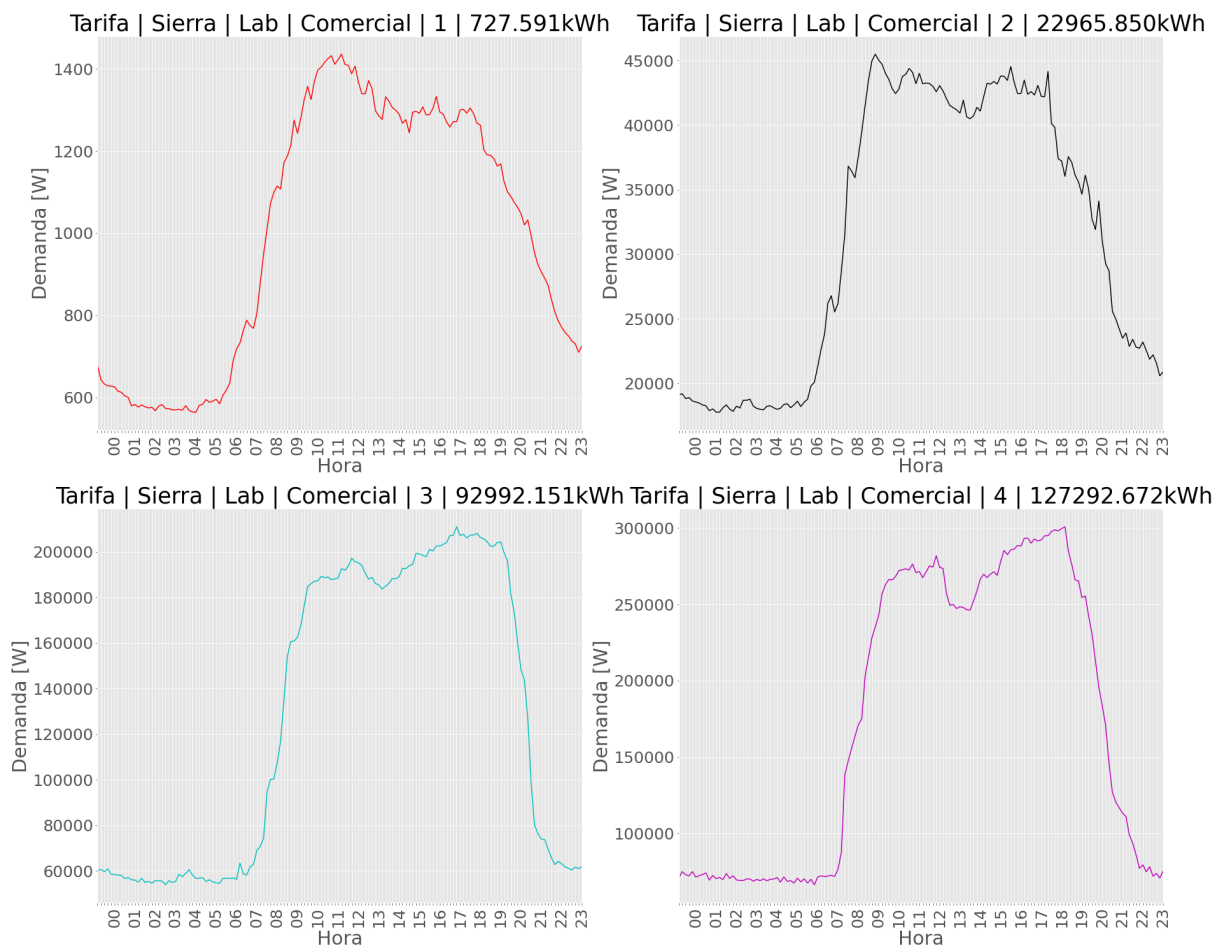
Perfiles de carga de cada Estrato de Mezcla Gaussiana



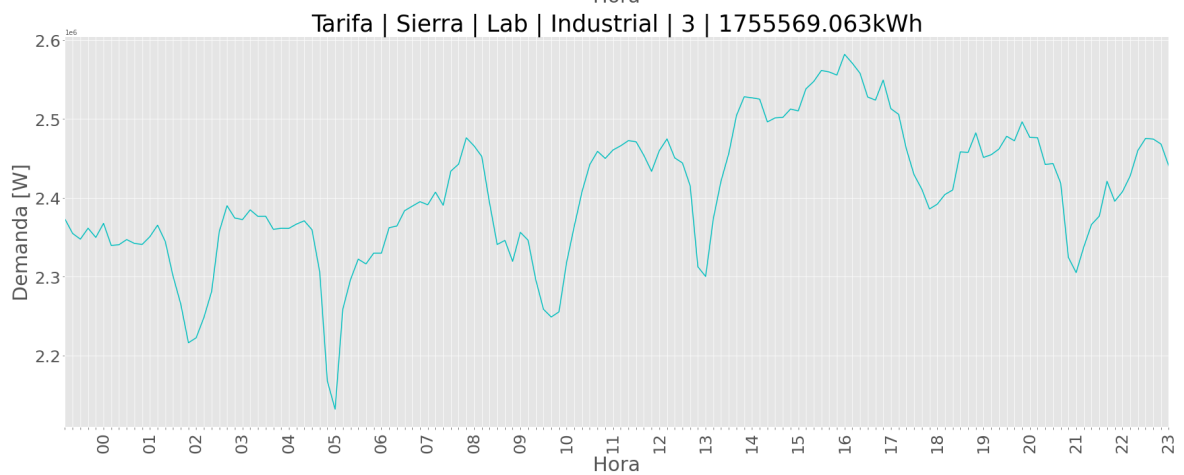
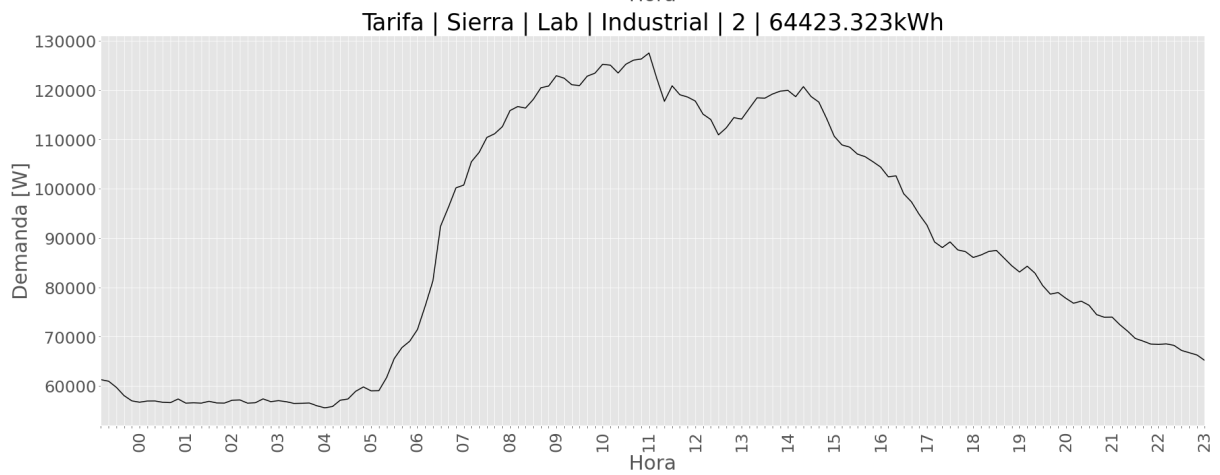
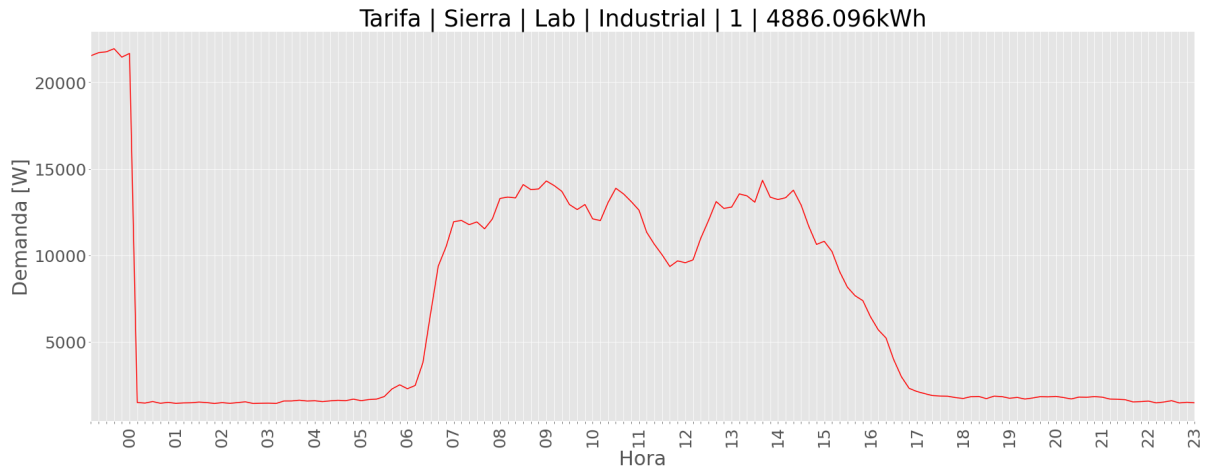
Perfiles de carga de cada Estrato de Mezcla Gaussiana



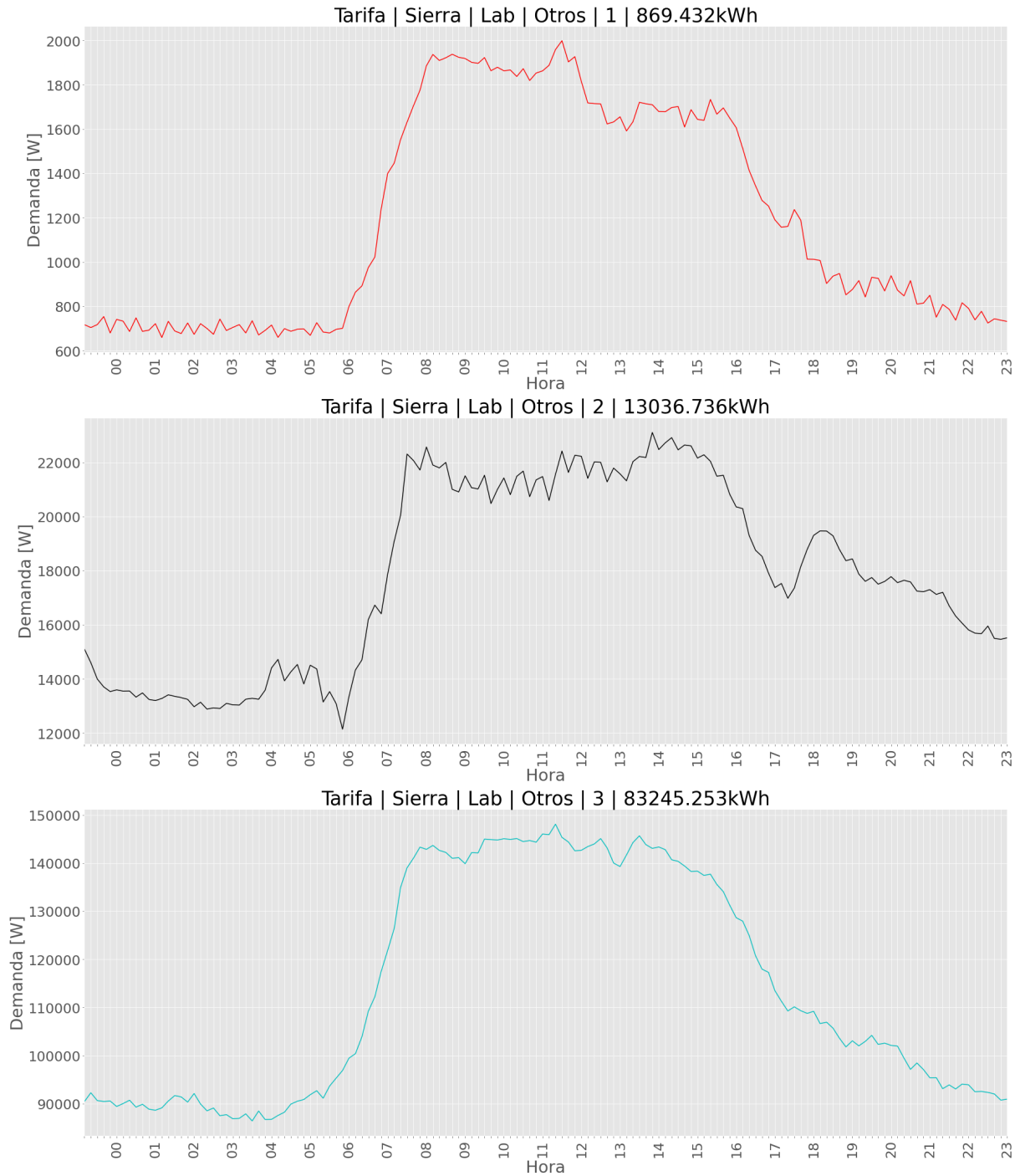
Perfiles de carga de cada Estrato de Mezcla Gaussiana



Perfiles de carga de cada Estrato de Mezcla Gaussiana



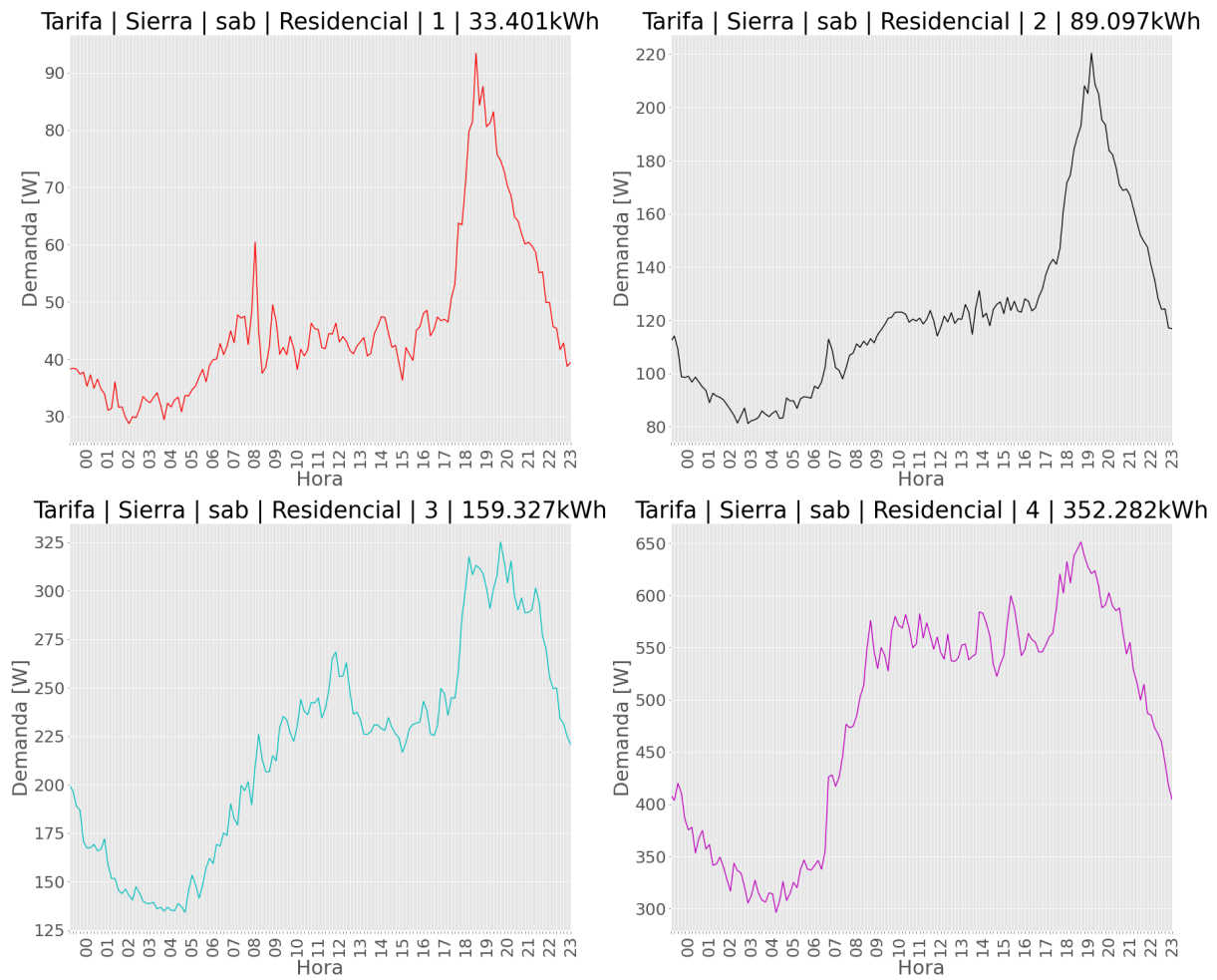
Perfiles de carga de cada Estrato de Mezcla Gaussiana



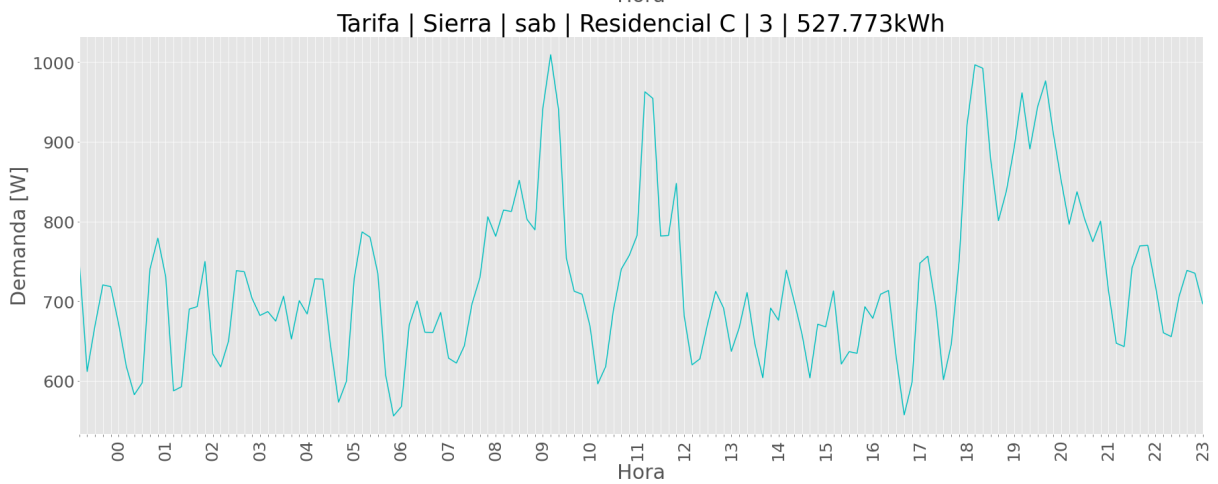
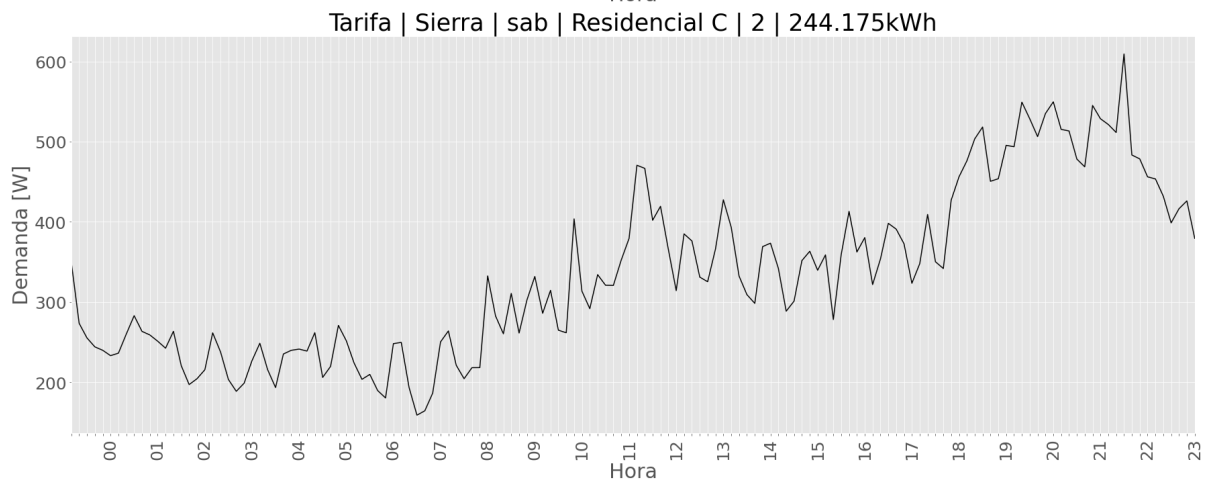
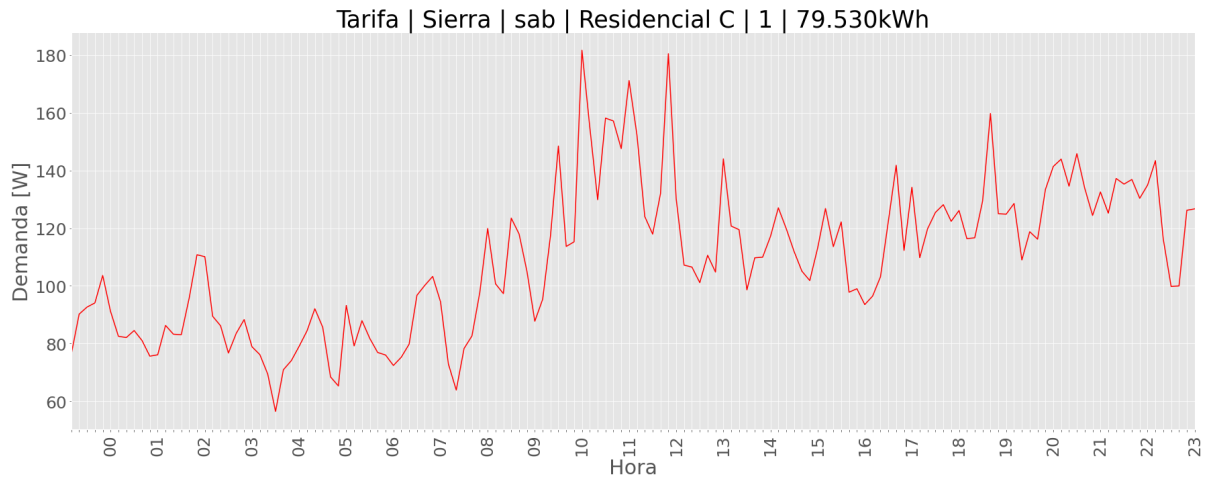
Datos de la Región Sierra en tipo de día sab

Estratos	Obser	EPmax [kW]	PPmax [kW]	PPmin [kW]	Ppro [kW]	E [kWh/mes]	Fc [%]
Residencial 1	423	2.77	0.09	0.03	0.05	33.4	49
Residencial 2	545	2.59	0.22	0.08	0.12	89.1	55.4
Residencial 3	464	2.65	0.33	0.13	0.22	159.3	67.1
Residencial 4	149	2.58	0.65	0.30	0.48	352.3	74.1
Resi Coc 1	29	1.61	0.18	0.06	0.11	79.5	59.9
Resi Coc 2	13	1.80	0.61	0.16	0.33	244.2	54.9
Resi Coc 3	2	1.38	1.01	0.56	0.11	527.8	71.6
Comercial 1	23	4.43	1.01	0.09	0.41	298.3	40.5
Comercial 2	21	3.50	0.90	0.44	0.64	465.3	71.2
Comercial 3	7	3.74	1.99	0.63	1.13	820.9	56.6
Comercial 4	5	2.82	2.12	0.73	1.49	1085.1	70.2
Industrial 1	73	333.80	49.64	28.34	36.43	26596.3	73.4
Industrial 2	9	5293.01	1649.35	1241.91	1501.10	1095805.5	91
Industrial 3	2	7080.13	6464.70	5202.34	6035.93	4406225.2	93.4
Otros 1	29	34.42	6.79	4.87	5.77	4208.9	84.9
Otros 2	6	176.95	93.64	71.85	80.15	58510.4	85.6
Otros 3	1	225.25	225.25	132.97	181.09	132194.5	80.4

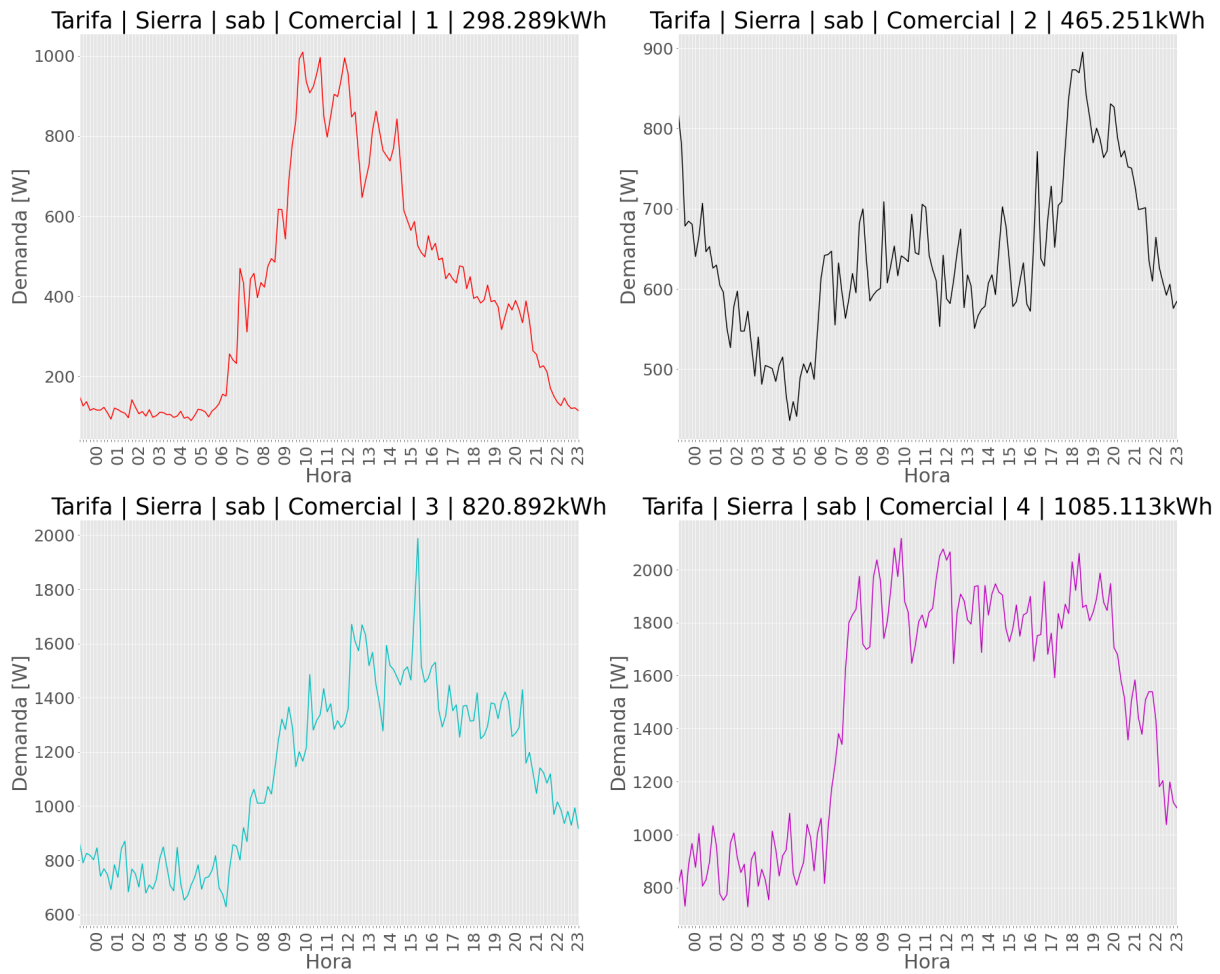
Perfiles de carga de cada Estrato de Mezcla Gaussiana



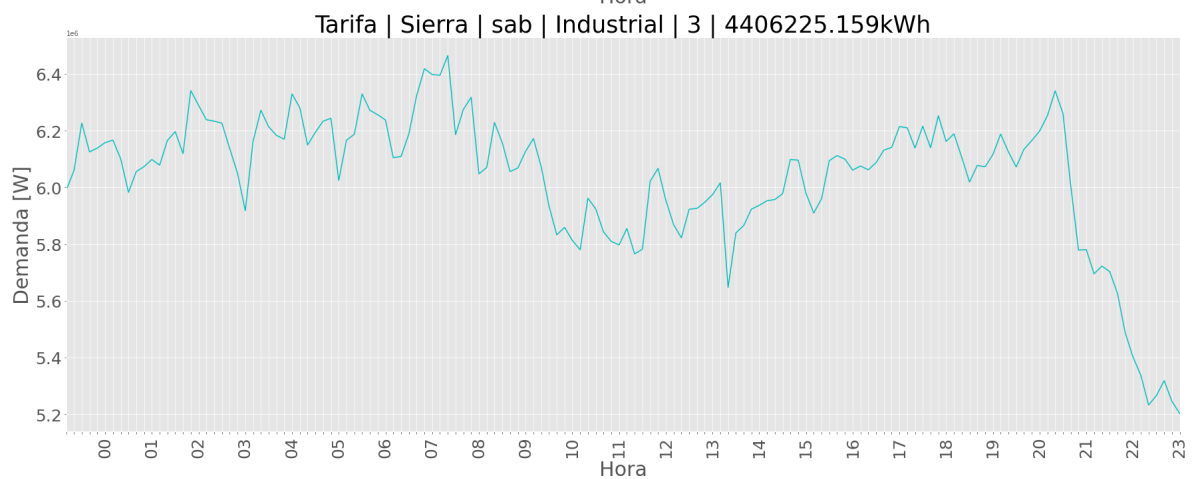
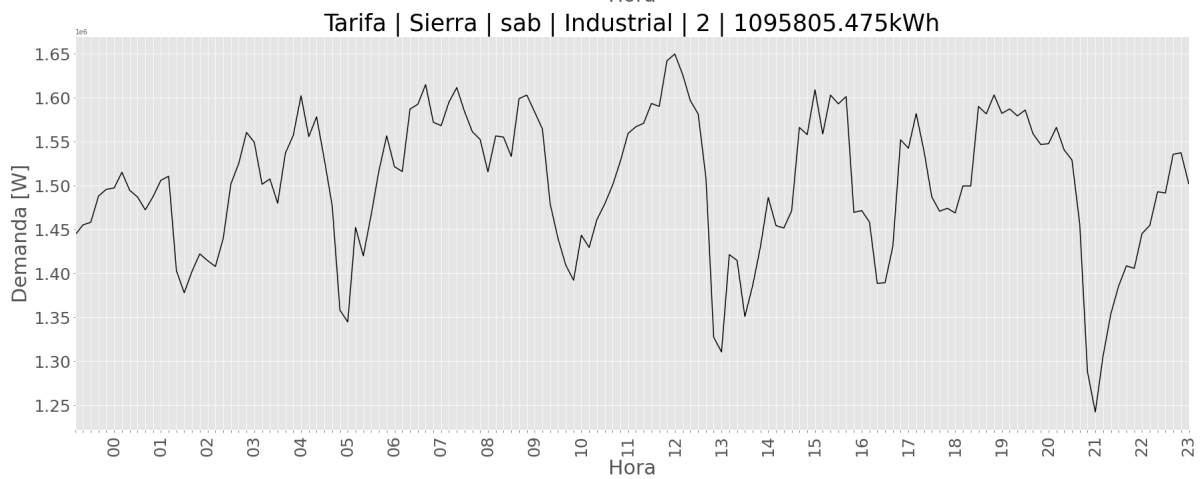
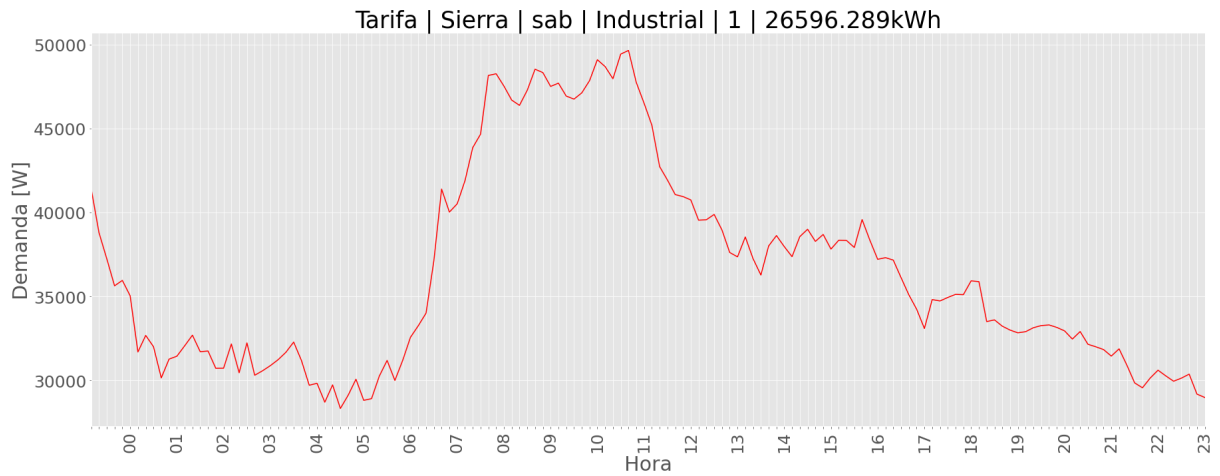
Perfiles de carga de cada Estrato de Mezcla Gaussiana



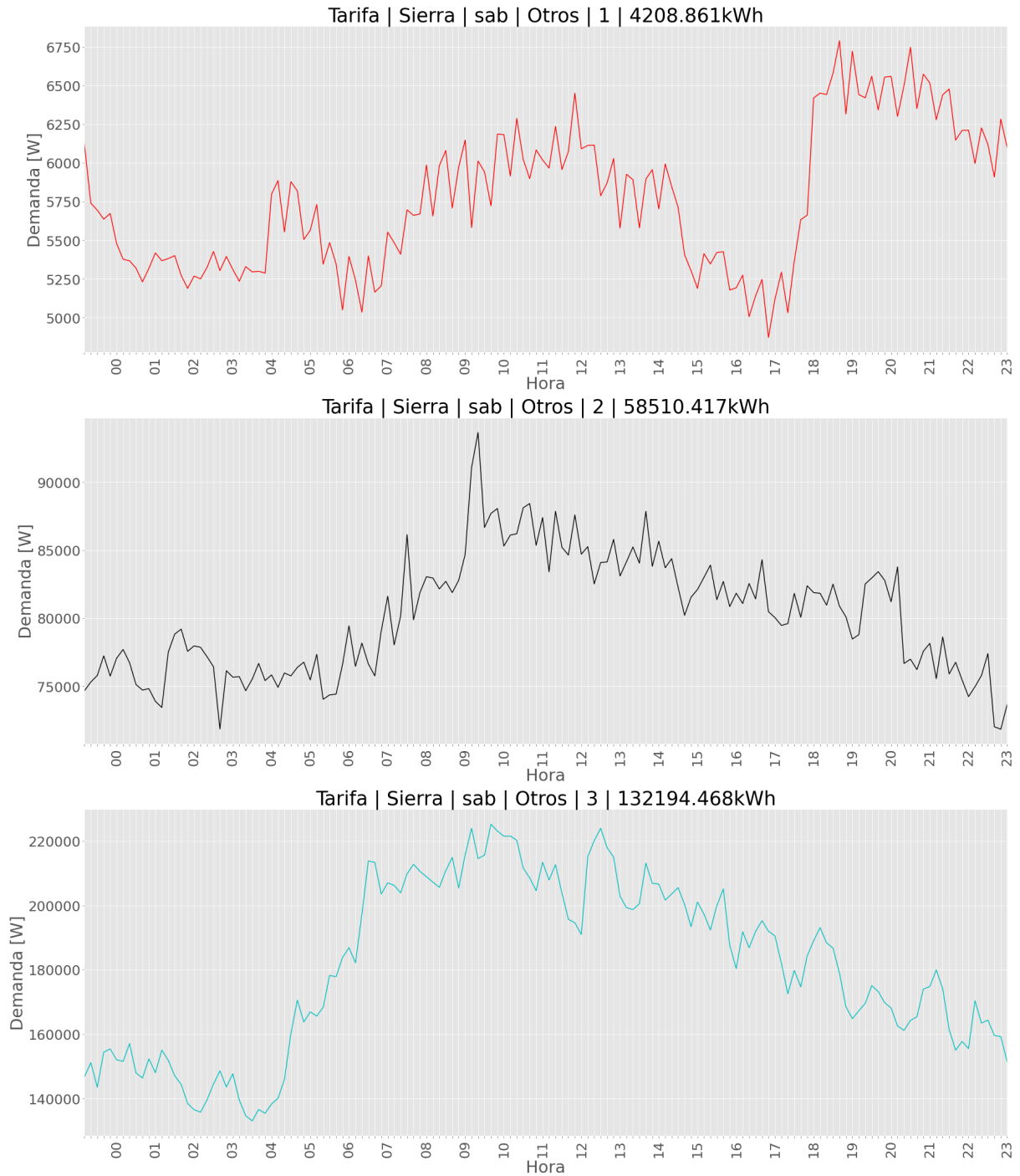
Perfiles de carga de cada Estrato de Mezcla Gaussiana



Perfiles de carga de cada Estrato de Mezcla Gaussiana



Perfiles de carga de cada Estrato de Mezcla Gaussiana

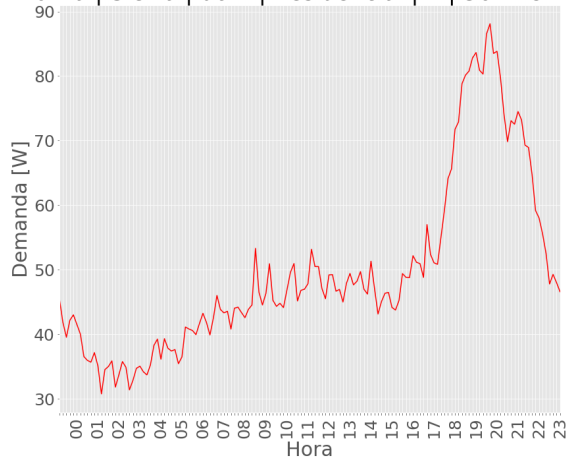


Datos de la Región Sierra en tipo de día dom

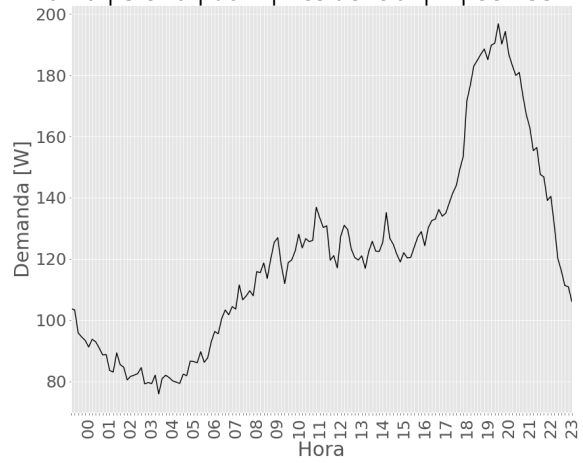
Estratos	Obser	EPmax [kW]	PPmax [kW]	PPmin [kW]	Ppro [kW]	E [kWh/mes]	Fc [%]
Residencial 1	475	2.37	0.09	0.03	0.05	36.3	56.4
Residencial 2	635	2.53	0.20	0.08	0.12	88.3	61.4
Residencial 3	499	2.57	0.37	0.15	0.23	167.2	61.3
Residencial 4	203	2.59	0.60	0.29	0.45	330.8	76
Resi Coc 1	31	1.20	0.18	0.06	0.09	66.2	51
Resi Coc 2	13	1.64	0.54	0.16	0.29	208.5	53.4
Resi Coc 3	2	0.98	0.88	0.21	0.65	475.7	73.9
Comercial 1	32	4.04	0.97	0.46	0.69	500.6	70.5
Comercial 2	12	4.03	2.05	0.87	1.44	1047.3	70.1
Comercial 3	5	4.73	3.30	2.16	2.57	1879.3	78
Comercial 4	1	5.64	5.64	1.62	3.71	2710.8	65.9
Industrial 1	86	245.73	33.56	27.84	30.61	22345.5	91.2
Industrial 2	7	2879.10	1390.75	1019.62	1285.00	938047.2	92.4
Industrial 3	3	5894.13	5355.45	4202.73	4839.84	3533079.7	90.4
Otros 1	37	31.34	6.86	5.03	5.78	4219.8	84.3
Otros 2	8	155.82	85.33	69.05	77.27	56405.6	90.6
Otros 3	1	211.17	211.17	133.62	174.26	127207.3	82.5

Perfiles de carga de cada Estrato de Mezcla Gaussiana

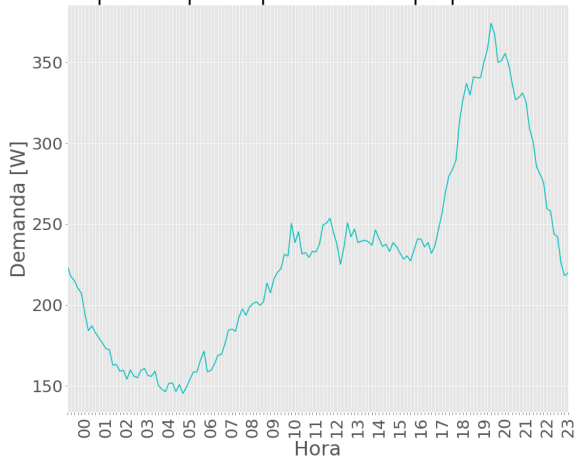
Tarifa | Sierra | dom | Residencial | 1 | 36.278kWh



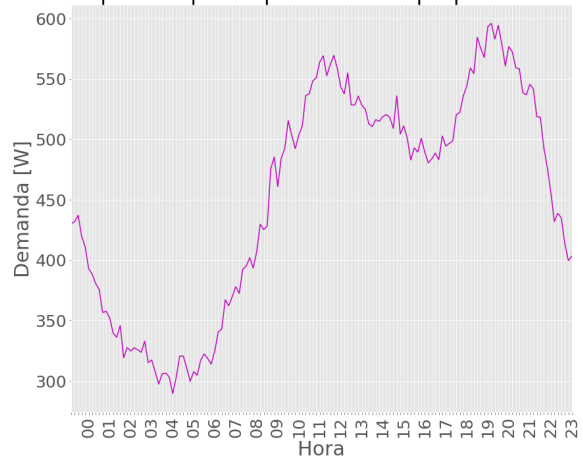
Tarifa | Sierra | dom | Residencial | 2 | 88.255kWh



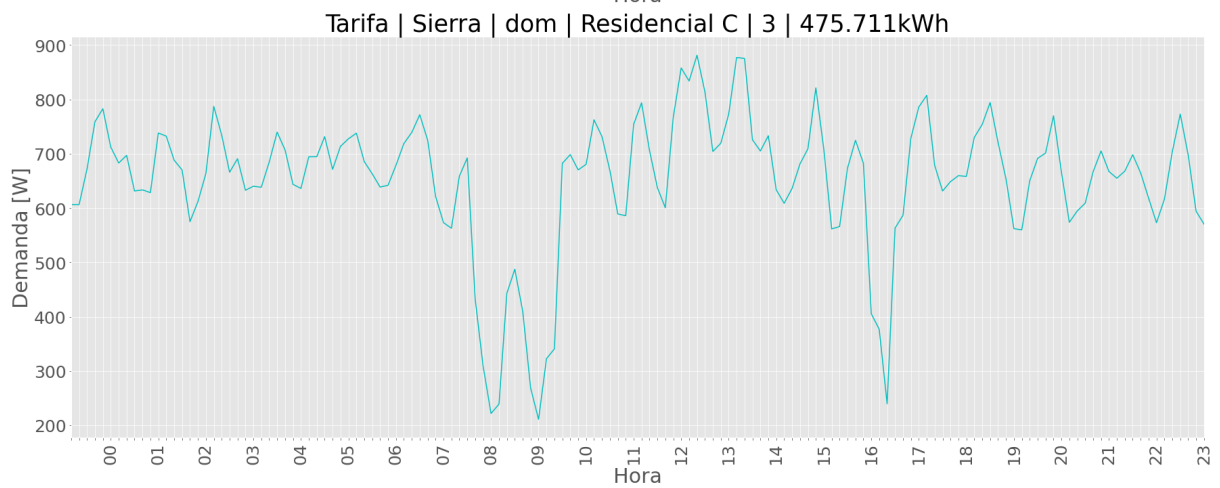
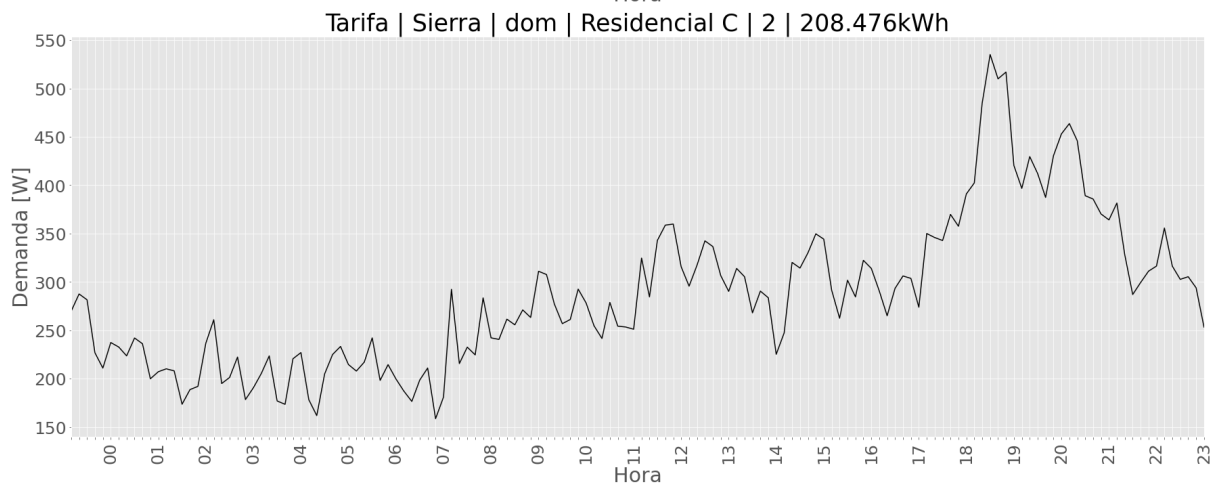
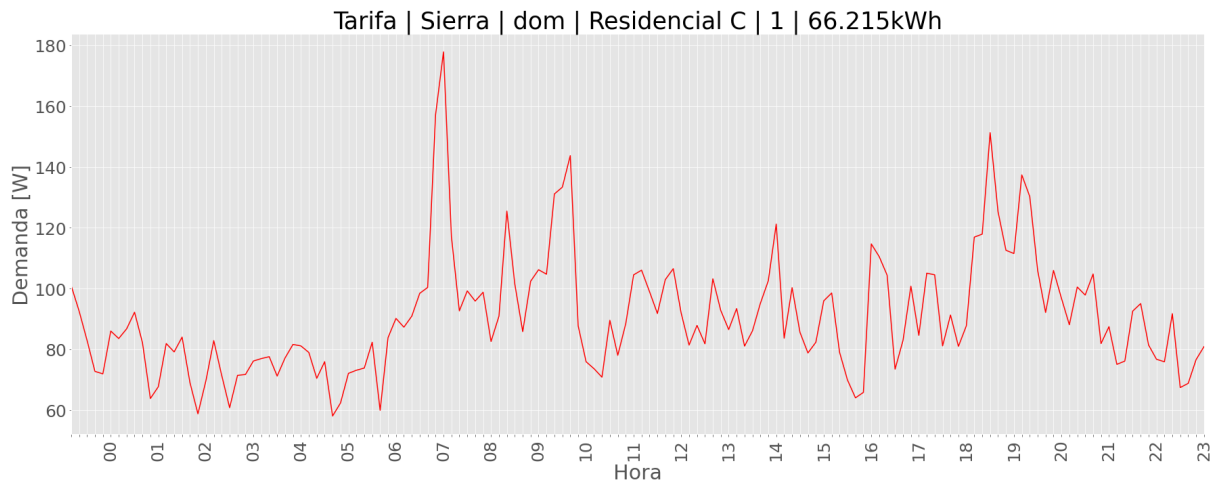
Tarifa | Sierra | dom | Residencial | 3 | 167.249kWh



Tarifa | Sierra | dom | Residencial | 4 | 330.752kWh

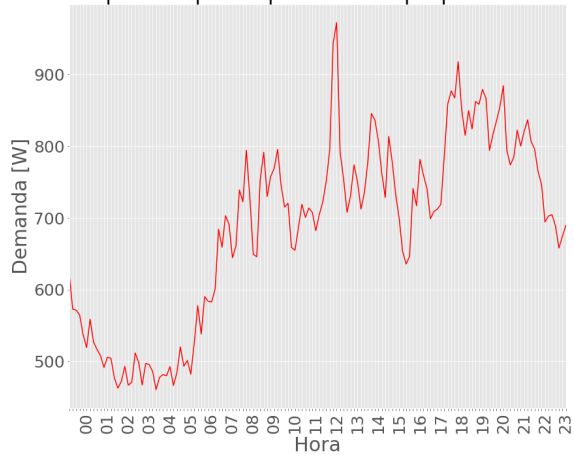


Perfiles de carga de cada Estrato de Mezcla Gaussiana

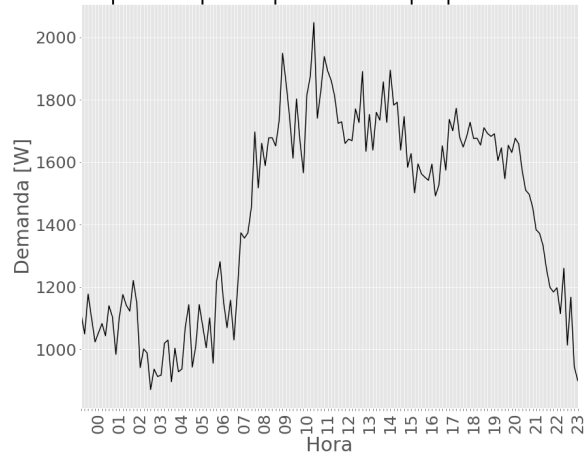


Perfiles de carga de cada Estrato de Mezcla Gaussiana

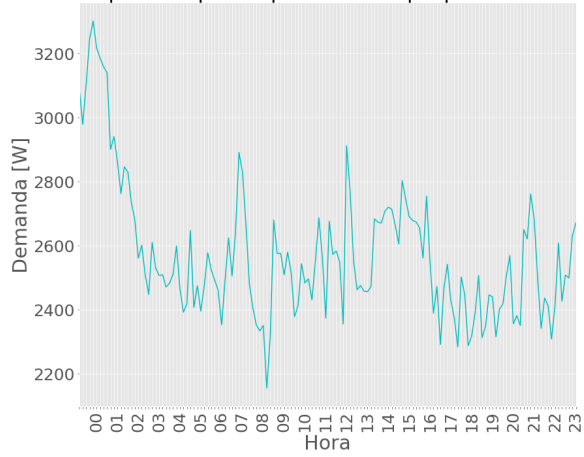
Tarifa | Sierra | dom | Comercial | 1 | 500.642kWh



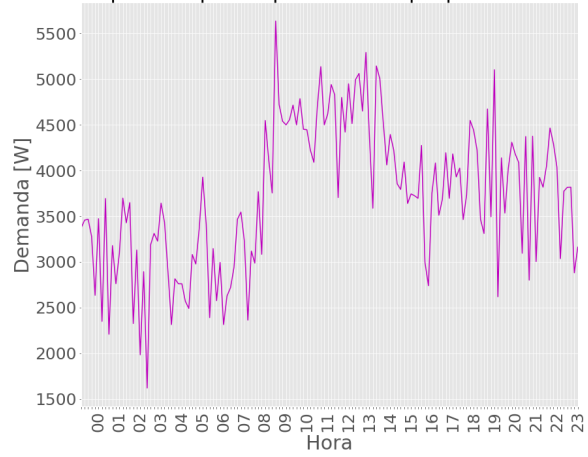
Tarifa | Sierra | dom | Comercial | 2 | 1047.263kWh



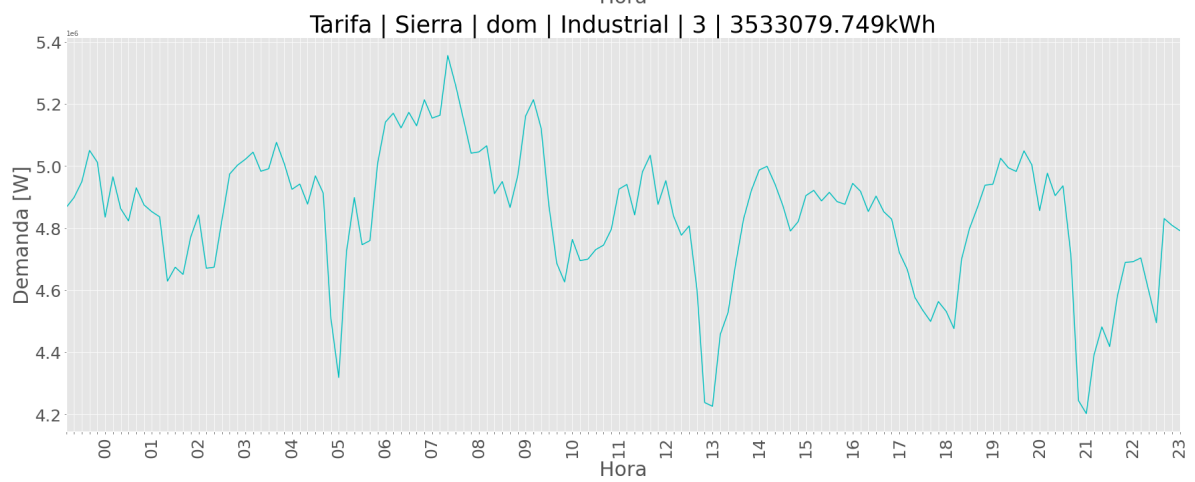
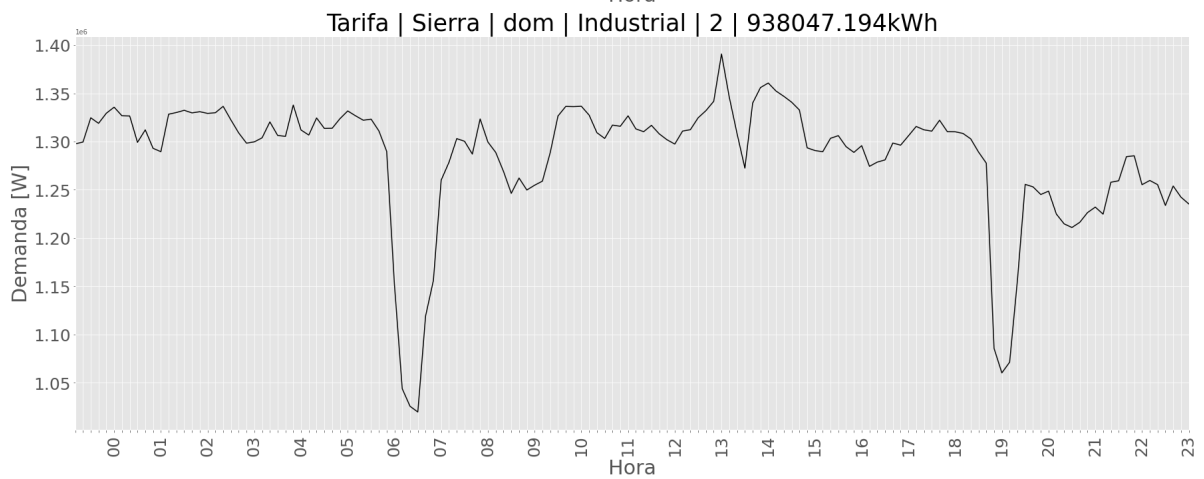
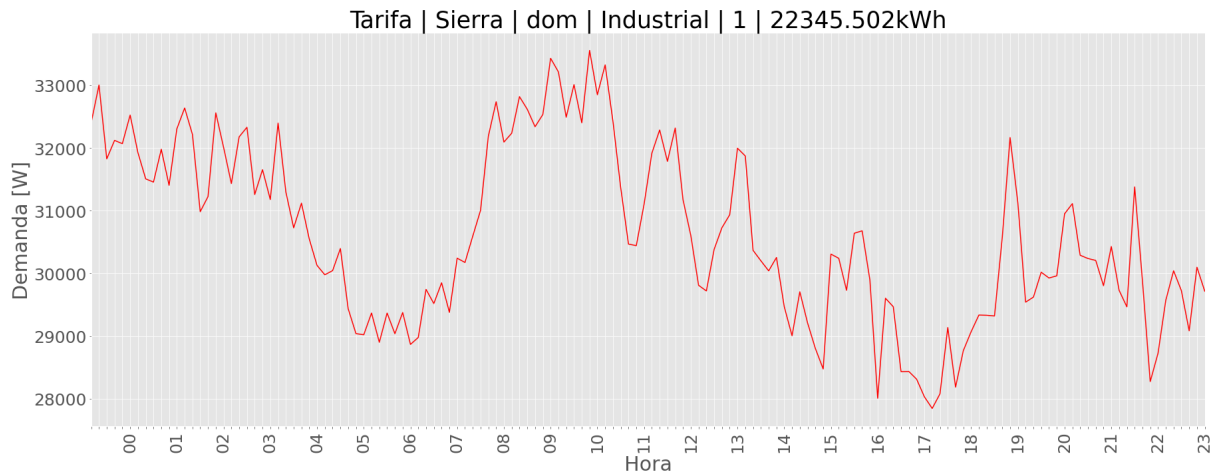
Tarifa | Sierra | dom | Comercial | 3 | 1879.340kWh



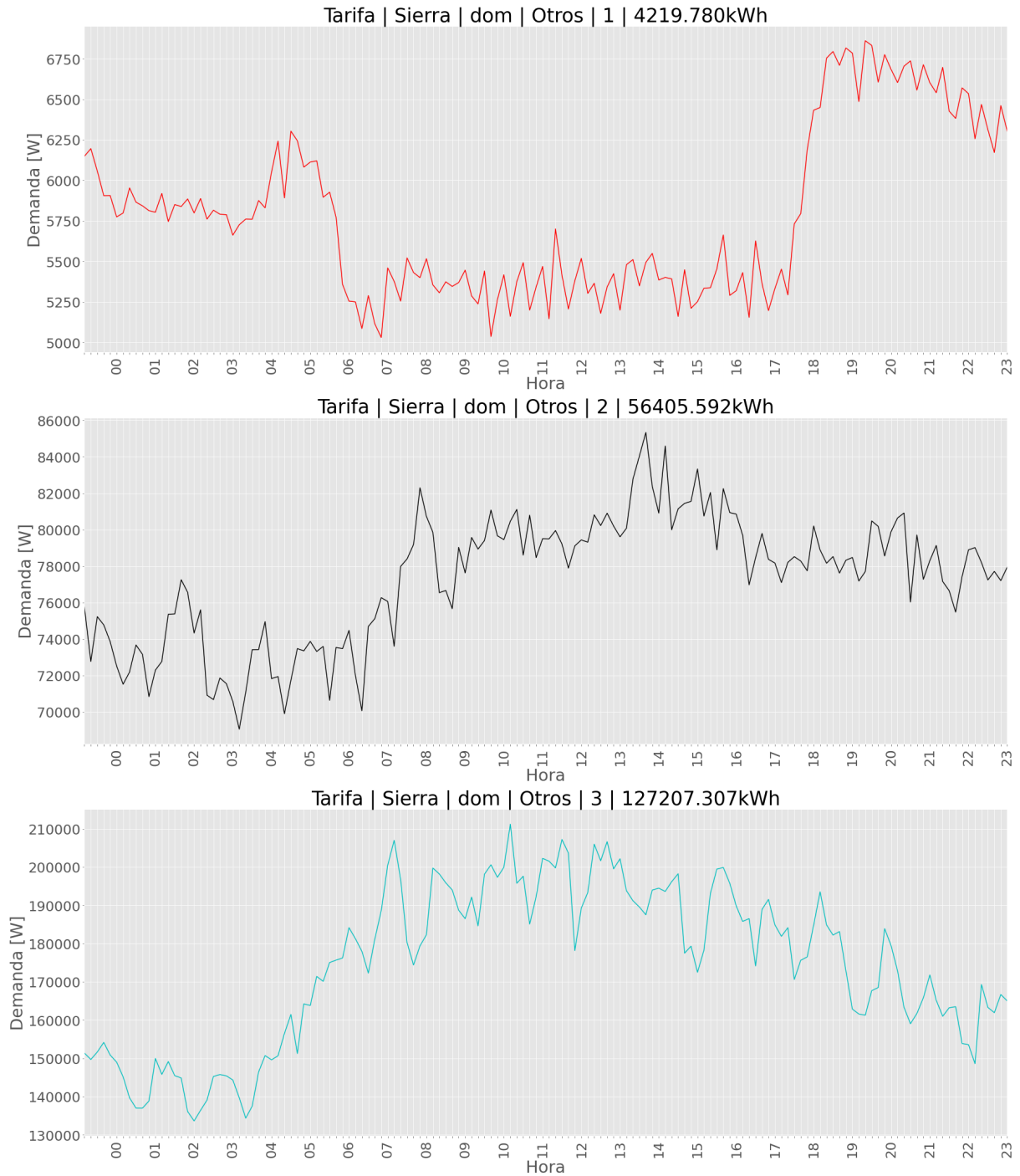
Tarifa | Sierra | dom | Comercial | 4 | 2710.804kWh



Perfiles de carga de cada Estrato de Mezcla Gaussiana



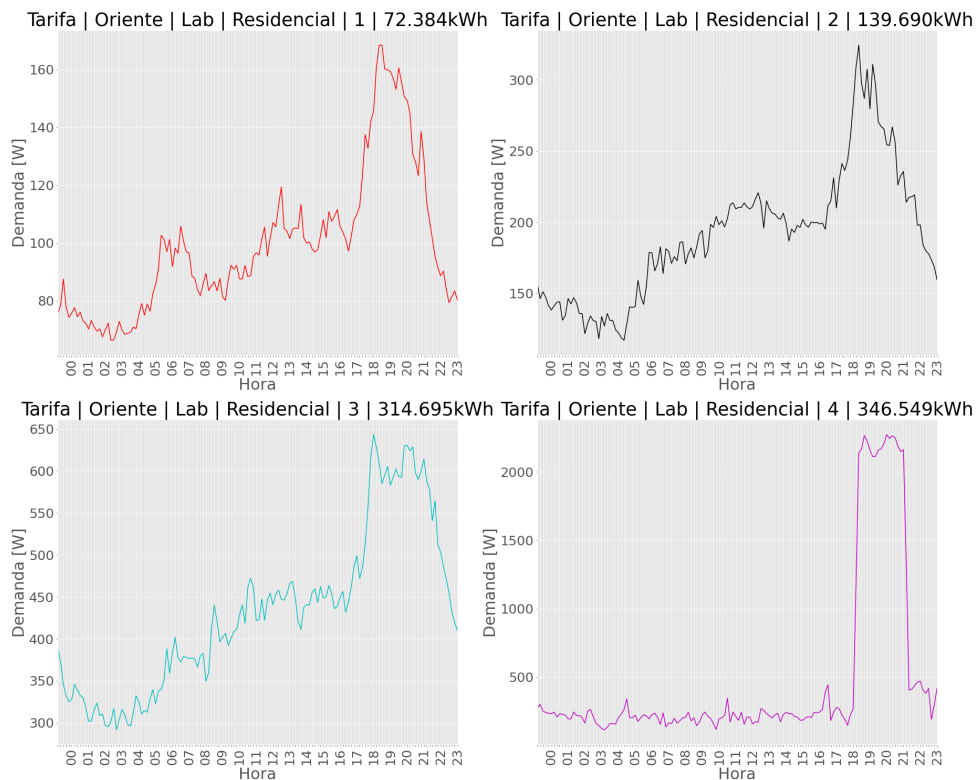
Perfiles de carga de cada Estrato de Mezcla Gaussiana



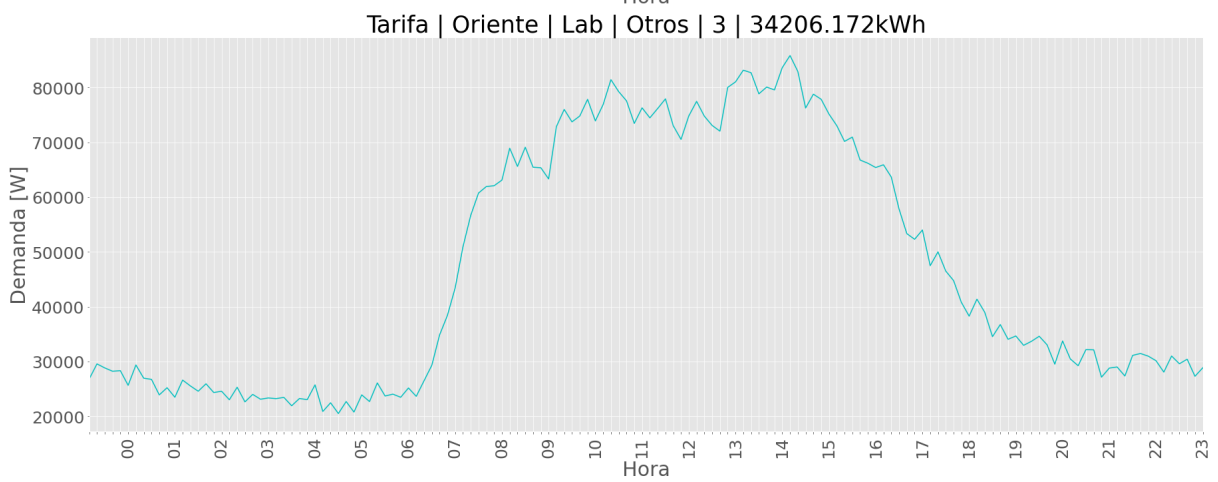
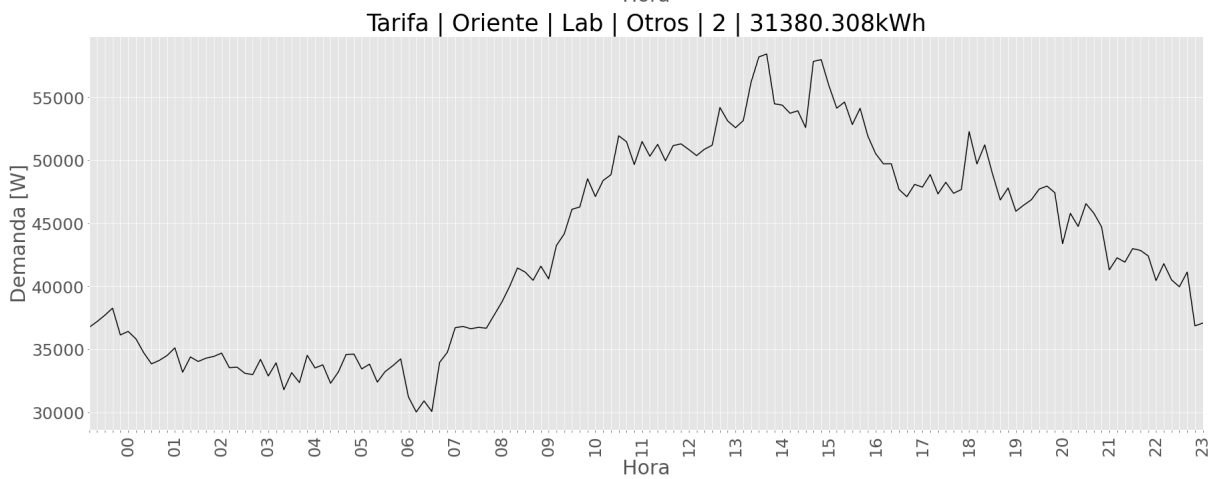
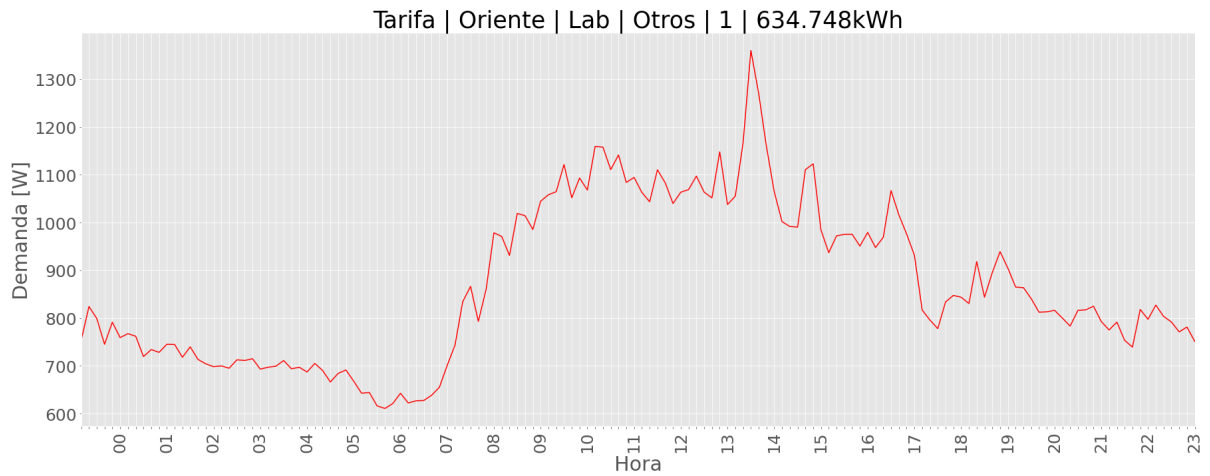
Datos de la Región Oriente en tipo de día Lab

Estratos	Obser	EPmax [kW]	PPmax [kW]	PPmin [kW]	Ppro [kW]	E [kWh/mes]	Fc [%]
Residencial 1	289	2.67	0.17	0.07	0.10	72.4	58.9
Residencial 2	229	2.67	0.32	0.12	0.19	139.7	59
Residencial 3	83	2.65	0.64	0.29	0.43	314.7	67
Residencial 4	3	2.49	2.27	0.12	0.48	346.5	20.9
Otros 1	37	10.88	1.36	0.61	0.87	634.7	63.9
Otros 2	8	78.62	58.42	30.01	42.99	31380.3	73.6
Otros 3	6	124.30	85.81	20.52	46.86	34206.2	54.6

Perfiles de carga de cada Estrato de Mezcla Gaussiana



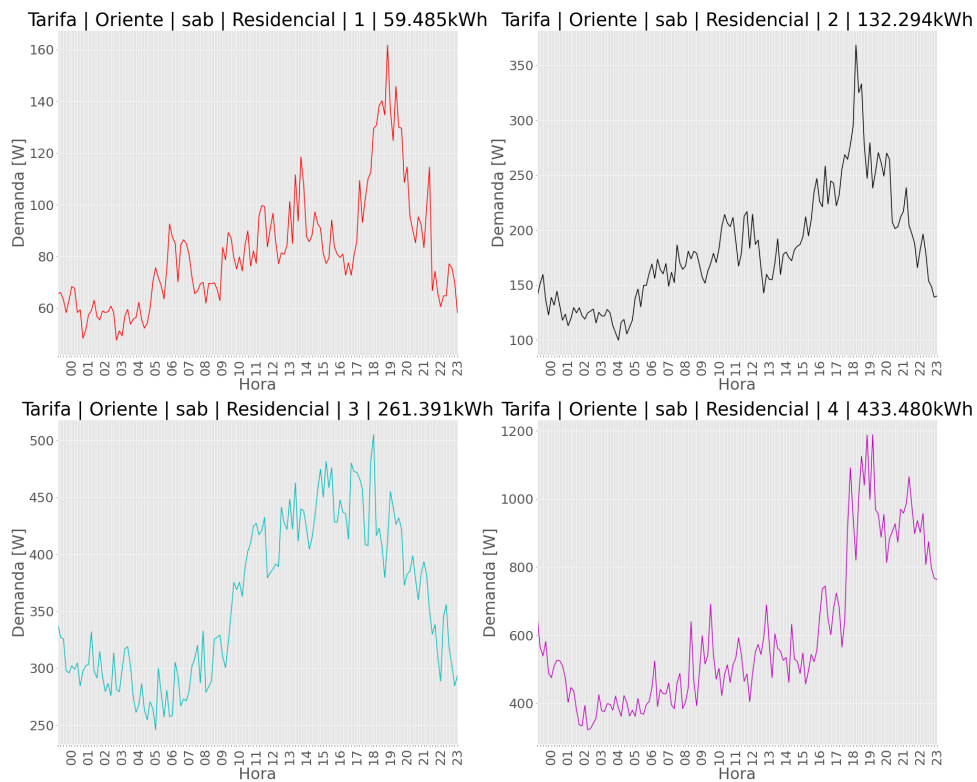
Perfiles de carga de cada Estrato de Mezcla Gaussiana



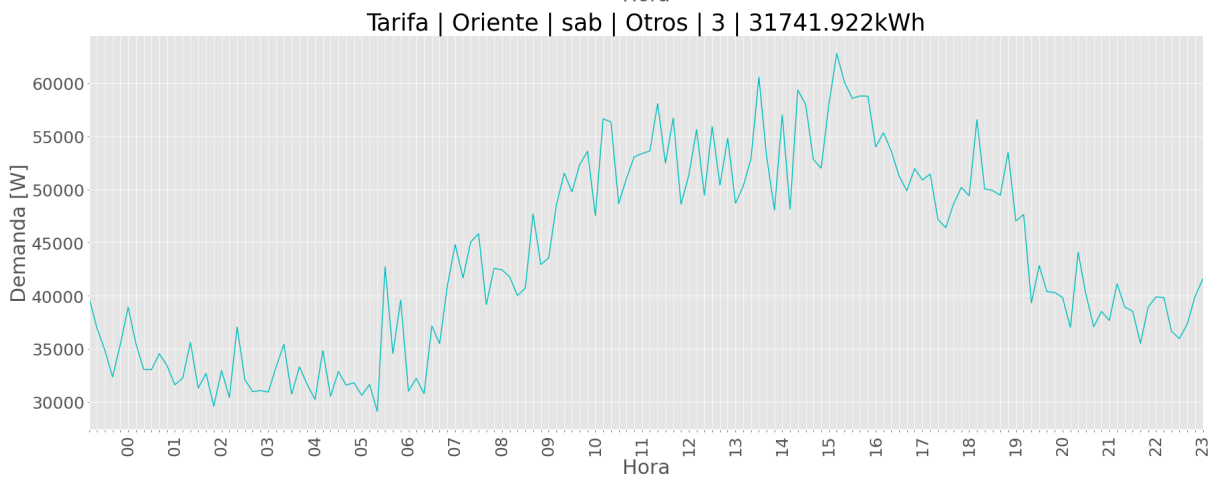
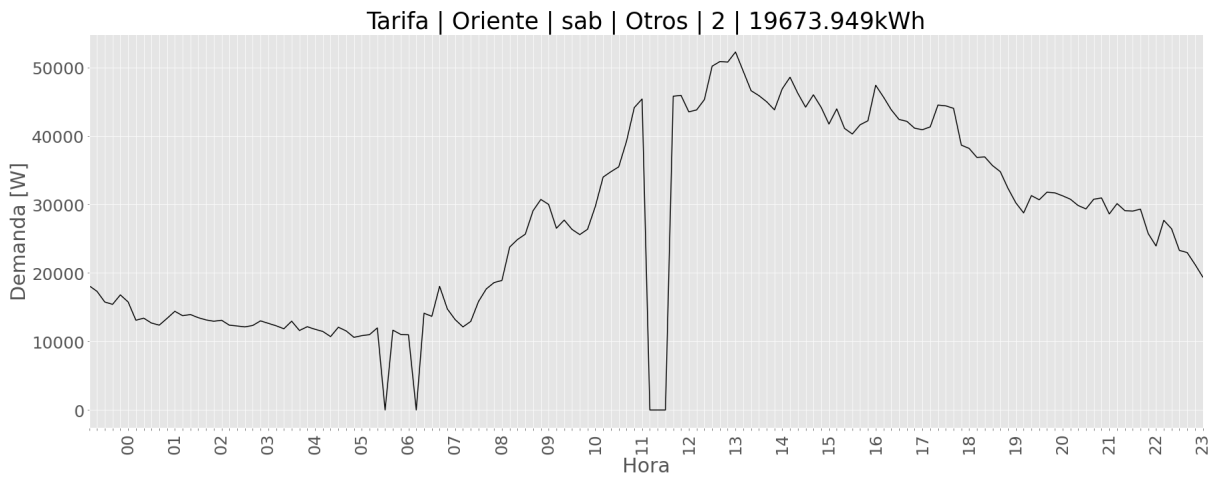
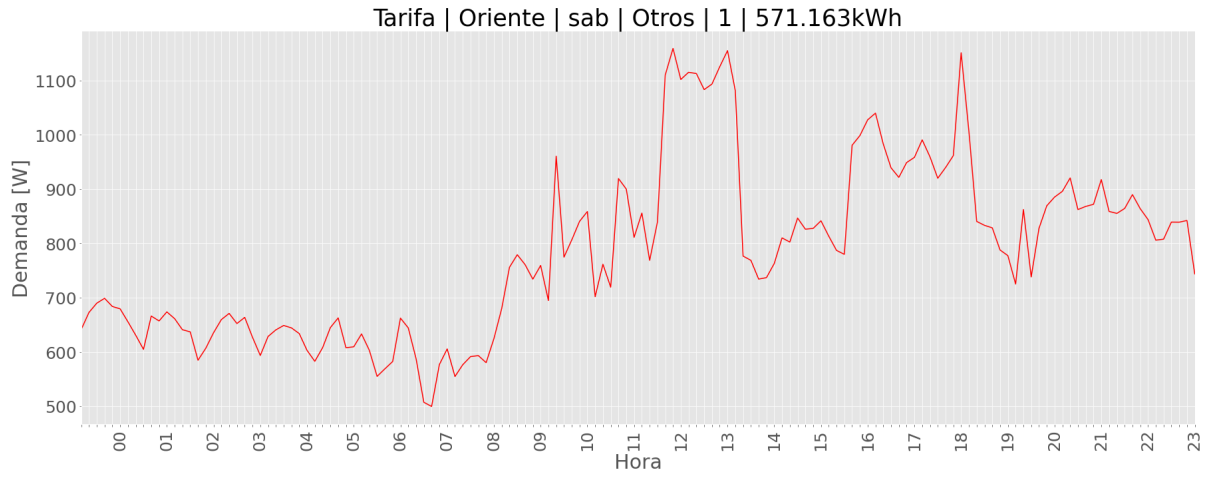
Datos de la Región Oriente en tipo de día sab

Estratos	Obser	EPmax [kW]	PPmax [kW]	PPmin [kW]	Ppro [kW]	E [kWh/mes]	Fc [%]
Residencial 1	74	2.34	0.16	0.05	0.08	59.5	50.4
Residencial 2	53	2.70	0.37	0.10	0.18	132.3	49.2
Residencial 3	13	1.37	0.51	0.27	0.36	261.4	70.9
Residencial 4	9	2.94	1.19	0.32	0.59	433.5	49.9
Otros 1	8	5.06	1.16	0.50	0.78	571.2	67.5
Otros 2	1	52.25	52.25	0.00	26.95	19673.9	51.6
Otros 3	2	79.96	62.79	29.11	43.48	31741.9	69.2

Perfiles de carga de cada Estrato de Mezcla Gaussiana



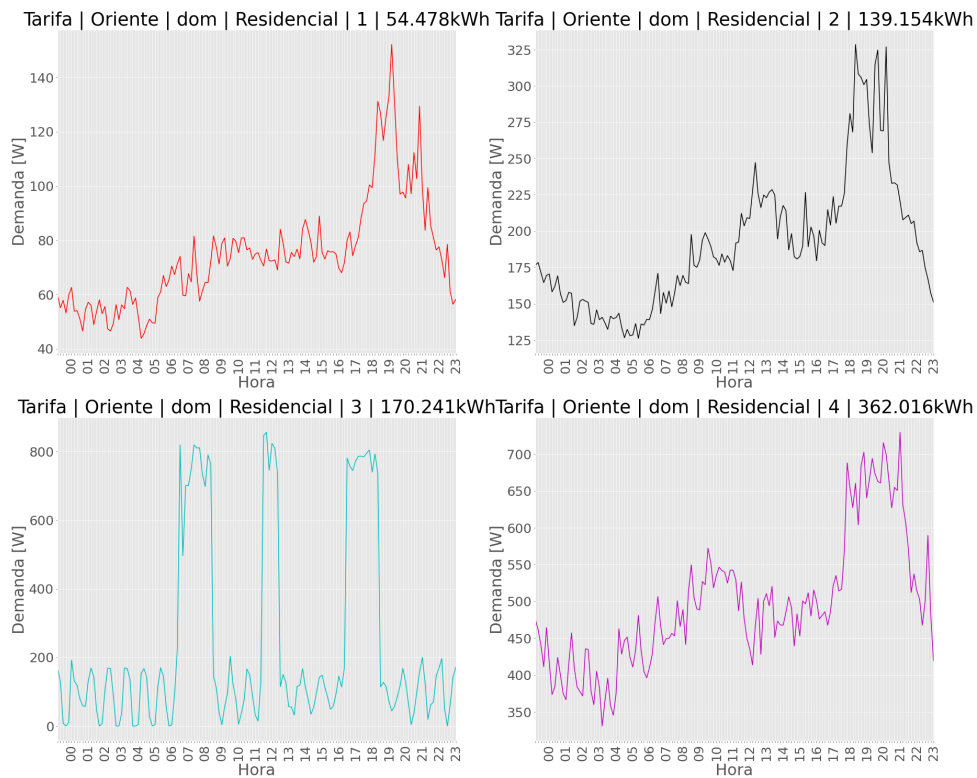
Perfiles de carga de cada Estrato de Mezcla Gaussiana



Datos de la Región Oriente en tipo de día dom

Estratos	Obser	EPmax [kW]	PPmax [kW]	PPmin [kW]	Ppro [kW]	E [kWh/mes]	Fc [%]
Residencial 1	98	2.38	0.15	0.04	0.08	54.5	49.1
Residencial 2	88	2.78	0.33	0.13	0.19	139.2	58
Residencial 3	3	0.92	0.86	0.00	0.23	170.2	27.2
Residencial 4	15	2.05	0.73	0.33	0.50	362	68
Otros 1	11	7.76	1.05	0.43	0.61	444.4	57.9
Otros 2	2	49.85	46.60	5.22	25.98	18964.3	55.7
Otros 3	2	84.73	64.95	28.53	44.65	32592.5	68.7

Perfiles de carga de cada Estrato de Mezcla Gaussiana



Perfiles de carga de cada Estrato de Mezcla Gaussiana

