

METODOLOGÍA DE CARACTERIZACIÓN DE PATRONES DE CONSUMO DE  
ENERGÍA ELÉCTRICA EN EL SISTEMA DE DISTRIBUCIÓN DE LA  
ELECTRIFICADORA DE SANTANDER

JUAN ANDRÉS MOJICA SUÁREZ  
ANDRÉS MAURICIO MUÑOZ ACOSTA

UNIVERSIDAD INDUSTRIAL DE SANTANDER  
FACULTAD DE INGENIERÍAS FÍSICO-MECÁNICAS  
ESCUELA DE INGENIERÍA ELÉCTRICA, ELECTRÓNICA Y DE  
TELECOMUNICACIONES  
ESPECIALIZACIÓN EN SISTEMAS DE DISTRIBUCIÓN DE ENERGÍA ELÉCTRICA  
BUCARAMANGA

2025

METODOLOGÍA DE CARACTERIZACIÓN DE PATRONES DE CONSUMO DE  
ENERGÍA ELÉCTRICA EN EL SISTEMA DE DISTRIBUCIÓN DE LA  
ELECTRIFICADORA DE SANTANDER

JUAN ANDRÉS MOJICA SUÁREZ  
ANDRÉS MAURICIO MUÑOZ ACOSTA

TRABAJO DE GRADO PARA OPTAR EL TÍTULO DE:  
ESPECIALISTAS EN SISTEMAS DE DISTRIBUCIÓN DE ENERGÍA ELÉCTRICA

DIRECTOR  
IVÁN DAVID SERNA SUÁREZ  
DOCTOR EN INGENIERÍA

UNIVERSIDAD INDUSTRIAL DE SANTANDER  
FACULTAD DE INGENIERÍAS FÍSICO-MECÁNICAS  
ESCUELA DE INGENIERÍA ELÉCTRICA, ELECTRÓNICA Y DE  
TELECOMUNICACIONES  
ESPECIALIZACIÓN EN SISTEMAS DE DISTRIBUCIÓN DE ENERGÍA ELÉCTRICA  
BUCARAMANGA

2025

## **AGRADECIMIENTOS**

Agradezco a Dios por darme la oportunidad de alcanzar esta meta personal. A mis padres, quienes han sido mi mayor fuente de motivación y apoyo constante. Expreso mi gratitud al director y a la evaluadora de esta monografía por sus valiosos comentarios, que fueron esenciales para lograr un trabajo de calidad. También agradezco a mi alma mater, la Universidad Industrial de Santander, por su formación y respaldo, y a la Electrificadora de Santander por su apoyo y por proporcionar la información necesaria para el desarrollo de este estudio. A mi compañero de trabajo y monografía , muchas gracias.

***Andrés Muñoz***

En primer lugar, agradezco profundamente a Dios por darme la fortaleza, sabiduría y perseverancia para alcanzar esta meta. A mis padres, por su amor incondicional, apoyo constante y por ser mi mayor inspiración. Expreso mi gratitud a la Universidad Industrial de Santander, por brindarme el espacio necesario para mi formación académica, y al director de la monografía, por su guía y compromiso en este proceso. Agradezco también a la Electrificadora de Santander, por facilitarme la información necesaria para el desarrollo de este proyecto. Finalmente, a mi compañero de monografía, por su dedicación y compañerismo.

***Juan Andres Mojica***

## TABLA DE CONTENIDO

	Pág.
INTRODUCCIÓN .....	11
1 OBJETIVOS .....	14
1.1 OBJETIVO GENERAL .....	14
1.2 OBJETIVOS ESPECÍFICOS .....	14
2 MARCO TEÓRICO Y CONCEPTUAL .....	15
2.1 PERFIL DE DEMANDA .....	15
2.2 MÉTODO DE AGRUPAMIENTO “ <i>K-MEANS</i> ” .....	19
2.3 MÉTODO DE REDUCCIÓN DE DIMENSIONALIDAD “ <i>T-SNE</i> ” .....	21
2.4 ÍNDICES DE VALIDEZ DE AGRUPAMIENTO “ <i>CVI’s</i> ” .....	22
3 ALGORITMO PARA LA CONSTRUCCIÓN DE PERFILES TÍPICOS DE DEMANDA EN UN SISTEMA DE DISTRIBUCIÓN CON DATOS SCADA .....	25
3.1 CONJUNTO DE DATOS .....	27
3.2 ESCALAMIENTO DE LOS DATOS .....	29
3.3 DETERMINAR NÚMERO DE CLÚSTERES ÓPTIMO .....	31
3.4 APLICAR MÉTODO <i>K-MEANS</i> .....	32
3.5 VALIDACIÓN CON <i>T-SNE</i> .....	32
3.6 EXTRACCIÓN DE CARACTERÍSTICAS.....	33
3.7 PROGRAMACIÓN EN <i>PYTHON</i> .....	35

4	PERFILES TÍPICOS DE DEMANDA EN EL SISTEMA DE DISTRIBUCIÓN DE ESSA	36
4.1	DESCRIPCIÓN DE CONJUNTO DE DATOS.....	36
4.2	CLÚSTERES ÓPTIMOS .....	39
4.3	APLICACIÓN DE <i>K-MEANS</i> PARA OBTENER PERFILES DE DEMANDA PREDOMINANTES.....	41
4.4	VALIDACIÓN CON <i>T-SNE</i> .....	46
4.5	AGRUPAMIENTO POR CARACTERÍSTICAS .....	50
4.6	COMPARACIÓN DE LAS DOS ETAPAS .....	54
5	RELACIÓN DE PERFILES TÍPICOS DE DEMANDA Y CARACTERÍSTICAS DE LOS USUARIOS .....	56
5.1	DESCRIPCIÓN DE LOS CLÚSTERES EN RELACIÓN CON LOS CIRCUITOS	56
5.2	CLASIFICACIÓN DE LOS TIPOS DE USUARIO POR CIRCUITO.....	59
5.3	UBICACIÓN GEOGRÁFICA DE LOS USUARIOS EN RELACIÓN CON LOS CLÚSTERES.....	60
6	CONCLUSIONES.....	64
7	RECOMENDACIONES .....	66
	BIBLIOGRAFÍA .....	68
	ANEXOS.....	72

## LISTA DE TABLAS

Tabla 1. Ejemplo conjunto de datos de demanda diaria. ....	28
Tabla 2. Clasificación de municipios por departamento del área de influencia ESSA .....	36
Tabla 3. Recuento por tiempo de usuario.....	39

## LISTA DE FIGURAS

Figura 1. Ejemplo de perfil de demanda diario .....	16
Figura 3. Diagrama de flujo.....	26
Figura 2. Perfiles de demanda de circuitos con datos normales vs escalamiento de datos.....	30
Figura 4. Área de influencia ESSA .....	38
Figura 5. Resultados índices de validez de agrupamiento “CVIS” .....	40
Figura 8. Perfiles típicos predominantes de demanda por clúster .....	42
Figura 6. Resultados validación <i>T-SNE</i> .....	48
Figura 7. <i>T-SNE</i> perplejidad 10 y tasa de aprendizaje automática .....	49
Figura 9. Resultados índices de validez de agrupamiento “CVIS” .....	50
Figura 10. <i>T-SNE</i> centroides - clúster .....	52
Figura 11. Perfiles típicos predominantes de demanda por clúster .....	53
Figura 12. Diagrama de <i>Sankey</i> . Transición de clústeres entre metodologías .	54
Figura 13. Clústeres por circuitos .....	57
Figura 14. Clústeres por cantidad de usuarios .....	58
Figura 15. Clústeres por tipos de usuarios .....	59
Figura 16. Mapa de Santander por tipo de clúster .....	62

## LISTA DE ANEXOS

Los anexos se encuentran adjuntos y se pueden consultar en el Repositorio Institucional.

**Anexo A.** Scripts de *Python*

**Anexo B.** Perfiles de demanda de circuitos ESSA

**Anexo C.** Clasificación de usuarios ESSA

**Anexo D.** Coordenadas de transformadores de distribución

## RESUMEN

**TÍTULO:** METODOLOGÍA DE CARACTERIZACIÓN DE PATRONES DE CONSUMO DE ENERGÍA ELÉCTRICA EN EL SISTEMA DE DISTRIBUCIÓN DE LA ELECTRICADORA DE SANTANDER

**AUTORES:** JUAN ANDRÉS MOJICA SUÁREZ  
ANDRÉS MAURICIO MUÑOZ ACOSTA

**PALABRAS CLAVE:** patrones de consumo, perfiles de demanda, SCADA, K-MEANS, t-SNE, clústeres, distribución eléctrica, ESSA, Python.

### DESCRIPCIÓN:

La demanda de energía en un sistema de potencia varía a lo largo del día, influenciada por los patrones de consumo de los usuarios. Entender estos patrones es clave para abordar los desafíos que enfrentan las empresas del sector eléctrico. Un sistema de distribución suele contar con equipos de medición que recopilan información de la red en intervalos regulares. En este contexto, gestionar esta información representa un desafío para las empresas distribuidoras de energía, que buscan aprovecharla para obtener beneficios. En esta monografía se presenta el diseño de una metodología para la caracterización de patrones de consumo de energía eléctrica en un sistema de distribución, utilizando datos recopilados a través del sistema SCADA. La metodología emplea técnicas de aprendizaje no supervisado, como K-MEANS y t-SNE, para identificar y clasificar los perfiles típicos de demanda de un conjunto de entrada según sus similitudes o características, en grupos denominados clústeres. Los resultados obtenidos en esta metodología permiten abordar objetivos clave, como mejorar los pronósticos de demanda, diseñar tarifas personalizadas, planificar la expansión de la infraestructura eléctrica y optimizar la gestión de la demanda en el sistema de distribución donde se aplique.

Adicionalmente, se presenta la implementación de esta metodología con datos provenientes de circuitos del sistema de distribución local de la Electricadora de Santander S.A. E.S.P. (ESSA). Su aplicación permitió clasificar los circuitos según patrones de consumo similares, identificando un número adecuado de clústeres. Al final, se llevó a cabo un análisis para explorar la relación entre las características de los usuarios, como su tipo (residencial, comercial o industrial), su ubicación geográfica y los clústeres obtenidos.

---

\* Monografía

\*\* Facultad de Ingenierías Físico-Mecánicas. Escuela de Ingenierías Eléctrica, Electrónica y Telecomunicaciones. Especialización en Sistema de Distribución de Energía Eléctrica.

Director: Iván David serna Suárez. Doctor en Ingeniería

## ABSTRACT

**TITLE:** METHODOLOGY FOR THE CHARACTERIZATION OF ELECTRIC ENERGY CONSUMPTION PATTERNS IN THE DISTRIBUTION SYSTEM OF ELECTRIFICADORA DE SANTANDER

**AUTHORS:** JUAN ANDRÉS MOJICA SUÁREZ  
ANDRÉS MAURICIO MUÑOZ ACOSTA

**KEY WORDS:** consumption patterns, typical demand profiles, *SCADA*, *K-Means*, *t-SNE*, clusters, electrical distribution, ESSA.

### DESCRIPTION:

Energy demand in a power system fluctuates throughout the day, influenced by user consumption patterns. Understanding these patterns is key to addressing the challenges faced by companies in the electric sector. A distribution system typically includes measurement equipment that collects network information at regular intervals. In this context, managing this information represents a challenge for energy distribution companies, which aim to leverage it for benefits. This monograph presents the design of a methodology for characterizing energy consumption patterns in a distribution system using data collected through the *SCADA* system. The methodology employs unsupervised learning techniques, such as *K-MEANS* and *t-SNE*, to identify and classify typical demand profiles from an input dataset based on their similarities or characteristics, grouping them into clusters. The results of this methodology support key objectives, such as improving demand forecasting, designing personalized tariffs, planning the expansion of electrical infrastructure, and optimizing demand management in the applied distribution system.

Additionally, the implementation of this methodology is demonstrated using data from circuits within the local distribution system of Electrificadora de Santander S.A. E.S.P. (ESSA). Its application enabled the classification of circuits based on similar consumption patterns, identifying an appropriate number of clusters. Finally, an analysis was conducted to explore the relationship between user characteristics, such as type (residential, commercial, or industrial), geographical location, and the resulting clusters.

---

\* Monograph

\*\* Faculty of Physical-Mechanical Engineering. School of Electrical, Electronic and Telecommunications Engineering. Specialization in Electric Power Distribution System. Director: Iván David Serna Suárez. PhD in Engineering

## INTRODUCCIÓN

La demanda de energía eléctrica en un sistema de distribución es variable en el tiempo y fluctúa según las necesidades de los usuarios. Esta variabilidad resalta la importancia de establecer los valores de demanda, en términos de potencia activa, para abordar diversas problemáticas en la ingeniería eléctrica. La identificación de los patrones de consumo de los usuarios puede ayudar a determinar estos valores, considerando el periodo horario o tipo de día que se esté estudiando.

Los patrones de consumo pueden ser analizados desde puntos centralizados de medición, como las subestaciones, o desde puntos más desagregados, como los transformadores de distribución o de mismos medidores individuales de los usuarios, si se cuenta con este tipo de tecnología. La adquisición de datos de estos medidores y su posterior procesamiento para elaborar perfiles típicos de demanda representan un desafío significativo para las empresas comercializadoras y distribuidoras de energía. Estos perfiles, conocidos en la literatura como *TLP (Typical Load Profile* por sus siglas en inglés) [1], [2], consisten en valores horarios de consumo de energía diario expresados en por unidad y están asociados a categorías específicas de consumidores y tipos de día. La Electrificadora de Santander S.A. E.S.P. (ESSA) cuenta con medición horaria desagregada para cada circuito de distribución en toda su área de influencia a través de su sistema *SCADA (Supervisory Control and Data Acquisition)*. Todos estos registros de medición son almacenados en bases de datos. Sin embargo, no cuenta con un procedimiento establecido para gestionar y tratar esta información, y posteriormente darle

valor mediante la caracterización de la demanda. Esta monografía busca establecer un método de caracterización del consumo de energía eléctrica de los usuarios de ESSA, utilizando los valores de potencia activa registrados en su sistema *SCADA* como fuente de datos. Dentro de esta caracterización, se pretende agrupar a los usuarios con patrones de consumo similares en segmentos más pequeños para establecer su perfil típico de demanda. Para realizar esto, es importante también determinar cuál es el número óptimo de agrupaciones a realizar. Esto dependerá de los comportamientos de consumo predominantes en los usuarios de ESSA.

Una vez que se hayan establecido los *TLP*, estos podrán ser asociados con cada usuario, transformador de distribución, circuito de distribución, subestación o zona del área de influencia de ESSA, según las necesidades específicas de los estudios a realizar. Esta información permitirá abordar las hipótesis relacionadas con los factores que influyen en la forma de la curva de demanda. Se considerarán variables como la cantidad y tipo de usuarios (residencial, comercial e industrial), la ubicación geográfica (municipios), las condiciones climáticas, entre otros.

La implementación de una metodología para caracterizar los patrones de consumo de energía eléctrica en sistemas de distribución representa un gran beneficio para las empresas involucradas en este sector. En este trabajo, se aporta un mayor valor al caso específico del sistema de distribución de ESSA. Comprender la variabilidad del consumo de energía eléctrica de los usuarios a lo largo del tiempo permitirá a ESSA y a otros distribuidores o comercializadores mejorar sus estudios eléctricos, realizar mejores predicciones de demanda [2], [3], [4], [5], [6], mejorar el dimensionamiento de sus redes

eléctricas [5], [7], [8], ofrecer tarifas preferenciales [8], [9], capturar el comportamiento de la demanda antes de la penetración de la generación distribuida a gran escala [10], entre otras aplicaciones.

Conocer los patrones de consumo aporta a una mejor planificación para el desarrollo futuro de la red eléctrica, brindando la oportunidad de tomar decisiones informadas sobre la expansión de la infraestructura y una determinación óptima de los recursos de generación [10], [11]. La comprensión más profunda de los patrones de consumo y los factores que influyen en ellos también permite realizar predicciones de demanda más precisas. Los grupos de usuarios con consumos similares tienden a exhibir perfiles de demanda similares a largo plazo. Por consiguiente, el empleo de algoritmos de pronóstico específicos para estos grupos puede reducir significativamente el margen de error [2], [4], [5]. Por otra parte, esta caracterización de los patrones de consumo puede utilizarse para diseñar ofertas comerciales adaptadas a las necesidades y comportamientos de consumo de los usuarios. Estas tarifas pueden incentivar la eficiencia energética, fomentar la adopción de energías renovables y promover el uso responsable de la energía. Todo ello contribuye a mejorar la satisfacción del cliente y la gestión de la demanda en la red eléctrica [6], [8], [9], [11].

# 1 OBJETIVOS

## 1.1 OBJETIVO GENERAL

Diseñar una metodología para caracterizar los patrones de consumo de energía eléctrica en el sistema de distribución de la Electrificadora de Santander S.A. E.S.P.

## 1.2 OBJETIVOS ESPECÍFICOS

- Diseñar un algoritmo para la construcción de perfiles típicos de demanda en un sistema de distribución utilizando datos del sistema SCADA mediante el método *K-Means*.
- Identificar los perfiles típicos de demanda predominantes en el sistema de distribución de ESSA aplicando el algoritmo diseñado.
- Establecer relación entre los perfiles típicos de demanda identificados en el sistema de distribución de ESSA y las características de los usuarios.

## 2 MARCO TEÓRICO Y CONCEPTUAL

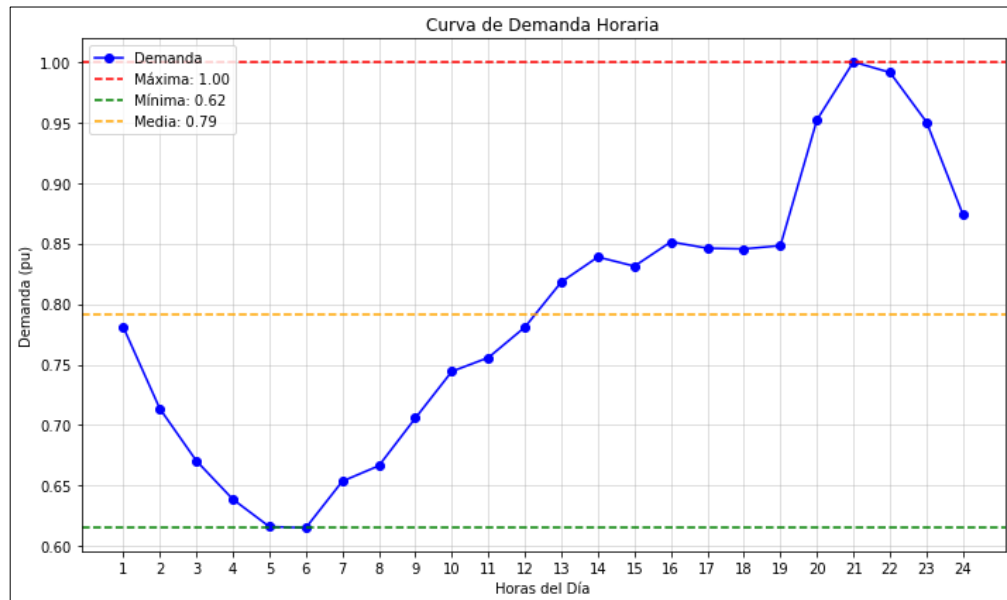
### 2.1 PERFIL DE DEMANDA

Los sistemas de distribución tienen la importante tarea de suministrar energía eléctrica a los usuarios proveniente de los sistemas de transmisión, garantizando un servicio ininterrumpido y cumpliendo con los estándares de calidad establecidos [12]. En este sentido, las empresas distribuidoras se han visto obligadas a estudiar el consumo de energía de sus usuarios para abordar diversas problemáticas que conlleva este servicio y de esta forma evitar posibles colapsos en el sistema.

La cantidad de energía que un grupo de usuarios necesita en un momento determinado se conoce como demanda, la cual se expresa en kilovatios (kWh). Esta demanda es variable en el tiempo debido a una serie de factores, tales como condiciones climáticas, variables sociales, económicas, así como los hábitos de consumo de los usuarios [13]. La variación de la demanda a lo largo de un periodo de tiempo se representa mediante lo que se conoce como la curva de carga, curva de demanda o perfil de demanda. Este periodo de tiempo puede ser diario, semanal, mensual o anual dependiendo del objetivo del estudio; a su vez este periodo es dividido en intervalos más pequeños [10]. La duración que se fije para esos intervalos dependerá mucho de las especificaciones que tengan los equipos de medición, aunque los periodos de tiempo mayormente usados son de 15, 30 minutos y 1 hora [6]. En este trabajo, se analiza el periodo de demanda a nivel diario,

con intervalos de una hora. En la Figura 1 se observa un ejemplo ilustrativo del perfil de demanda diaria.

Figura 1. Ejemplo de perfil de demanda diario



Fuente: Autores

En la literatura, el perfil de demanda suele representarse de manera normalizada, lo que implica graficar un vector con valores en un rango de  $[0,1]$ , es decir, en escala de por unidad, usualmente con base en la potencia máxima del período de análisis, aunque también se puede aplicar el “*Feature scaling*” como método de normalización [14]. A partir de este perfil, se pueden calcular diversos parámetros para extraer información relevante de la curva, entre los cuales se incluyen:

- **Energía consumida:** la energía durante el periodo o parte de él, obtenida calculando el área bajo la curva.

- **Demanda máxima:** la magnitud más alta de consumo durante el periodo estudiado.
- **Demanda mínima:** la magnitud más baja de consumo durante el periodo representado.
- **Demanda media:** el promedio de todos los valores de demanda en el periodo.
- **Factor de carga:** la relación entre la demanda media y la demanda máxima del sistema en un periodo dado.

Analizar estos parámetros mencionados nos puede revelar características importantes de los consumos de los usuarios. Por ejemplo, el factor de carga indica el grado en que el pico de la carga es sostenido durante el periodo. Un factor de carga de 1 indica que la demanda máxima se mantiene constante, mientras que un factor de carga alto (por ejemplo, 0.9) sugiere que la curva de carga tiene pocas variaciones. En contraste, un factor de carga bajo (por ejemplo, 0.2) indica muchas variaciones en la curva de carga, con picos y valles pronunciados [15].

A pesar de la naturaleza aleatoria de la demanda, como se mencionó anteriormente, las curvas de demanda pueden exhibir patrones recurrentes, lo que ha despertado el interés de los investigadores para estudiar y caracterizar estos comportamientos. Según se detalla en [6], las publicaciones relacionadas con el estudio de los patrones de demanda comenzaron a surgir a finales de la década de los 50's. Desde entonces, estos estudios han buscado aplicaciones diversas a la caracterización de los patrones de consumo, como el pronóstico de demanda, el

análisis de uso final de energía, eficiencia energética, detección de anomalías en el consumo, predicción basada en el reconocimiento de patrones, identificación del comportamiento de los consumidores y la representación de los patrones de consumo.

Recientes artículos han adoptado la denominación de "*Typical Load Profile*" (*TLP*) para describir la curva de demanda que representa los patrones de consumo de grupos específicos de usuarios, clasificados según categorías como residencial, industrial o comercial, así como períodos temporales como invierno o verano, además de días hábiles y fines de semana [1], [2]. La obtención de datos para construir estos perfiles puede realizarse a través de medidores inteligentes, aunque algunos estudios han preferido utilizar datos del sistema *SCADA* para análisis de sistemas de distribución [4], [16]. Estos *TLP* pueden ser agrupados en conjuntos distintos mediante métodos de agrupamiento (conocidos en inglés como clustering), donde cada conjunto exhibe características de consumo similares. Las técnicas de agrupación más comúnmente empleadas en este contexto incluyen "*Follow the leader*", lógica borrosa, agrupamiento jerárquico, *K-Means*, y redes neuronales artificiales, entre otras [8]. En esta monografía se empleará el método *K-Means* debido a su amplia utilización en diversas disciplinas y en distintos tipos de datos, incluyendo la clasificación de patrones de consumo de energía eléctrica, así como a los resultados favorables que ha mostrado [4], [17]. A continuación, se muestra una breve descripción de este método.

## 2.2 MÉTODO DE AGRUPAMIENTO “*K-MEANS*”

El método de agrupación *K-Means* es un algoritmo popular utilizado en aprendizaje automático y minería de datos para la partición de conjuntos de datos en un número predefinido de grupos. Su objetivo es agrupar los datos en  $k$  grupos, minimizando la varianza dentro de cada grupo y maximizando la varianza entre grupos. Este método utiliza la distancia euclidiana como métrica para calcular la similitud entre los puntos de datos y los centroides de los grupos [18]. La distancia euclidiana entre dos puntos  $p$  y  $q$  se calcula utilizando la siguiente fórmula:

$$d(p, q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$$

El algoritmo *K-Means* sigue estos pasos básicos:

1. **Inicialización:** Seleccionar  $k$  centroides iniciales. Esto puede hacerse de forma aleatoria o utilizando algún método heurístico.
2. **Asignación de puntos a grupos:** Asignar cada punto de datos al grupo cuyo centroide es el más cercano.
3. **Actualizar centroides:** Calcular los nuevos centroides de cada grupo basados en las observaciones asignadas a él.
4. **Repetir:** Iterar los pasos 2 y 3 hasta que los centroides ya no cambien significativamente o se alcance un criterio de convergencia predefinido.

La determinación de un número óptimo de 'clusters' ( $k$ ) es un paso crucial para utilizar el algoritmo *K-Means* de manera efectiva. Existen diversos métodos que

pueden emplearse para estimar el valor óptimo de  $k$ , siendo uno de los más utilizados el método del codo (*Elbow Method*). Este enfoque implica trazar la suma de las distancias al cuadrado entre cada punto de datos y su centroide asignado para distintos valores de  $k$ . En este método, buscamos el valor de  $k$  donde la disminución en la suma de las distancias al cuadrado se desacelera, formando una curva que recuerda a un codo. Este punto de inflexión se considera como el número óptimo de clústeres [18], [19].

Se seleccionó el método de agrupamiento *K-Means* para esta monografía debido a varias razones. En primer lugar, su simplicidad y facilidad de implementación lo convierten en una herramienta accesible y efectiva para la tarea en cuestión. *K-Means* es conocido por ser un algoritmo rápido y robusto, que ofrece eficiencia computacional, especialmente en conjuntos de datos grandes y de alta dimensión. Además, este método preserva la estructura temporal de los datos, lo que es crucial para mantener la integridad de la información original. *K-Means* también se adapta bien a conjuntos de datos de diversos tamaños, asegurando una rápida convergencia [20]. Por último, es uno de los métodos más citados en la literatura de sistemas de potencia relacionada con el agrupamiento de patrones de consumo de energía eléctrica, lo que refuerza su idoneidad en este contexto [17].

Por otro lado, *t-SNE* es una técnica de análisis de datos que, aunque tiene objetivos y funcionalidades distintas a *K-Means*, puede complementarse con este último en un flujo de trabajo. En este enfoque, *K-Means* se utiliza para identificar los clústeres en los datos, mientras que *t-SNE* se emplea para explorar y visualizar esos clústeres

en un espacio de menor dimensión. El algoritmo *t-SNE* se explicará con mayor detalle en la siguiente sección.

### **2.3 MÉTODO DE REDUCCIÓN DE DIMENSIONALIDAD “*T-SNE*”**

El algoritmo *t-distributed Stochastic Neighbor Embedding (t-SNE)* es una técnica avanzada para la reducción de dimensionalidad, diseñada para facilitar la visualización de datos de alta dimensionalidad al proyectarlos en espacios de menor dimensión, como 2D o 3D. Su principal fortaleza es la capacidad de revelar la estructura subyacente de los datos al preservar las relaciones de proximidad entre puntos en el espacio original [21], [22].

El desempeño de *t-SNE* está altamente influenciado por la configuración de sus parámetros principales, como la perplejidad y la tasa de aprendizaje. La perplejidad, aunque matemáticamente compleja de definir, puede interpretarse como una medida de dispersión que afecta el balance entre las relaciones locales y globales en los datos. Por otro lado, la tasa de aprendizaje controla el tamaño de los pasos durante cada iteración del algoritmo: valores demasiado altos pueden generar comportamientos erráticos que dificulten la convergencia, mientras que valores muy bajos pueden ralentizar el proceso o llevar al algoritmo a quedar atrapado en mínimos locales.

A través de esta técnica *t-SNE*, se simplifica la representación de los datos, mejorando la capacidad de detectar y analizar patrones y grupos ocultos que

pueden estar presentes en datos complejos y multidimensionales. Al aplicar *t-SNE*, es posible transformar datos en un espacio de menor dimensión donde los grupos o patrones de agrupamiento se vuelven visualmente evidentes. Esto permite una comprensión más intuitiva y esclarecedora de los resultados de análisis o agrupamiento, especialmente cuando los datos presentan relaciones complejas o variables de gran diversidad.

La visualización que proporciona *t-SNE* no solo apoya la validación de la calidad del algoritmo de agrupamiento, permitiendo ver si los datos han sido correctamente separados en grupos con sentido lógico, sino que también genera información adicional. Al examinar los patrones visuales en el espacio reducido, se puede obtener una perspectiva más profunda sobre la estructura y relaciones internas de los datos, lo cual enriquece el contexto más amplio de la investigación. Además, junto con el uso de matrices de similitud, *t-SNE* permite corroborar la consistencia de los grupos y explorar conexiones que podrían pasar desapercibidas en otras representaciones [21].

#### **2.4 ÍNDICES DE VALIDEZ DE AGRUPAMIENTO “CVI’s”**

En el análisis de agrupamiento, un proceso de aprendizaje no supervisado, la evaluación objetiva de la calidad de los clústeres resultantes constituye un desafío importante. A diferencia de los métodos supervisados, en los cuales se puede evaluar el rendimiento mediante la comparación con etiquetas conocidas, en el agrupamiento no se dispone de una referencia específica para validar los

resultados. Por esta razón, es necesario emplear métodos que permitan medir la calidad de los grupos de manera cuantitativa. De forma intuitiva, un buen método de agrupamiento debe garantizar que cada clúster sea compacto (es decir, los elementos dentro de un clúster deben estar lo más cercanos posibles entre sí) y que exista una separación significativa entre los diferentes clústeres (de modo que los clústeres estén claramente diferenciados y distantes unos de otros).

Para evaluar estas propiedades y la calidad general del agrupamiento, se pueden utilizar diversas métricas conocidas como Índices de Validez de Agrupamiento (*Cluster Validity Indices, CVIs*). Estos índices permiten analizar y validar la coherencia y la estructura interna de los agrupamientos obtenidos, proporcionando herramientas objetivas para interpretar los resultados del análisis de clustering [17].

Algunos de los índices más utilizados son:

- **Índice de Silueta:** Este índice evalúa qué tan bien está cada punto agrupado dentro de su propio clúster en comparación con los clústeres vecinos. El valor del índice de silueta varía entre -1 y 1, donde los valores cercanos a 1 indican que los puntos están bien agrupados dentro de su clúster, y alejados de otros clústeres. Un valor cercano a 0 sugiere que el punto está en el límite entre dos clústeres, mientras que los valores negativos indican que el punto podría estar mal asignado a su clúster.
- **Índice *Davies-Bouldin*:** Este indicador mide la dispersión de cada clúster y la distancia entre clústeres. Se calcula como la relación entre la dispersión dentro de un clúster y la distancia entre los centroides de dos clústeres. Un

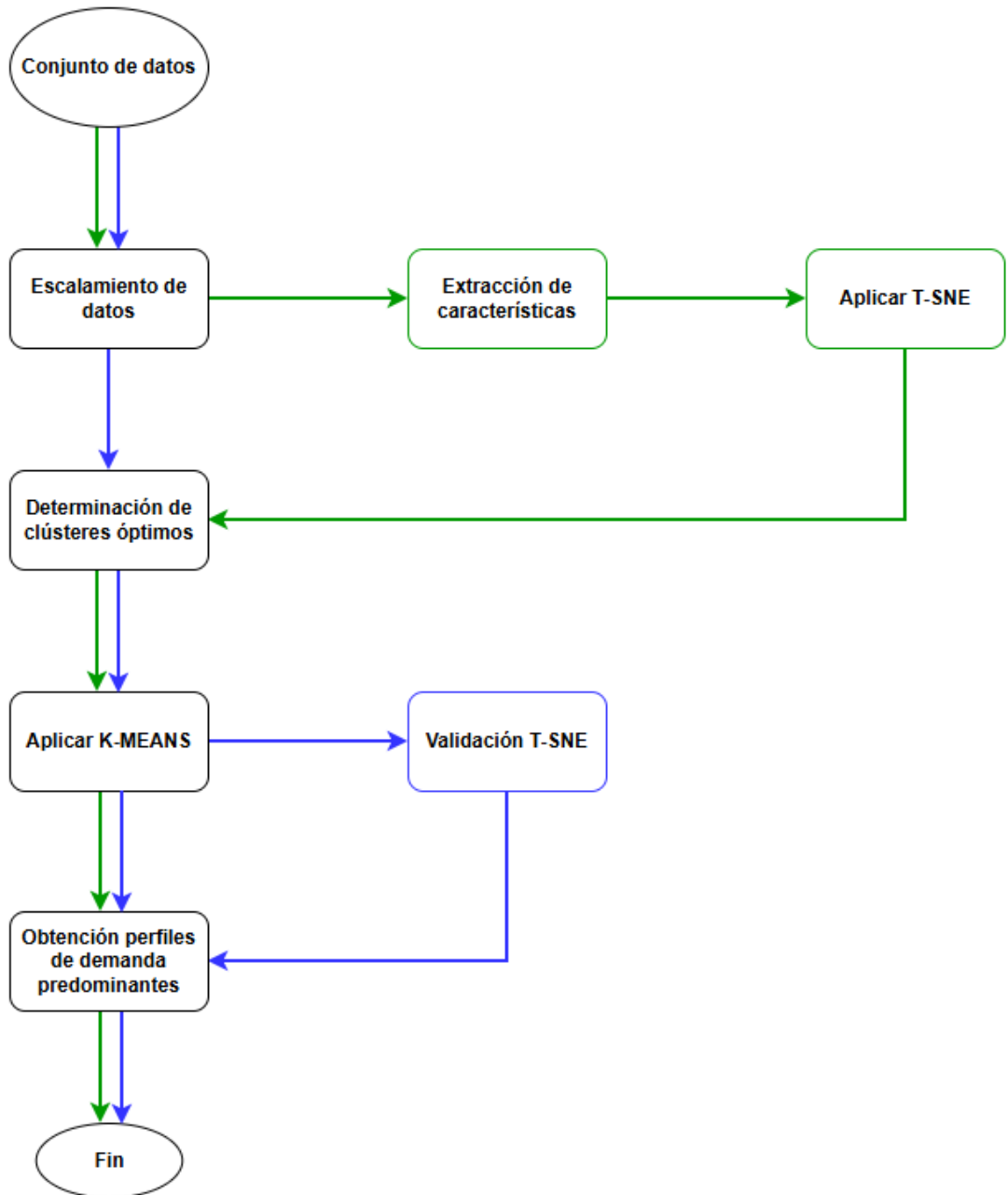
valor bajo del índice de *Davies-Bouldin* indica una buena estructura de agrupamiento, ya que sugiere que los clústeres están bien separados y son relativamente compactos. En cambio, valores altos sugieren que los clústeres son más dispersos y están menos diferenciados.

- **Error Cuadrático Medio (MSE):** Aunque normalmente se utiliza en problemas de regresión y predicción, el Error Cuadrático Medio (MSE) también se emplea en la evaluación de agrupamientos. En este contexto, mide la diferencia entre las distancias de los puntos a sus centroides dentro de cada clúster. El MSE se calcula como el promedio de los errores al cuadrado, es decir, la diferencia entre el valor real (distancia real del punto al centroide) y el valor predicho (distancia estimada al centroide) elevado al cuadrado, promediado para todos los puntos. Un MSE bajo indica que los puntos están más cerca de sus respectivos centroides, lo que refleja una mayor cohesión dentro de los clústeres [17], [20].

### **3 ALGORITMO PARA LA CONSTRUCCIÓN DE PERFILES TÍPICOS DE DEMANDA EN UN SISTEMA DE DISTRIBUCIÓN CON DATOS SCADA**

Para facilitar la comprensión de la metodología, se comienza presentando el diagrama de flujo de la Figura 2 que describe cada una de las etapas del proceso. Luego, se detalla cada etapa de manera individual, proporcionando un análisis más profundo de los pasos involucrados. Posteriormente, se describe el conjunto de datos que debe ser utilizado, destacando sus propiedades y especificaciones clave. Finalmente, se resumen las herramientas empleadas en el desarrollo y la aplicación del algoritmo.

Figura 2. Diagrama de flujo del algoritmo para la construcción de perfiles típicos de demanda



Fuente: Autores

La metodología propuesta contempla dos flujos paralelos, cada uno basado en un enfoque específico para generar entradas diferentes al algoritmo *K-MEANS*. El primer flujo utiliza los perfiles de demanda originales, previamente escalados, mientras que el segundo emplea características de los perfiles de demanda reducidos mediante *t-SNE* a dos dimensiones. Como resultado, se obtienen dos grupos de perfiles de demanda predominantes, permitiendo seleccionar la representación que mejor se ajuste a la necesidad del caso estudio.

### 3.1 CONJUNTO DE DATOS

El conjunto datos debe corresponder a mediciones reales de potencia activa o corriente de circuitos alimentadores de distribución y abarcar un solo día, con mediciones horarias que comprendan un total de 24 periodos. Todos los registros deben conservar los mismos prefijos en las unidades de medida; es decir, todos los datos deben estar expresados en Megavatios, kilovatios o Amperios.

La fuente de datos puede estar almacenada en una hoja de cálculo y debe mantener la siguiente estructura de columnas:

- **ESTACIÓN:** Nombre de la subestación eléctrica a la cual pertenece el circuito alimentador.
- **CAMPO:** Nombre del circuito alimentador del cual se tomó la información.
- **HORA:** Periodo del día en el que se capturó la medición.
- **MEDIDA:** Valor de potencia activa o de la corriente.

En algunos casos, para optimizar los resultados, se pueden emplear datos preprocesados que reflejen mejor el comportamiento típico del circuito alimentador. Esto es útil, ya que en un día determinado pueden presentarse situaciones como fallas en el circuito, operación en suplencia con otro alimentador, o pérdida de información en los equipos de medida. Las empresas distribuidoras de energía suelen disponer de este tipo de datos para sus análisis operativos. En esta monografía se emplea un procedimiento interno de ESSA que procesa un conjunto de medidas dentro de un rango de tiempo determinado. Este método establece un valor típico por circuito alimentador para cada periodo horario, tomando como referencia el tercer cuartil. De este modo, los valores atípicos bajos, como fallas parciales o completas, quedan por debajo del tercer cuartil, mientras que los valores de mayor magnitud, asociados a cargas adicionales por suplencia, se ubican por encima de este umbral.

A continuación, se muestra un ejemplo de conjunto de datos de entrada correspondiente a un único circuito alimentador:

Tabla 1. Ejemplo conjunto de datos de demanda diaria.

<b>ESTACION</b>	<b>CAMPO</b>	<b>HORA</b>	<b>MEDIDA [MW]</b>
<b>SAcuarel</b>	36-501	00:00	1.1972
<b>SAcuarel</b>	36-501	01:00	1.1465
<b>SAcuarel</b>	36-501	02:00	1.0948
<b>SAcuarel</b>	36-501	03:00	1.0908
<b>SAcuarel</b>	36-501	04:00	1.0935
<b>SAcuarel</b>	36-501	05:00	1.1694
<b>SAcuarel</b>	36-501	06:00	1.3062
<b>SAcuarel</b>	36-501	07:00	1.5559
<b>SAcuarel</b>	36-501	08:00	1.8952
<b>SAcuarel</b>	36-501	09:00	1.8887

<b>SAcuarel</b>	36-501	10:00	2.0500
<b>SAcuarel</b>	36-501	11:00	2.0847
<b>SAcuarel</b>	36-501	12:00	2.0574
<b>SAcuarel</b>	36-501	13:00	1.7112
<b>SAcuarel</b>	36-501	14:00	1.9049
<b>SAcuarel</b>	36-501	15:00	1.9439
<b>SAcuarel</b>	36-501	16:00	1.8078
<b>SAcuarel</b>	36-501	17:00	1.6224
<b>SAcuarel</b>	36-501	18:00	1.5761
<b>SAcuarel</b>	36-501	19:00	1.8497
<b>SAcuarel</b>	36-501	20:00	1.8045
<b>SAcuarel</b>	36-501	21:00	1.6759
<b>SAcuarel</b>	36-501	22:00	1.4644
<b>SAcuarel</b>	36-501	23:00	1.2732

Fuente: Base de datos ESSA.

### 3.2 ESCALAMIENTO DE LOS DATOS

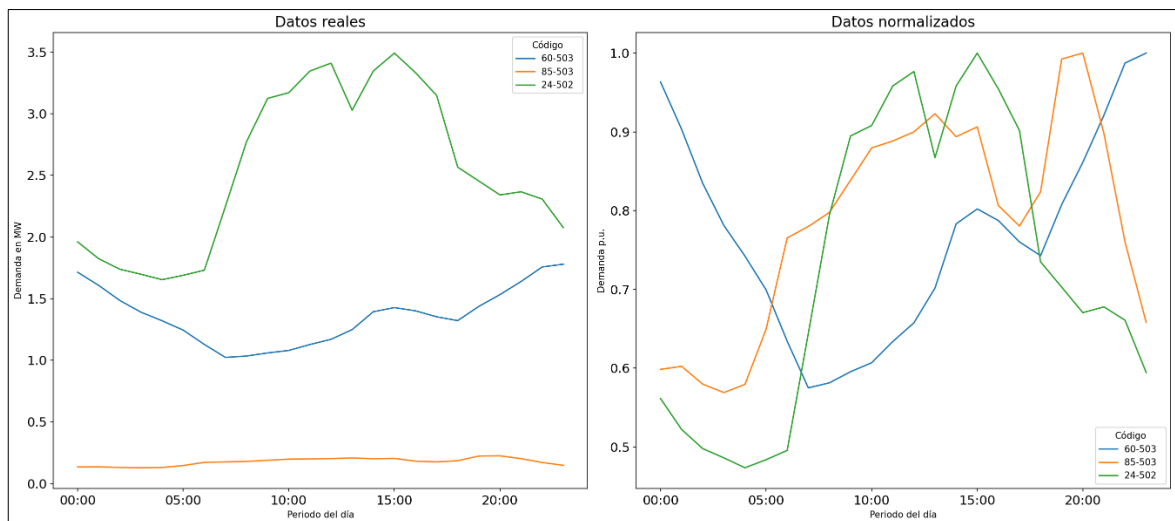
La mayoría de los estudios sobre patrones de consumo de energía eléctrica en el estado del arte se enfocan en la silueta o forma del consumo diario de la demanda, en lugar de su magnitud. Este enfoque facilita la identificación de características y similitudes, lo que a su vez permite realizar agrupamientos más precisos. Por ello, antes de iniciar cualquier análisis, es fundamental escalar o normalizar los datos.

Existen diversas técnicas para normalizar los datos; una de ellas es el escalamiento por valor absoluto máximo. Este método transforma cada valor del conjunto de datos de manera que el valor absoluto máximo se normalice a 1.0. En otras palabras, cada dato se divide entre el valor absoluto más alto de su respectiva característica, lo que produce un rango de valores entre -1 y 1.

Este método ha sido seleccionado para la presente metodología por su simplicidad conceptual, su facilidad de implementación y su capacidad para preservar la dispersión de los datos. En consecuencia, se aplicará este tipo de escalamiento al conjunto de datos de entrada, escalando cada curva de demanda de cada circuito alimentador de esta manera. Todos los gráficos presentados mantendrán esta escala, al igual que las salidas generadas a partir de esta metodología, que se expresarán consistentemente en el mismo formato.

La Figura 3 muestra una representación gráfica que ilustra cómo el escalamiento de los datos influye en la comparación de varios perfiles de demanda:

Figura 3. Perfiles de demanda de circuitos con dato reales vs. datos normalizados



Fuente: Autores

### 3.3 DETERMINAR NÚMERO DE CLÚSTERES ÓPTIMO

Uno de los principales desafíos al aplicar algoritmos de agrupamiento como *K-MEANS* es determinar el número óptimo de clústeres. Para abordar esta tarea, se empleará un enfoque que combina el método de la curva de codo con el cálculo de índices de validación de agrupamiento, tales como el índice de silueta, el índice de *Davies-Bouldin* y el error cuadrático medio. Estos métodos se complementan entre sí y permiten evaluar la calidad de las agrupaciones desde diferentes perspectivas, facilitando la selección del número óptimo de clústeres.

El proceso inicia aplicando *K-MEANS* en un rango de valores de clústeres, típicamente de 2 a 20. Para cada agrupamiento generado, se evalúa la variabilidad explicada, construyendo la curva de codo que identifica el punto de inflexión donde la mejora en la compactación de los clústeres comienza a estabilizarse. De manera simultánea, se calculan los índices de validación, los cuales se representan gráficamente para este mismo rango de clústeres. Esto permite una comparación visual y cuantitativa que respalda la elección final del número de agrupaciones.

En este proceso, cada perfil de demanda del conjunto de datos de entrada se considera como una variable que determina su asignación en uno de los  $k$  grupos. La distancia euclidiana se utiliza como medida de similitud, calculando la suma de las distancias entre cada periodo horario del perfil y el correspondiente valor promedio de su curva centroide. Este análisis conjunto asegura que la selección del

número de clústeres sea robusta y respalde de manera confiable la identificación de patrones representativos de consumo de energía.

### **3.4 APLICAR MÉTODO *K-MEANS***

Una vez determinados los  $k$  grupos óptimos para segmentar los perfiles de demanda, se incorporará este dato de entrada en la función *K-MEANS*. Esta técnica de agrupamiento identificará los perfiles de demanda predominantes, agrupando los circuitos alimentadores con características de demanda similares y asignándoles etiquetas según esta clasificación. Para una mejor interpretación de los resultados, el método generará gráficos que representen los subgrupos y cada perfil de demanda en relación con su respectivo centroide.

### **3.5 VALIDACIÓN CON *T-SNE***

Se utiliza el método de reducción de dimensionalidad *t-SNE* como herramienta complementaria para validar los clústeres generados con *K-Means*. Dado que los resultados de *t-SNE* dependen significativamente de la configuración de sus hiperparámetros, en particular la perplejidad y la tasa de aprendizaje, se exploran diferentes combinaciones de valores: perplejidades entre 5 y 100, y tasas de aprendizaje entre 10 y 1000. Para cada combinación, se generan gráficos, seleccionando finalmente aquel que muestra una mejor convergencia.

Aplicar *t-SNE* sobre los perfiles de demanda puede facilitar la visualización de las agrupaciones y permite evaluar si el número óptimo de clústeres ha sido

correctamente determinado. El gráfico resultante de *t-SNE* puede revelar si las agrupaciones son claras y bien definidas, lo que indicaría una segmentación adecuada, o si hay puntos ambiguos que sugieren pertenecer a otros clústeres o incluso a un grupo nuevo. Es importante destacar que *t-SNE*, por sí solo, no tiene información sobre las etiquetas de clúster generadas por *K-MEANS*, ya que trabaja únicamente con la estructura de similitudes derivada de los datos.

### 3.6 EXTRACCIÓN DE CARACTERÍSTICAS

Como segunda etapa de la metodología, se incorpora un proceso de extracción de características, basado en estudios previos sobre patrones de consumo de la demanda que proponen técnicas matemáticas para simplificar y representar los perfiles de demanda de manera más eficiente. Este enfoque busca reducir la cantidad de parámetros necesarios para analizar el comportamiento de la demanda, pasando de 24 valores por cada circuito alimentador a un conjunto de características más manejable.

Las características escogidas para ser utilizadas en esta etapa fueron:

- **Desviación estándar:** Mide la variación o dispersión de los valores horarios de la demanda con respecto al promedio. Una mayor desviación estándar indica una mayor variabilidad en el perfil.
- **Promedio:** Es el valor promedio del consumo horario durante las 24 horas. Representa el nivel típico de demanda.

- **Asimetría:** Mide el sesgo o la inclinación de la distribución del perfil de demanda. Un valor positivo indica que el consumo es mayor en las horas más altas del día, mientras que un valor negativo sugiere que el consumo se concentra en horas bajas.
- **Curtosis:** Mide qué tan concentrados están los valores del perfil de demanda alrededor del promedio. Una curtosis alta indica la presencia de picos extremos en el consumo, mientras que una curtosis baja sugiere que los valores están más uniformemente distribuidos, sin consumos extremos significativos.
- **Potencia medio día:** Suma del consumo en el intervalo de 11:00 a 13:00 horas. Representa el consumo característico en las horas centrales del día, asociadas a la transición entre la mañana y la tarde.
- **Potencia alta noche:** Suma del consumo en la madrugada (0:00 a 3:00) y la noche tardía (22:00 a 23:00). Refleja el consumo característico en los extremos nocturnos.
- **Potencia inicio día:** Suma del consumo en el intervalo de 4:00 a 10:00 horas. Representa el consumo característico al inicio de las actividades diarias.
- **Potencia tarde:** Suma del consumo en el intervalo de 14:00 a 18:00 horas. Refleja la demanda típica durante las horas vespertinas.

- **Potencia noche temprana:** Suma del consumo en el intervalo de 19:00 a 21:00 horas. Representa el consumo en las primeras horas de la noche, antes del descenso característico en la actividad.

Estas nueve características fueron seleccionadas para sintetizar la información de los perfiles y facilitar un análisis posterior de agrupamiento. Dado que se manejan múltiples características, se considera útil aplicar *t-SNE* como procedimiento de reducción de dimensionalidad para proyectarlas en un espacio de dos dimensiones. Posteriormente, se sigue el procedimiento descrito en la sección de Determinación del número óptimo de clústeres y Aplicar método *K-MEANS*. Esta es la segunda vez que se aplica *t-SNE* en la metodología: la primera, como se mencionó en la sección 3.5, se utilizó para validación, y en esta etapa se emplea como método de tratamiento del conjunto de datos inicial, lo que permite proceder con la aplicación de *K-MEANS*.

### **3.7 PROGRAMACIÓN EN PYTHON**

La metodología descrita anteriormente se implementó utilizando un script de *Python*, lenguaje que fue seleccionado por su facilidad de uso y amplia adopción en el ámbito científico y técnico. El desarrollo se apoyó principalmente en la biblioteca *Scikit-learn* [23], que proporcionó las herramientas necesarias para llevar a cabo los procesos de normalización de datos, aplicación de *T-SNE* y *K-MEANS*, así como el cálculo de los *CVI*'s. El código desarrollado se incluye como anexo para facilitar su consulta y reproducibilidad.

## 4 PERFILES TÍPICOS DE DEMANDA EN EL SISTEMA DE DISTRIBUCIÓN DE ESSA

En este capítulo se presentan los resultados derivados de la aplicación de la metodología para la caracterización de perfiles de demanda, descrita en el Capítulo 3. Para ello, se empleó un conjunto de datos provenientes del sistema de distribución local de ESSA. En primer lugar, se detalla la descripción de los datos de entrada, y posteriormente se muestran los resultados obtenidos en cada etapa de la metodología.

### 4.1 DESCRIPCIÓN DE CONJUNTO DE DATOS

ESSA opera en un área de influencia que abarca 101 municipios distribuidos en seis departamentos de Colombia, como se detalla en la Tabla 2.

Tabla 2. Clasificación de municipios por departamento del área de influencia ESSA

Departamento	Número de municipios
Santander	87
Norte de Santander	5
Cesar	4
Bolívar	2
Boyacá	2
Antioquia	1
<b>Total</b>	<b>101</b>

Fuente: Base de datos ESSA

En la Figura 4 se presenta un mapa geográfico del área de influencia de ESSA, proporcionando una visión clara de la extensión y alcance del sistema de distribución local.

Figura 4. Área de influencia ESSA



Fuente: Base de datos ESSA

Actualmente, ESSA cuenta con aproximadamente 935824 usuarios, clasificados en las categorías de oficiales, industriales, comerciales y residenciales, cuya distribución se presenta en la Tabla 3.

Tabla 3. Recuento por tipo de usuario

<b>Tipo de usuario</b>	<b>Recuento</b>
Residencial	843777
Industrial	11238
Comercial	74969
Oficial	5840
<b>Total</b>	<b>935824</b>

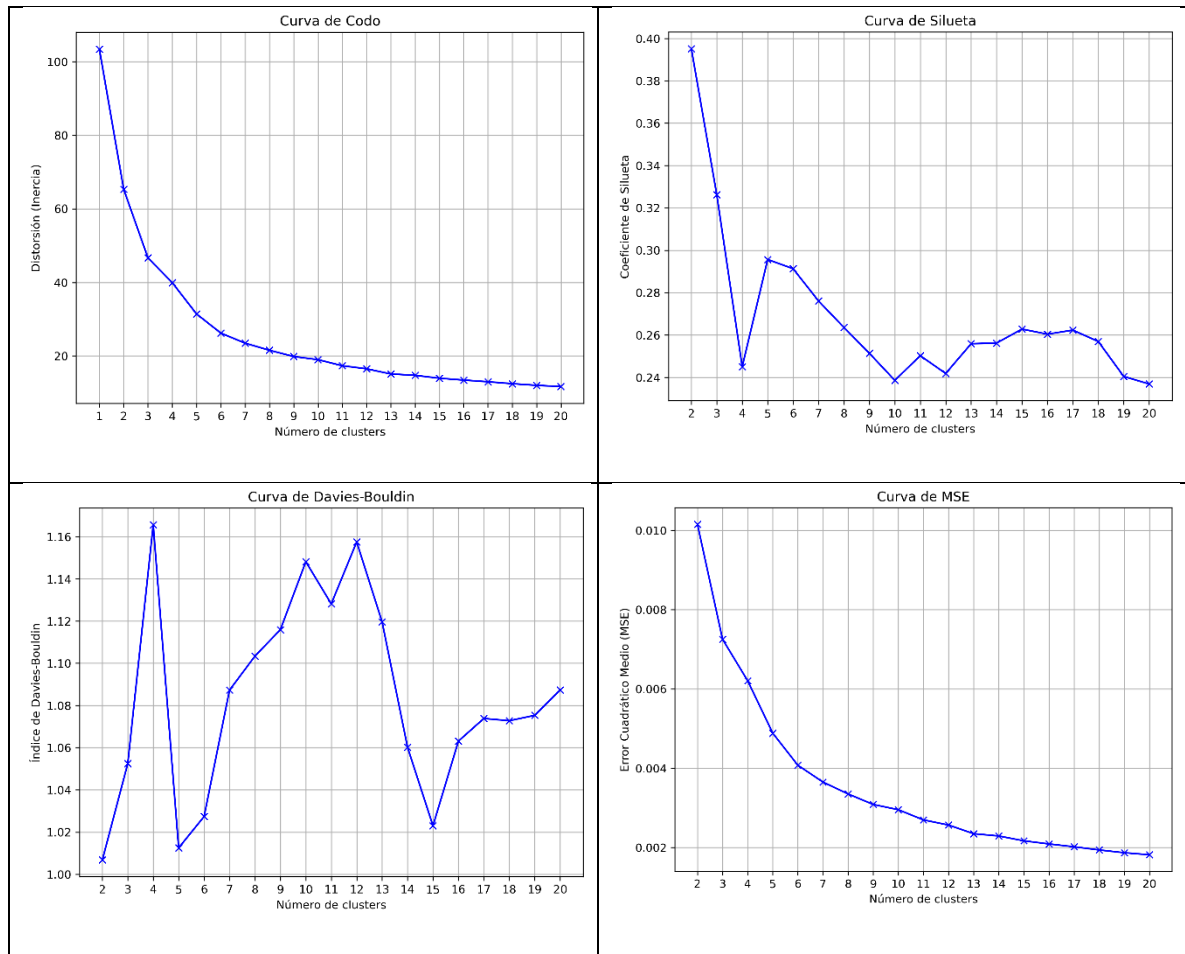
Fuente: Base de datos ESSA

La base de datos utilizada para aplicar en esta metodología proviene de información recopilada a través del sistema SCADA de ESSA. Estos sistemas permiten el monitoreo y la adquisición de datos en tiempo real y posteriormente se almacena en repositorios para su consulta. El conjunto de datos incluye registros detallados de las variables eléctricas de 268 circuitos alimentadores pertenecientes al sistema de distribución local. Cada circuito alimentador puede abastecer desde 60 hasta 15300 usuarios, con un promedio estimado de aproximadamente 3350 usuarios.

#### **4.2 CLÚSTERES ÓPTIMOS**

El conjunto de datos descrito anteriormente se somete a los pasos detallados en la sección 3.3 para determinar el número óptimo de clústeres. Este proceso genera cuatro curvas distintas: la curva de codo, la curva de silueta, la curva de *Davies-Bouldin* y la curva de error cuadrático medio (MSE). Los resultados obtenidos se presentan en la Figura 5 y son descritos posteriormente:

Figura 5. Resultados índices de validez de agrupamiento “CVIS”



Fuente: autores

- **Curva de codo:** La inercia decrece rápidamente entre 2 y 4 clústeres, mostrando un cambio menos pronunciado a partir de 5 clústeres. Esto sugiere que configuraciones entre 4 y 6 clústeres podrían ser un buen balance entre cohesión interna y simplicidad del modelo.
- **Coeficiente de silueta:** Este indicador es mayor en configuraciones bajas (como 2 y 3 clústeres) debido a la alta cohesión de grupos grandes y bien separados. Sin embargo, mantiene valores competitivos en configuraciones

intermedias como 4, 5, y 6 clústeres, lo que indica que estas soluciones aún mantienen buena separación y cohesión.

- **Índice *Davies-Bouldin*:** No presenta un mínimo claro en el rango analizado (2 a 20 clústeres), pero valores moderados como 5 a 6 clústeres ofrecen una calidad razonable sin un aumento significativo en la dispersión ni pérdida de separación. Configuraciones mayores no parecen aportar beneficios relevantes según este indicador.
- **Error cuadrático medio (MSE):** El MSE disminuye considerablemente entre 2 y 4 clústeres, con reducciones más sutiles a partir de 5 clústeres. Esto sugiere que la mayor parte de la variabilidad se explica antes de aumentar excesivamente el número de grupos.

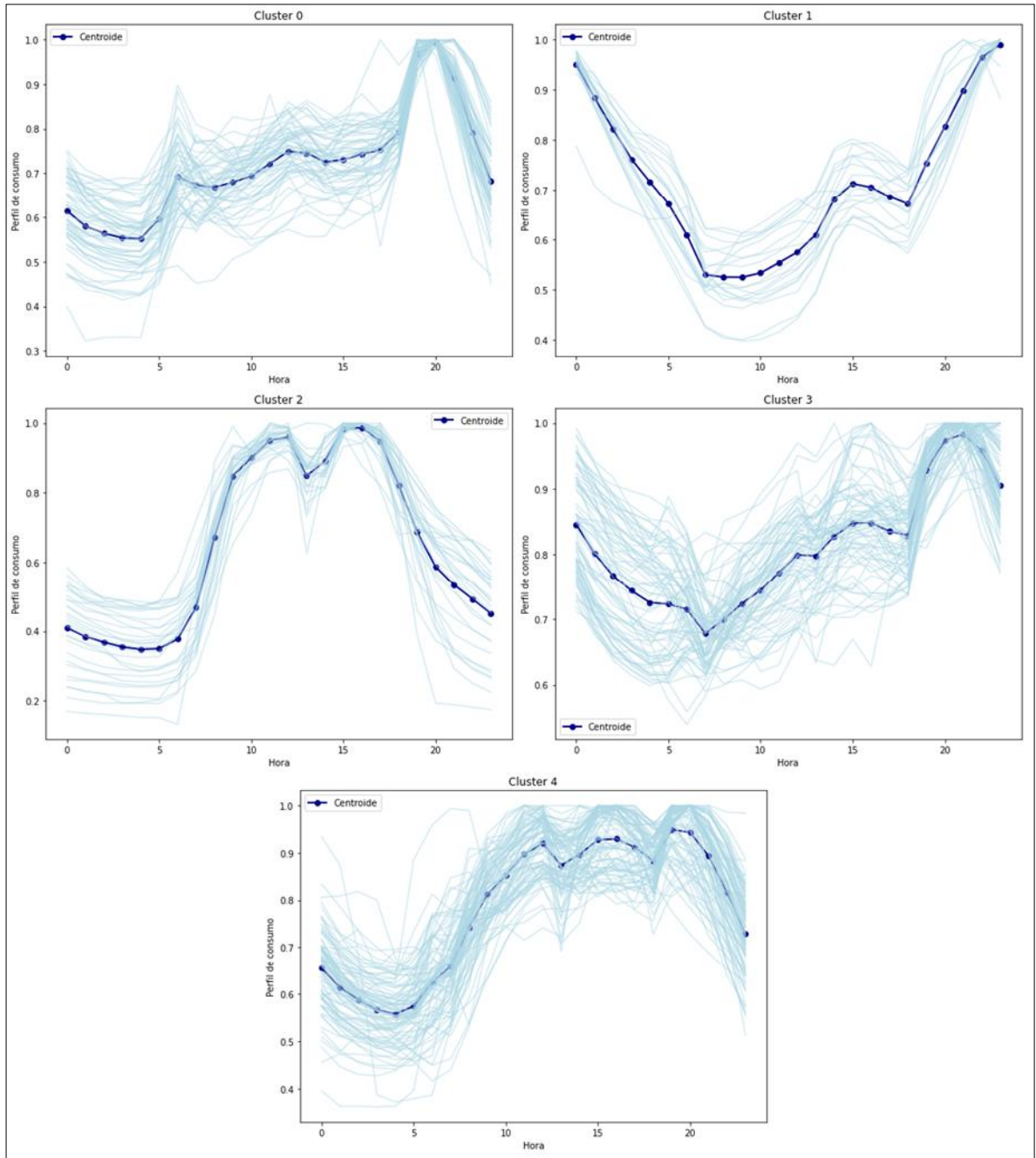
A partir de los resultados obtenidos, se seleccionan 5 clústeres como el número adecuado, ya que representan una solución eficiente que maximiza la calidad del agrupamiento sin incrementar innecesariamente la complejidad del modelo.

### **4.3 APLICACIÓN DE *K-MEANS* PARA OBTENER PERFILES DE DEMANDA PREDOMINANTES**

En la Figura 6 se presentan los típicos perfiles de demanda correspondientes a cada clúster seleccionado. En cada gráfico, se muestran múltiples curvas de demanda superpuestas, junto con una curva central destacada (centroide), que representa el comportamiento promedio o característico del consumo para ese grupo. Esta curva

del centroide sintetiza los patrones de consumo predominantes en el clúster, permitiendo identificar tendencias comunes y variaciones dentro del grupo.

Figura 6. Perfiles típicos predominantes de demanda por clúster



Fuente: Autores

A continuación, se describe un análisis más detallado para cada clúster, observando por periodos cómo cambia el consumo y sugiriendo posibles interpretaciones en base a patrones de carga típicos.

**Clúster 0:** El patrón de consumo inicia en un nivel moderado durante la medianoche y desciende gradualmente, alcanzando su valor mínimo alrededor de las 4:00 horas. Este comportamiento es característico de la baja actividad operativa típica de la madrugada, con procesos limitados o en pausa.

A partir de las 4:00 horas, el consumo muestra un incremento acelerado hasta las 6:00 horas. Posteriormente, la curva se estabiliza manteniendo una tendencia positiva durante el resto de la mañana. Entre las 12:00 y las 14:00 horas, se observa un ligero cambio en esta tendencia, posiblemente asociado a pausas o descansos para almorzar. Sin embargo, después de este período, el consumo retoma su incremento hasta las 18:00 horas.

A partir de las 18:00 horas, la curva registra un aumento significativo que culmina en el valor máximo del día alrededor de las 20:00 horas, reflejando un periodo de máxima actividad energética. Finalmente, tras alcanzar su punto máximo, el consumo comienza a descender de manera progresiva hasta finalizar el día.

**Clúster 1:** A diferencia de los demás clústeres, este patrón de consumo se caracteriza por iniciar en un nivel alto durante la medianoche, mostrando una disminución constante y progresiva hasta alcanzar su valor mínimo alrededor de las

8:00 horas. Este comportamiento sugiere una actividad nocturna que va reduciéndose gradualmente a medida que avanza la madrugada, posiblemente por el uso de cargas relacionadas con el confort térmico, tales como ventiladores y aires acondicionados.

A partir de las 8:00 horas, la curva presenta un cambio notable, con un incremento progresivo que se extiende hasta las 15:00 horas. Este aumento gradual podría estar relacionado con la reactivación de actividades operativas o productivas que se intensifican durante las primeras horas del día y también debido a aumento de la temperatura.

Entre las 15:00 y las 18:00 horas, se observa una ligera disminución en el consumo, lo que podría reflejar una pausa temporal o cese de actividades. Sin embargo, a partir de las 18:00 horas, la tendencia cambia excesivamente, mostrando un marcado incremento que culmina en el valor máximo del día hacia las 22:00 horas. Este pico nocturno sugiere un período de alta actividad energética, posiblemente asociado con las actividades domésticas típicas de las horas de la noche.

**Clúster 2:** Entre las 00:00 horas y las 04:00 horas, el consumo se mantiene relativamente bajo y estable, representando el valor mínimo de toda la curva y reflejando la baja actividad típica de la madrugada. A partir de las 4 hasta las 12 horas, se observa un aumento significativo en la demanda, lo que sugiere el inicio de actividades o la apertura de operaciones. Entre las 12 y las 13 horas, la curva presenta una leve disminución en el consumo, posiblemente asociado a una pausa

o descanso en las actividades productivas, como el almuerzo. Sin embargo, entre las 13 y las 16 horas, la demanda vuelve a incrementarse de manera considerable, alcanzando el valor máximo del día, lo que puede reflejar una intensificación de las actividades en esta franja horaria. Finalmente, a partir de las 16:00 horas, la curva comienza a descender gradualmente, sugiriendo una fase de cierre o reducción de las operaciones conforme se aproxima la noche.

**Clúster 3:** Entre las 0:00 y las 7:00 horas, el consumo empieza alto a medianoche y disminuye progresivamente hasta alcanzar su mínimo. Este descenso refleja una reducción en la actividad, como terminación de procesos nocturnos. A partir de las 7:00 horas, el consumo aumenta gradualmente hasta las 15:00 horas, lo que sugiere un periodo de preparación o un incremento en las actividades operativas.

Entre las 15:00 y las 18:00 horas, el consumo disminuye ligeramente, posiblemente debido a la terminación de actividades diurnas. Entre las 18:00 y las 21:00 horas, el consumo asciende hasta alcanzar su máximo, y es el momento donde se encuentra la mayor actividad energética posiblemente asociada con las actividades domésticas típicas de las horas de la noche.

Después de las 21:00 horas, el consumo comienza a disminuir gradualmente, reflejando una reducción en las actividades operativas hacia el final del día.

**Clúster 4:** Entre las 0:00 y las 4:00 horas, el consumo comienza en un nivel moderado y desciende de forma progresiva, alcanzando su valor mínimo alrededor

de las 4:00 horas. Este comportamiento refleja la baja actividad operativa característica de la madrugada. A partir de las 4:00 horas, el consumo empieza a incrementarse de manera gradual y sostenida hasta las 12:00 horas, lo que podría estar relacionado con el inicio y desarrollo de actividades operativas, típico de un período de producción creciente.

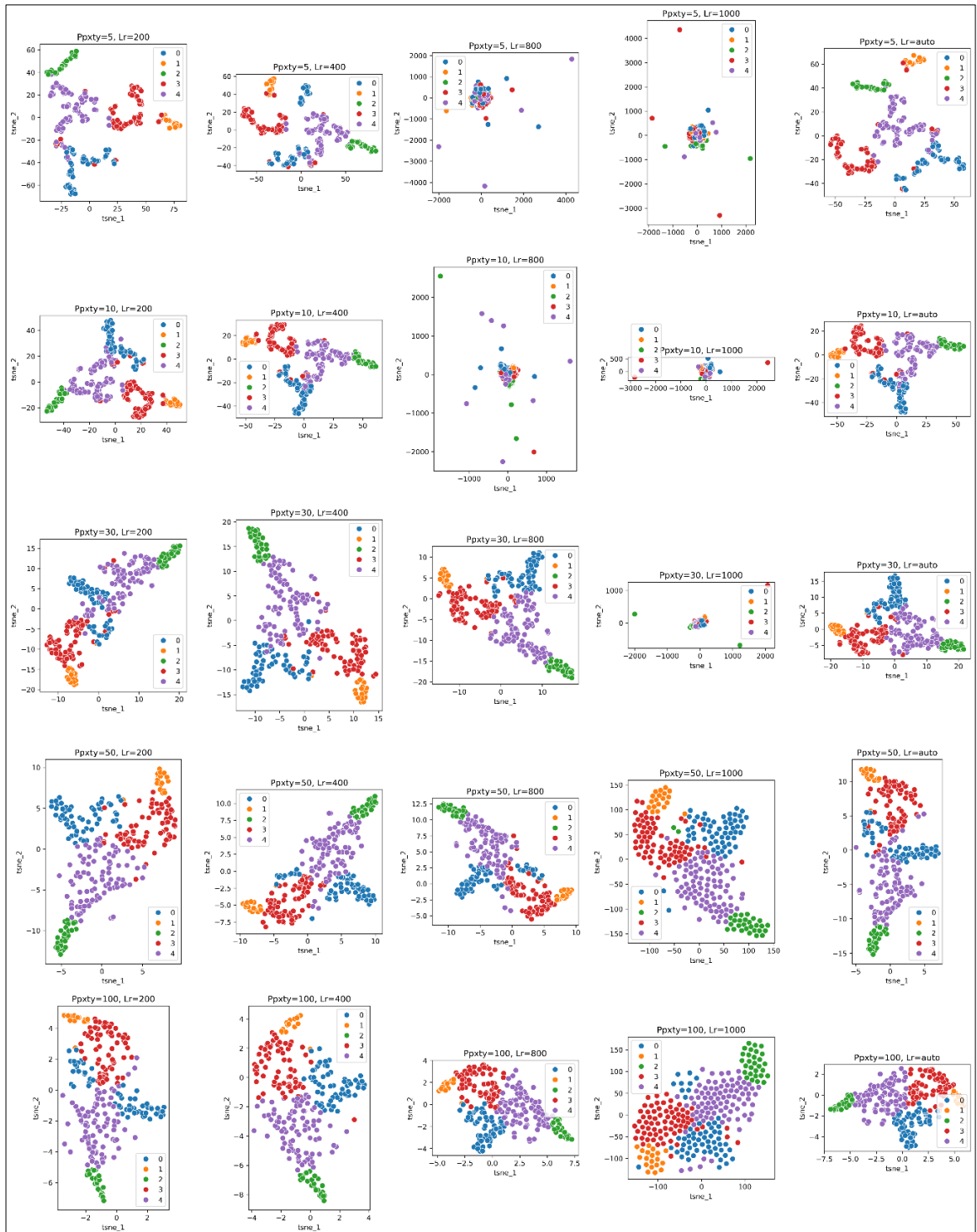
Entre las 12:00 y las 19:00 horas, el consumo permanece relativamente estable, con algunas fluctuaciones leves, incluyendo disminuciones puntuales a las 13:00 y las 18:00 horas. Estas reducciones podrían estar asociadas a pausas en las actividades, como el tiempo de almuerzo y la posible finalización de operaciones, respectivamente. Posteriormente, entre las 19:00 y las 20:00 horas, se observa un repunte significativo en la demanda, alcanzando el valor máximo del día. Finalmente, a partir de las 20:00 horas, el consumo comienza a descender de manera progresiva hasta la medianoche, indicando una reducción en las operaciones hacia el cierre del día.

#### **4.4 VALIDACIÓN CON *T-SNE***

Finalmente, se lleva a cabo una validación adicional para complementar y confirmar que la selección de cinco clústeres fue la adecuada. Para esto, se continúa con el proceso descrito en la sección 3.5. Como resultado, se genera un gráfico compuesto por 25 subgráficos que muestran los resultados de aplicar el algoritmo *t-SNE* a los perfiles de demanda, utilizando diferentes valores de perplejidad y tasa de aprendizaje. Para facilitar la visualización de los clústeres, se asignaron colores

específicos: 0 (azul), 1 (naranja), 2 (verde), 3 (rojo) y 4 (púrpura). Este esquema de colores se mantendrá consistentemente a lo largo de esta monografía. Los resultados obtenidos para el conjunto de datos analizado se presentan en la Figura 7.

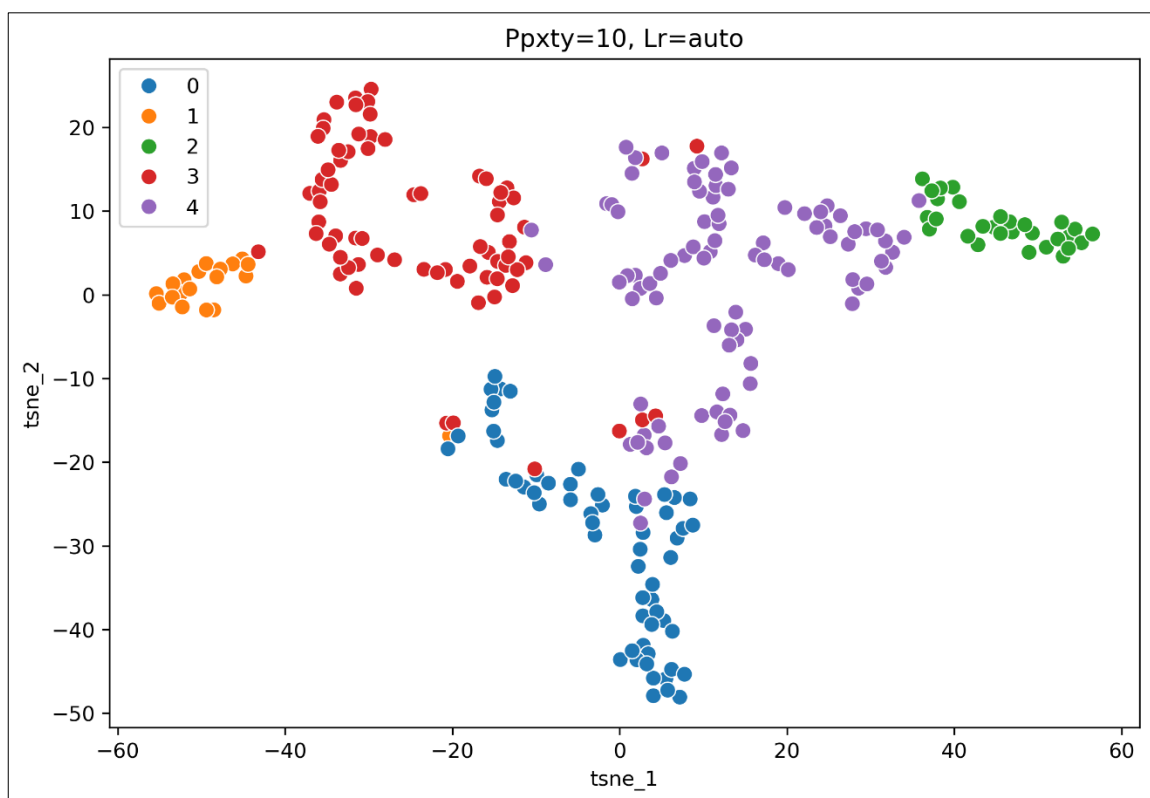
Figura 7. Resultados validación *T-SNE*



Fuente: Autores

El arreglo con una perplejidad de 10 y una tasa de aprendizaje ajustada automáticamente fue seleccionado como el de mejor convergencia tras una inspección visual de las gráficas. A continuación, se presentan en detalle los resultados obtenidos con este arreglo.

Figura 8. *T-SNE* perplejidad 10 y tasa de aprendizaje automática



Fuente: Autores

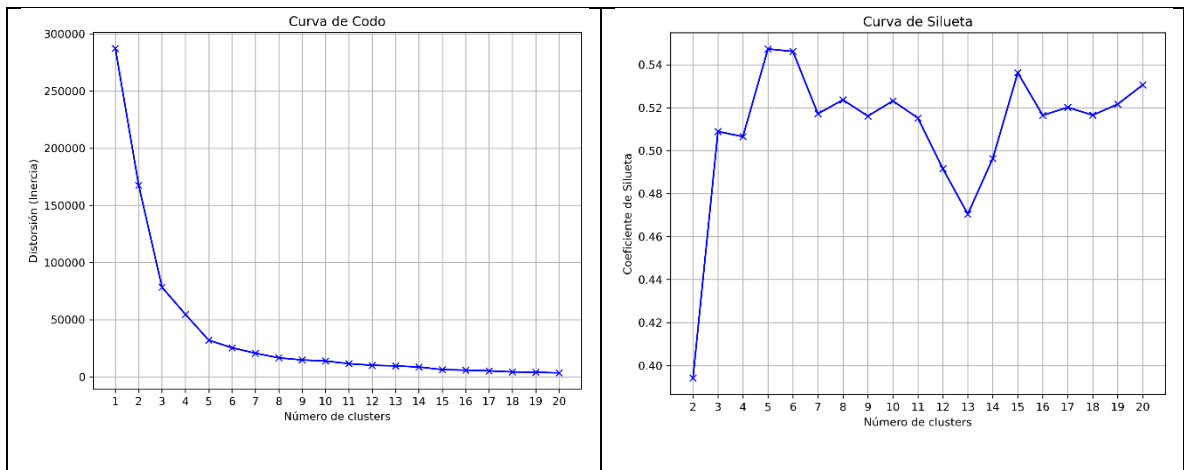
En la Figura 8 se puede observar que, aunque el método de reducción de dimensionalidad *t-SNE* no tiene conocimiento previo sobre los clústeres seleccionados, se logra una separación clara entre ellos. Sin embargo, se identifican posibles mejoras en la separación de los clústeres 0 y 4, así como entre los clústeres

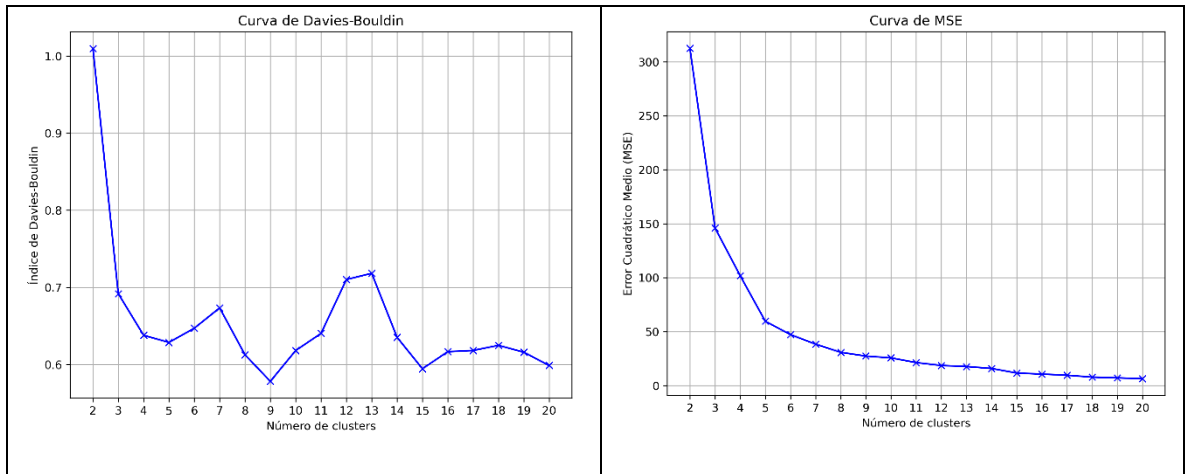
0 y 3. Por otro lado, los clústeres 1 y 2 parecen ser bastante distintos entre sí, sin mostrar superposición.

#### 4.5 AGRUPAMIENTO POR CARACTERÍSTICAS

Como se explicó en la sección 3.6, esta etapa tiene como objetivo agrupar los perfiles de demanda mediante la extracción de características específicas de cada perfil. Las características calculadas se procesan utilizando *t-SNE*, y a partir de la matriz resultante se aplica el algoritmo *K-MEANS*. Con base en este procedimiento, se generan nuevamente gráficos de índices de validación de agrupamiento. A continuación, se presentan los resultados obtenidos.

Figura 9. Resultados índices de validez de agrupamiento “CVI's”

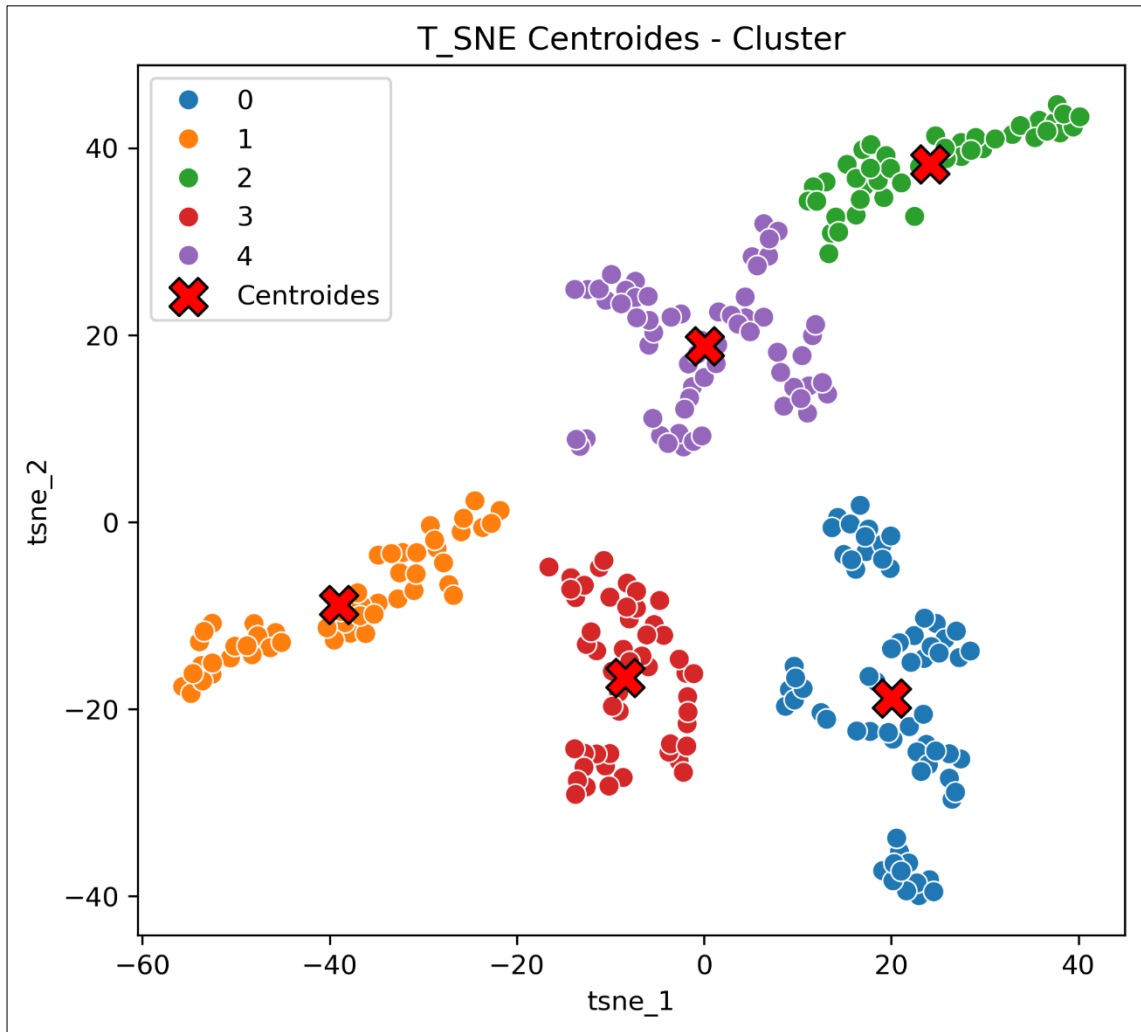




Fuente: Autores

En la Figura 9 se observa que, en las cuatro métricas analizadas, el número adecuado de clústeres vuelve a ser cinco. Esto se evidencia en que: la curva del codo se estabiliza en ese punto, la curva de silueta alcanza su valor máximo, se identifica como uno de los mínimos locales en la curva de *Davies-Bouldin*, y los deltas en la curva MSE muestran una disminución considerable. Con el número adecuado de clústeres seleccionado, se aplica el algoritmo *K-MEANS* a la matriz generada por *t-SNE*. Los resultados se presentan en el siguiente gráfico de dispersión:

Figura 10. T-SNE centroides - clúster

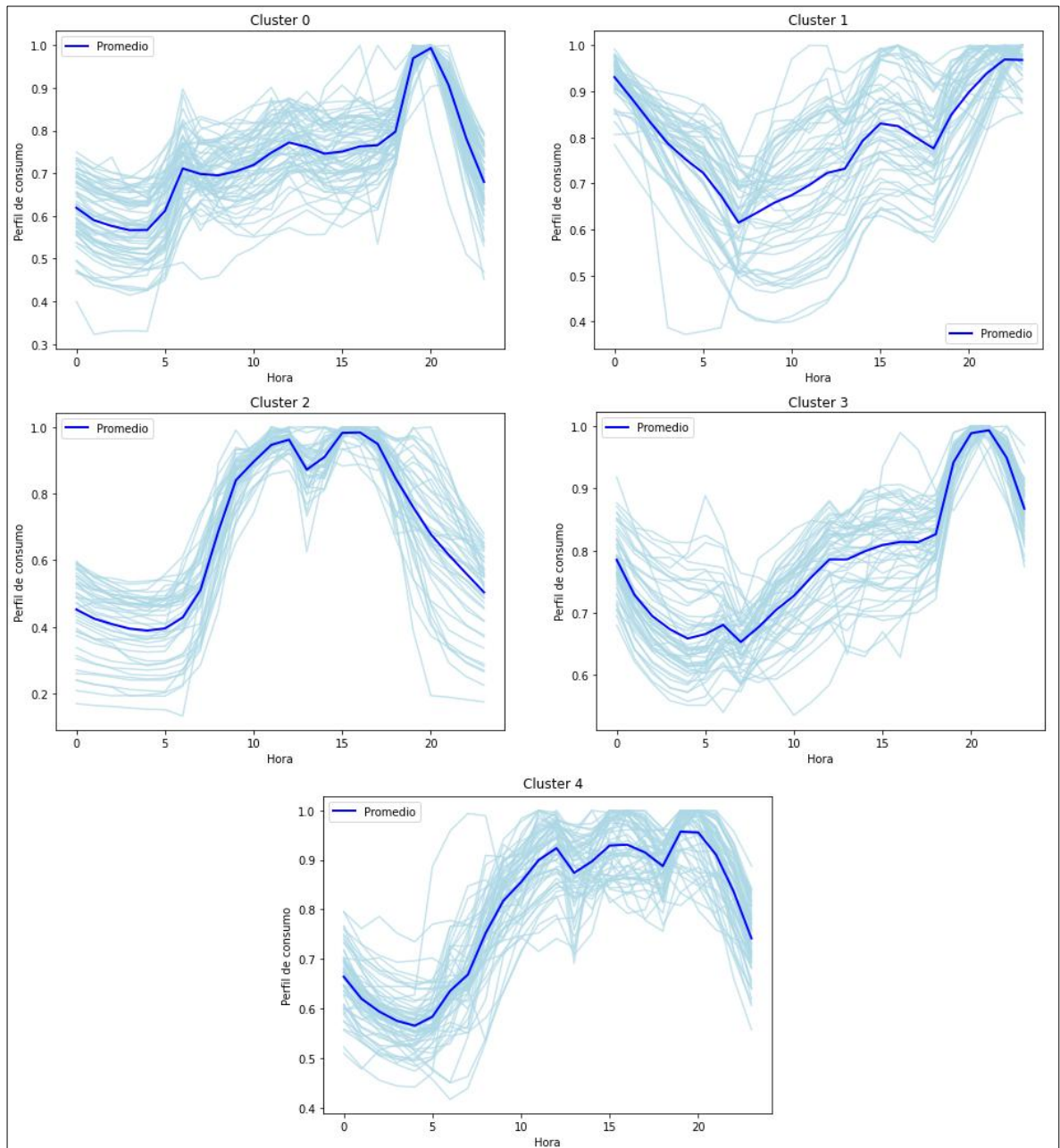


Fuente: autores

En la Figura 10 se observa que los clústeres están claramente agrupados y diferenciados utilizando la escala de colores seleccionada. La marca en forma de equis roja indica el centroide de cada clúster. No se detecta superposición entre los datos, lo que sugiere que las características elegidas reflejan comportamientos distintos en los patrones de consumo.

Los perfiles de demanda predominantes utilizando esta etapa paralela se muestran en la Figura 11.

Figura 11. Perfiles típicos predominantes de demanda por clúster



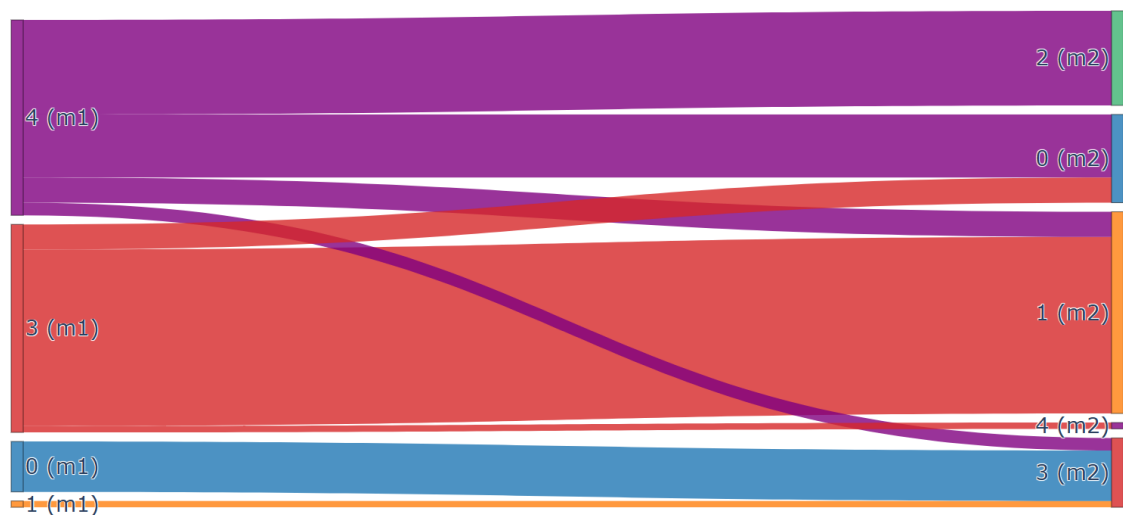
Fuente: Autores

#### 4.6 COMPARACIÓN DE LAS DOS ETAPAS

Al comparar las Figura 6 y Figura 11, se observa que los centroides o curvas predominantes de cada clúster son prácticamente idénticos en ambas etapas. Sin embargo, no todos los circuitos fueron asignados al mismo clúster en los dos métodos. Esto podría atribuirse a que los perfiles de demanda de algunos circuitos presentan características que los hacen similares a más de un patrón predominante de consumo.

El análisis revela que 73 circuitos alimentadores cambiaron de clúster entre las metodologías, representando el 27% del total. Este cambio se ilustra en el diagrama de *Sankey* de la Figura 12, que detalla las transiciones de los clústeres de la metodología 1 (m1) a los clústeres de la metodología 2 (m2). Un porcentaje de cambio más elevado podría señalar la necesidad de ajustar la segmentación actual, ya sea mediante la creación de un nuevo clúster o la consolidación de los existentes.

Figura 12. Diagrama de *Sankey*: Transición de clústeres entre metodologías



Fuente: Autores

En la Figura 12 se observa un mayor flujo del clúster tipo 3 del primer método al clúster tipo 1 del segundo método. Al analizar los patrones de consumo predominantes de cada caso (Figura 6 y Figura 11), se puede notar una similitud aparente entre estos clústeres, aunque con algunas diferencias. Esta comparación explica las razones detrás de los cambios observados.

## 5 RELACIÓN DE PERFILES TÍPICOS DE DEMANDA Y CARACTERÍSTICAS DE LOS USUARIOS

En este capítulo, se analiza la relación entre los patrones típicos de consumo de energía, identificados y descritos en la sección 4.4, y las características de los usuarios. El análisis explora la influencia de la ubicación geográfica y la clasificación de los usuarios (residencial, comercial o industrial) en dichos patrones. Para ello, se incluye una descripción estadística detallada de cada clúster identificado. Para simplificar el análisis, se consideraron únicamente los resultados obtenidos con el primer método, que aplica *K-MEANS* directamente sobre los perfiles de demanda, dejando a un lado el enfoque basado en el agrupamiento por características.

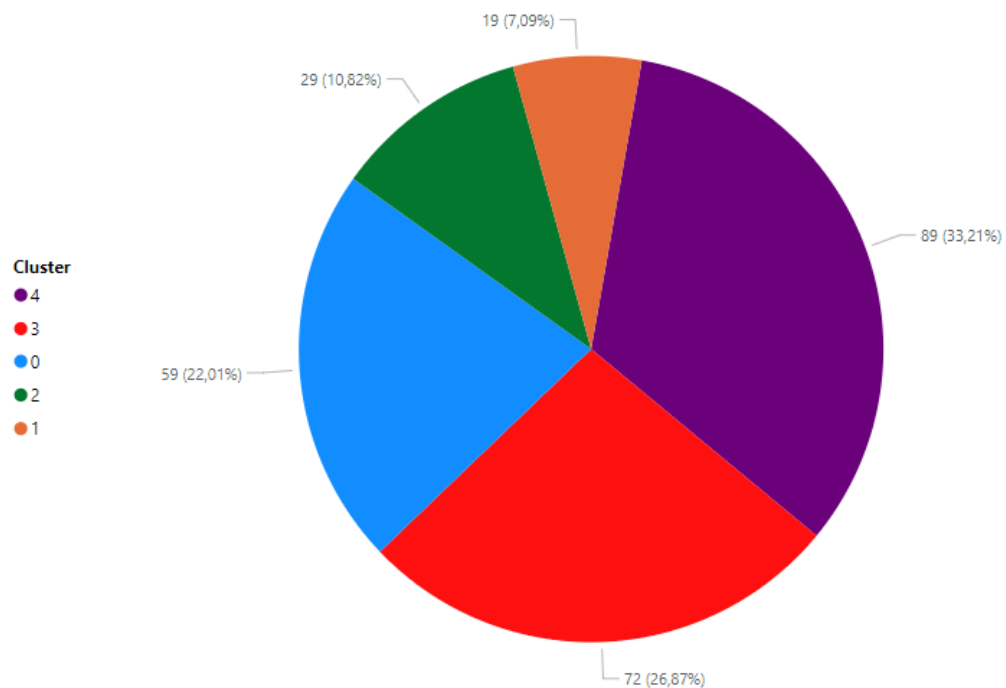
### 5.1 DESCRIPCIÓN DE LOS CLÚSTERES EN RELACIÓN CON LOS CIRCUITOS

En la Figura 13 se ilustra la distribución porcentual de los circuitos alimentadores analizados en función de los clústeres de patrones de consumo identificados. Cada segmento representa la proporción de circuitos asociados a un clúster específico, destacando las siguientes observaciones:

- **Clúster 4:** Representa el mayor porcentaje, con un 33.21% del total de circuitos analizados, lo que sugiere que este patrón de consumo es el más predominante en el sistema.
- **Clúster 3:** Agrupa el 26.87% de los circuitos, siendo el segundo más común.

- **Clúster 0:** Incluye el 22.01% de los circuitos, ocupando el tercer lugar en representación.
- **Clúster 2:** Tiene una menor participación, con el 10.82% de los circuitos.
- **Clúster 1:** Es el menos representado, con solo el 7.09% del total.

Figura 13. Clústeres por circuitos

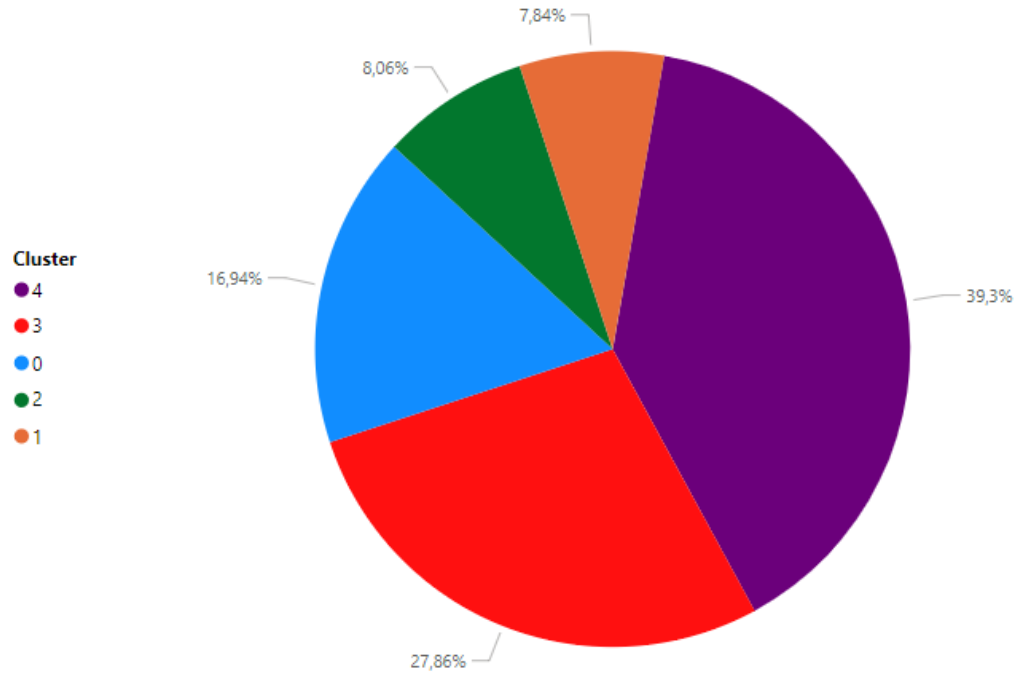


Fuente: Autores

Otra perspectiva de análisis es considerar un nivel más desagregado, enfocado en la cantidad de usuarios. Este enfoque se utiliza principalmente para evitar sesgos en la información, ya que algunos circuitos alimentadores pueden tener un mayor número de usuarios que otros. La Figura 14 muestra la distribución porcentual de

los usuarios analizados en función de los clústeres de patrones de consumo identificados.

Figura 14. Clústeres por cantidad de usuarios



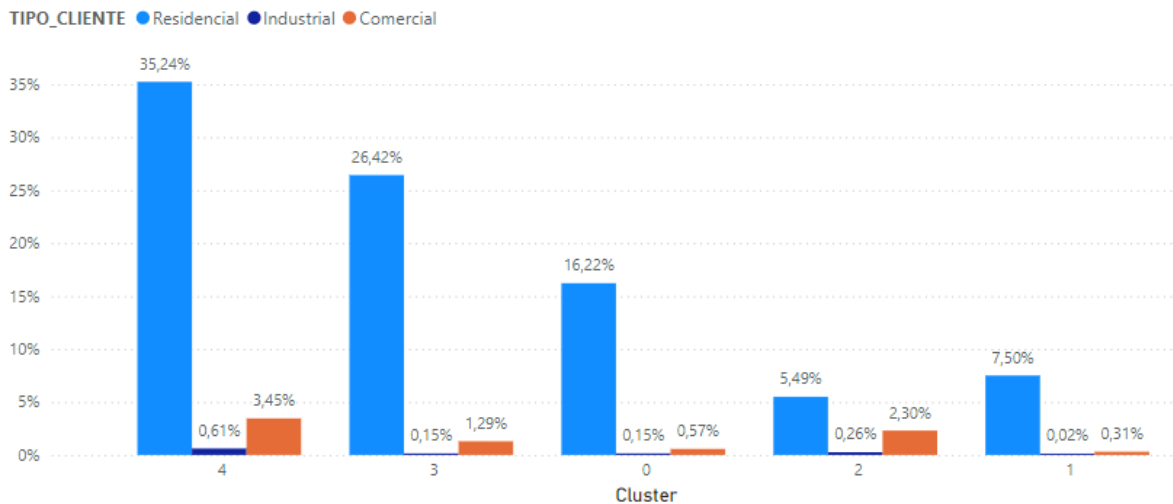
Fuente: Autores

Se observa que, en ambos casos, el orden de la proporción de los clústeres no se ve afectado, y los porcentajes se mantienen prácticamente iguales. Esto sugiere que las variaciones en la cantidad de usuarios entre circuitos podrían no tener un impacto significativo en la clasificación basada en usuarios.

## 5.2 CLASIFICACIÓN DE LOS TIPOS DE USUARIO POR CIRCUITO

La Figura 15 presenta, mediante una gráfica de barras, la distribución porcentual de la clasificación de los usuarios (residencial, comercial e industrial) en cada clúster. Los usuarios residenciales predominan en todos los clústeres, con mayor representación en el Clúster 4 (35.24%) y el Clúster 3 (26.42%). Los usuarios comerciales, que tienen una menor participación, destacan en el Clúster 4 (3.45%) y el Clúster 2 (2.30%). Por último, los usuarios industriales que tienen una presencia mínima en todos los clústeres, con valores ligeramente más altos en el Clúster 2 (2.30%) y el Clúster 4 (0.61%).

Figura 15. Clústeres por tipos de usuarios



Fuente: Autores

La distribución de los tipos de usuario por clúster influye significativamente en los patrones de consumo presentados en la Figura 15. Aunque todos los clústeres incluyen los tres tipos de usuarios (residencial, comercial e industrial), predominan

ciertas tendencias que determinan el comportamiento específico de cada patrón de consumo.

En el caso del clúster 2, la mayor proporción de usuarios comerciales e industriales genera picos de demanda durante la mañana y la tarde, reflejando el impacto de las actividades propias de los horarios laborales. De manera similar, el clúster 4 muestra un comportamiento comparable. Sin embargo, la alta presencia de usuarios residenciales introduce un aumento adicional en el consumo al inicio de la noche, característico de este tipo de clientes.

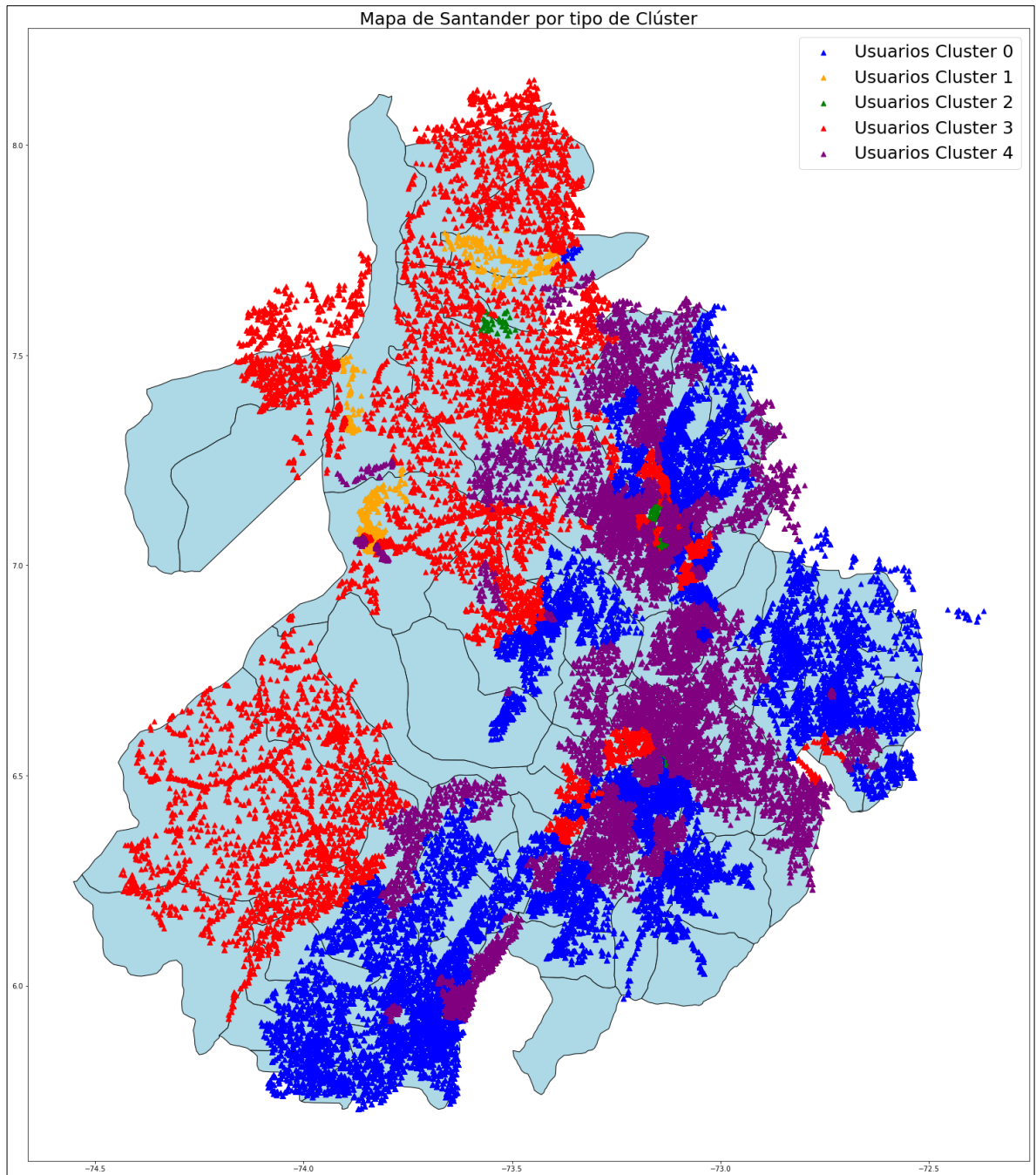
Por otro lado, los clústeres 0 y 3 representan perfiles típicos de usuarios residenciales, con consumos bajos y estables durante el día y picos significativos al caer la noche. A diferencia de estos, el clúster 1, aunque también dominado por usuarios residenciales, presenta un comportamiento particular con mayores niveles de consumo durante la noche, posiblemente influenciado por necesidades específicas de confort o hábitos nocturnos de uso energético

### **5.3 UBICACIÓN GEOGRÁFICA DE LOS USUARIOS EN RELACIÓN CON LOS CLÚSTERES**

La ubicación geográfica de los usuarios puede influir notablemente en los patrones de consumo diario, debido a factores como el clima, la actividad económica y las características demográficas, que varían según la región, municipio o localidad. Por ello, esta sección analiza la distribución geográfica de los usuarios de ESSA dentro del área de influencia descrita en la sección 4.1, en relación con los patrones de

consumo identificados en la sección 4.4. Para ello, se diseñó un mapa político que delimita los contornos municipales y utiliza puntos de colores para representar el tipo de clúster al que pertenecen los transformadores de distribución asociados a cada circuito alimentador. Este mapa se presenta en la Figura 16:

Figura 16. Mapa de Santander de transformadores de distribución por tipo de clúster



Fuente: Autores

En la figura 16 se observa una aparente separación geográfica de los clústeres. En la zona occidental predominan los clústeres 3 (rojo) y 1 (amarillo), mientras que en la zona oriental se localizan los clústeres 0 (azul) y 4 (morado). También se aprecian algunos puntos dispersos de estos clústeres a lo largo de todo el mapa. Esta distribución podría explicarse por las diferencias geográficas y climáticas de la región: en el occidente se encuentra el Valle del Magdalena Medio, caracterizado por su clima cálido, mientras que en el oriente predominan áreas montañosas y de páramo, donde el clima es más frío. En los principales municipios, como Bucaramanga, se observa la presencia de 4 de los 5 clústeres, mientras que en Barrancabermeja se identifican 3. Esto podría indicar que, en una población urbana más grande, se presentan diversos comportamientos de consumo.

## 6 CONCLUSIONES

Se diseñó una metodología capaz de caracterizar los patrones de consumo de energía eléctrica en cualquier sistema de distribución que cuente con registros de medición provenientes del sistema *SCADA*. Esta metodología, que incorpora técnicas como el escalamiento de datos, el algoritmo *K-MEANS* y el método *t-SNE*, identifica y clasifica los perfiles típicos de demanda de un conjunto de entrada según sus similitudes o características. Los resultados obtenidos pueden ser una herramienta valiosa para aplicaciones como la mejora de la distribución de energía, la optimización de los pronósticos de demanda o el diseño de tarifas preferenciales ajustadas a las necesidades y comportamientos de consumo de los usuarios. Esta metodología, al ser desarrollada en el lenguaje de programación *Python*, ofrece una implementación sencilla y de alta adaptabilidad.

La aplicación de esta metodología al conjunto de datos de los circuitos alimentadores de ESSA ha permitido identificar cinco grupos principales de perfiles típicos de demanda. Este número se considera adecuado, respaldado por diversos índices de validación. Los clústeres reflejan la diversidad en los patrones de consumo de los usuarios, destacando comportamientos característicos en distintos períodos del día. Los clústeres 0 y 3, mayoritariamente residenciales, presentan picos de consumo nocturnos asociados a actividades domésticas, con comportamientos estables durante el día. El clúster 1 se caracteriza por un alto consumo nocturno y de madrugada, posiblemente vinculado a necesidades de confort térmico o hábitos nocturnos. El clúster 2, relacionado con usuarios

comerciales y productivos, muestra picos de demanda en la mañana y la tarde, reflejando horarios laborales típicos. Por su parte, el clúster 4 combina características residenciales y comerciales, destacándose por un consumo diversificado con picos tanto en la mañana como en la noche. Estos resultados ofrecen una base para optimizar la gestión de la demanda energética, adaptándose a las particularidades de cada patrón identificado.

La identificación de clústeres destaca la predominancia del Clúster 4 en términos de circuitos y usuarios, confirmando su importancia dentro del sistema. Cada clúster se define según la proporción de usuarios residenciales, comerciales e industriales. Los clústeres con una mayor participación de usuarios residenciales (Clústeres 0, 1, 3 y 4) exhiben patrones característicos de picos nocturnos, mientras que aquellos dominados por usuarios comerciales e industriales (Clústeres 2 y 4) presentan picos durante horarios laborales. Este análisis revela que las actividades específicas de los usuarios tienen una influencia directa en los patrones de consumo. La distribución geográfica de los clústeres muestra una correlación con las condiciones climáticas y geográficas de la región. Por ejemplo, los clústeres 3 y 1 predominan en municipios del Valle del Magdalena Medio, caracterizados por un clima cálido, mientras que los clústeres 0 y 4 son más comunes en zonas de mayor altitud, montañosas y de páramo del oriente de Santander. Estos hallazgos resaltan cómo factores externos, como el clima y la topografía, impactan los patrones de consumo energético.

## 7 RECOMENDACIONES

- Se recomienda a ESSA capturar y almacenar los patrones de consumo identificados en su sistema. Esta práctica proporcionará información valiosa para estudios sobre los cambios en los patrones de consumo a lo largo del tiempo, los cuales pueden verse influenciados por factores como la aparición de nuevas tecnologías, diferentes condiciones sociales y económicas, y la integración de fuentes de energía renovables no convencionales, entre otros.
- Se sugiere la aplicación de esta metodología con mediciones más desagregadas, como las obtenidas a través de transformadores de distribución o los medidores de los usuarios. Este enfoque podría facilitar la identificación de nuevos perfiles típicos de demanda, mejorando así la comprensión de los hábitos de consumo de los usuarios.
- Existen diversos algoritmos de agrupamiento innovadores, como *K-Medoid*, Agrupamiento Jerárquico, *DBSCAN*, entre otros, que podrían mejorar los resultados obtenidos con *K-Means*. Por esta razón, se recomienda, como complemento a este estudio, aplicar estas metodologías adicionales para explorar diferentes perspectivas, con el fin de encontrar cual algoritmo de agrupamiento se ajusta mejor a la caracterización de demanda.
- Se recomienda complementar el método de agrupamiento por características con un algoritmo de optimización, con el objetivo de seleccionar el conjunto de características que, utilizando el menor número de métricas posibles,

proporcione la información suficiente para realizar agrupamientos adecuados de los tipos de consumo.

## BIBLIOGRAFÍA

- [1] Y.-I. Kim, J.-M. Ko, and S.-H. Choi, *Methods for Generating TLPs (Typical Load Profiles) for Smart Grid-Based Energy Programs*. IEEE, 2011.
- [2] S. A. Azad, A. B. M. S. Ali, and P. Wolfs, "Identification of typical load profiles using K-means clustering algorithm," in *Asia-Pacific World Congress on Computer Science and Engineering, APWC on CSE 2014*, Institute of Electrical and Electronics Engineers Inc., 2014. doi: 10.1109/APWCCSE.2014.7053855.
- [3] A. Grandjean, J. Adnot, and G. Binet, "A review and an analysis of the residential electric load curve models," Apr. 2012. doi: 10.1016/j.rser.2012.08.013.
- [4] B. Akperi and P. Matthews, "Analysis of Clustering Techniques on Load Profiles for Electrical Distribution," 2014.
- [5] M. M. Navarro, M. N. Young, Y. T. Prasetyo, and R. Nadlifatin, "Customer Load Profile Clustering Using K-means Algorithm: A Case Study in an Electric Distribution Company in the Philippines Amidst the COVID-19 Pandemic," in *IEEE International Conference on Industrial Engineering and Engineering Management*, IEEE Computer Society, 2022, pp. 625–629. doi: 10.1109/IEEM55944.2022.9989809.
- [6] X. S. Guerrero and G. E. Escrivá, "Caracterización de la demanda de energía mediante patrones estocásticos en las Redes Eléctricas Inteligentes," 2020.

- [7] D. Fischer, A. Härtl, and B. Wille-Hausmann, "Model for electric load profiles with high time resolution for German households," *Energy Build*, vol. 92, pp. 170–179, Apr. 2015, doi: 10.1016/j.enbuild.2015.01.058.
- [8] G. Chicco, "Overview and performance assessment of the clustering methods for electrical load pattern grouping," *Energy*, vol. 42, no. 1, pp. 68–80, 2012, doi: 10.1016/j.energy.2011.12.031.
- [9] G. Chicco, R. Napoli, and F. Piglione, "Comparisons among clustering techniques for electricity customer classification," *IEEE Transactions on Power Systems*, vol. 21, no. 2, pp. 933–940, Apr. 2006, doi: 10.1109/TPWRS.2006.873122.
- [10] M. R. R. Aravena, "Elaboración de perfiles de demanda a nivel de distribución para efectos de consideración en planificación energética de largo plazo," UNIVERSIDAD DE CHILE, Santiago de Chile, 2019.
- [11] M. Mora-Alvarez, P. Contreras-Ortiz, X. Serrano-Guerrero, and G. Escrivá-Escriva, "Characterization and Classification of Daily Electricity Consumption Profiles: Shape Factors and k-Means Clustering Technique," 2018, doi: 10.1051/e3sconf/2018.
- [12] L. F. G. Gómez, "DISEÑO E IMPLEMENTACION DE UNA TÉCNICA PARA LA CONSTRUCCIÓN DE LAS CURVAS DE DEMANDA MÁXIMA DIVERSIFICADA PARA EMPRESAS DEL SECTOR ELÉCTRICO COLOMBIANO," 2010.

- [13] K. A. H. Hernández and J. S. C. Cruz, “ANÁLISIS DE LA CURVA DE DEMANDA ELÉCTRICA PARA USUARIOS RESIDENCIALES ESTRATO 4 EN LA CIUDAD DE BOGOTÁ ANTE DIFERENTES ESCENARIOS DE LOS HÁBITOS DE CONSUMO,” 2017.
- [14] I. P. Panapakidis, M. C. Alexiadis, and G. K. Papagiannis, “Electricity customer characterization based on different representative load curves,” in *9th International Conference on the European Energy Market, EEM 12*, 2012. doi: 10.1109/EEM.2012.6254761.
- [15] S. R. Castaño, *Redes de Distribución de Energía*, Tercera Edición. Manizales: UNIVERSIDAD NACIONAL DE COLOMBIA SEDE MANIZALES, 2004.
- [16] B. Akperi and P. Matthews, “Analysis of Customer Profiles on an Electrical Distribution Network,” 2014. doi: 10.1109/UPEC.2014.6934624.
- [17] A. Rajabi, M. Eskandari, M. J. Ghadi, L. Li, J. Zhang, and P. Siano, “A comparative study of clustering techniques for electrical load pattern segmentation,” *Renewable and Sustainable Energy Reviews*, vol. 120, Mar. 2020, doi: 10.1016/j.rser.2019.109628.
- [18] Á. Gonzalo, “Segmentación utilizando K-means en Python,” 2019. [Online]. Available: <https://machinelearningparatodos.com/segmentacion-utilizando-k-means-en-python/>
- [19] A. Barrios, “Tutorial del Algoritmo K-Means en Python | by Anthony Barrios | LatinXinAI | Medium,” 2023. [Online]. Available:

<https://medium.com/latinxinai/tutorial-del-algoritmo-k-means-en-python-d8055751e2f3>

- [20] G. E. Okereke, M. C. Bali, C. N. Okwueze, E. C. Ukekwe, S. C. Echezona, and C. I. Ugwu, "K-means clustering of electricity consumers using time-domain features from smart meter data," *Journal of Electrical Systems and Information Technology*, vol. 10, no. 1, Jan. 2023, doi: 10.1186/s43067-023-00068-3.
- [21] V. Michalakopoulos, E. Sarmas, I. Papias, P. Skaloumpakas, V. Marinakis, and H. Doukas, "A machine learning-based framework for clustering residential electricity load profiles to enhance demand response programs," *Appl Energy*, vol. 361, May 2024, doi: 10.1016/j.apenergy.2024.122943.
- [22] L. Van Der Maaten and G. Hinton, "Visualizing Data using t-SNE," 2008.
- [23] F. Pedregosa FABIANPEDREGOSA *et al.*, "Scikit-learn: Machine Learning in Python Gaël Varoquaux Bertrand Thirion Vincent Dubourg Alexandre Passos PEDREGOSA, VAROQUAUX, GRAMFORT ET AL. Matthieu Perrot," 2011. [Online]. Available: <http://scikit-learn.sourceforge.net>.

## ANEXOS

Los anexos se encuentran adjuntos y se pueden consultar en el Repositorio Institucional.

**Anexo A:** Scripts de *Python*

**Anexo B:** Perfiles de demanda de circuitos ESSA

**Anexo C:** Clasificación de usuarios ESSA

**Anexo D:** Coordenadas de transformadores de distribución