**Logotipo

Descripción generada automáticamente**

**Pronóstico de demanda de energía eléctrica en un mercado de comercialización en Colombia.**

Juan Esteban Rojas Serna.

Mateo Castañeda Ríos.

Monografía presentada para optar al título de Especialista en Analítica y Ciencia de Datos

Asesor  
David Manuel Villanueva Valdés, Magíster (M.Sc) en Ingeniería de Software

Universidad de Antioquia  
Facultad de Ingeniería

Especialización en Analítica y Ciencia de Datos

Medellín, Antioquia, Colombia

2023

|  |  |
| --- | --- |
| **Cita** | (Rojas Serna & Castañeda Ríos, 2023) |
| **Referencia**  **Estilo APA 7 (2020)** | Rojas Serna, J.E., & Castañeda Ríos, M. (2023). Pronóstico de demanda de energía eléctrica en un mercado de comercialización en Colombia Trabajo de grado especialización, Especialización en Analítica y Ciencia de Datos. Universidad de Antioquia, Medellín, Colombia. |

**** 

Especialización en Analítica y Ciencia de Datos, CohorteVI.

Centro de Investigación Ambientales y de Ingeniería (CIA).

|  |  |
| --- | --- |
|  | Diagrama  Descripción generada automáticamente con confianza media |

Centro de Documentación Ingeniería (CENDOI)

**Repositorio Institucional:** http://bibliotecadigital.udea.edu.co

Universidad de Antioquia - www.udea.edu.co

Rector: John Jairo Arboleda Céspedes.

Decano: Julio Cesar Saldarriaga Molina

Jefe departamento: Diego José Luis Botia Valderrama

El contenido de esta obra corresponde al derecho de expresión de los autores y no compromete el pensamiento institucional de la Universidad de Antioquia ni desata su responsabilidad frente a terceros. Los autores asumen la responsabilidad por los derechos de autor y conexos.

**Tabla de contenido**

Contenido

[Resumen 7](#_Toc151223222)

[Abstract 8](#_Toc151223223)

[1. Descripción del problema 9](#_Toc151223224)

[1.1. Problema de negocio 10](#_Toc151223225)

[1.2. Aproximación desde la analítica de datos 10](#_Toc151223226)

[1.3. Origen de los datos 11](#_Toc151223227)

[1.4. Métricas de desempeño 13](#_Toc151223228)

[2. Objetivos 13](#_Toc151223229)

[2.1. Objetivo general 13](#_Toc151223230)

[2.2. Objetivos específicos 14](#_Toc151223231)

[3. Datos 15](#_Toc151223232)

[3.1. Datos originales 15](#_Toc151223233)

[3.1.1. Histórico de demanda de Energía Eléctrica. 15](#_Toc151223234)

[3.1.2. Histórico de variables climatológicas. 17](#_Toc151223235)

[3.1.3. Histórico de indicadores macroeconómicos. 18](#_Toc151223236)

[3.1.4. Histórico de usuarios conectados al sistema. 20](#_Toc151223237)

[3.2. Datsets 21](#_Toc151223238)

[3.2.1. Conjunto de datos de entrenamiento y de validación. 22](#_Toc151223239)

[3.3. Analítica descriptiva 23](#_Toc151223240)

[Referencias 27](#_Toc151223241)

**Lista de tablas**

[Tabla 1. Estructura del dataset Histórico de demanda de Energía Eléctrica (demanda.csv) 16](#_Toc151223242)

[Tabla 2. *Descripción del archivo demanda.csv* 16](#_Toc151223243)

[Tabla 3. Estructura del dataset con el histórico de variables climatológicas (clima.csv) 17](#_Toc151223244)

[Tabla 4. Descripción del archivo clima.csv 18](#_Toc151223245)

[Tabla 5. Descripción del dataset TRM descargado del Banco de la Republica 18](#_Toc151223246)

[Tabla 6. Características del dataset TRM descargado del Banco de la Republica 18](#_Toc151223247)

[Tabla 7. Descripción del Dataset descargado del DANE sobre el IPC e IPP 19](#_Toc151223248)

[Tabla 8. Dataset resultante de usuarios después de unir todos los archivos del SUI 21](#_Toc151223249)

**Lista de figuras**

[Figura 1. Desviación mensual por mercado de comercialización de la demanda real vs la demanda pronosticada. Elaboración propia a partir de información pública de XM. 9](#_Toc151223250)

[Figura 2. Estructura de archivos con información histórica de la demanda en repositorio de XM 16](#_Toc151223251)

[Figura 3. Lugar en la página del Banco de la Republica en donde se puede Descargar la información histórica de la TRM. Adicional el grafico del comportamiento de la TRM ofrecida por la página. 19](#_Toc151223252)

[Figura 4. Estructura de la página del SUI donde se descarga la información de los usuarios conectados a la red energía. 20](#_Toc151223253)

[Figura 5. Diagrama de cajas y bigotes: i) Demanda de energía horaria por año, ii) Clientes Residenciales por año, iii) Clientes no Residenciales por año y iv) Total usuarios por año. 23](#_Toc151223254)

[Figura 6. Histograma de frecuencia demanda horaria de energía. 24](#_Toc151223255)

[Figura 7. Curva horaria de la demanda de energía real. 24](#_Toc151223256)

[Figura 8. Curva de demanda media horaria y su evolución anual. 25](#_Toc151223257)

[Figura 9. Curva de temperatura media horaria y su evolución anual. 25](#_Toc151223258)

[Figura 10. i) Media de la demanda de energía horaria por mes del año, ii) Temperatura media por mes del año 26](#_Toc151223259)

[Figura 11. Diagrama de dispersión Demanda horaria Vs Temperatura. 26](#_Toc151223260)

**Siglas, acrónimos y abreviaturas**

**ARIMA** AutoRegressive Integrated Moving Average

**CND** Centro Nacional de Despacho

**CNO** Consejo Nacional de Operación

**CREG** Comisión de Regulación de Energía y Gas

**DANE** Departamento Administrativo Nacional de Estadística

**IDEAM** Instituto de Hidrología, Meteorología y Estudios Ambientales

**IPC** Índice de Precios al Consumidor

**IPP** Índice de Precios al Productor

**MC** Mercado de comercialización

**NREL** National Renewable Energy Laboratory

**SIN** Sistema Interconectado Nacional

**SUI** Sistema Único de Información de la Superintendencia de Servicios Públicos Domiciliarios

**TRM** Tasa Representativa del Mercado

**UCP** Unidad de Control de Pronóstico de demanda

# Resumen

El presente trabajo ha sido desarrollado con la finalidad de aplicar modelos de Machine Learning para ser aplicados al pronóstico horario de la demanda de energía eléctrica para un mercado de comercialización específico en Colombia, específicamente el correspondiente al departamento de Antioquia, que permitan disminuir el porcentaje de desviación de los pronósticos con respecto a la demanda de energía real vs los pronósticos realizados actualmente y/o que dicho nivel de desviación permanezca por debajo del +/-4%. Para esto se ha empleado diferentes fuentes de información abiertas para la obtención de los datos históricos relacionados con la demanda de energía eléctrica, cantidad de usuarios conectados al sistema en cada uno de los municipios del departamento, temperatura ambiente, precipitaciones, entre otras variables climatológicas de cada uno de los municipios, así como variables macroeconómicas a nivel nacional.

Dado que los datos provienen de diferentes fuentes y presentan una granularidad temporal diferente, se hace necesario la aplicación de diferentes metodologías para la preparación y homogenización de los datos para ser anexados en un único dataset que brinde información sobre todas las variables mencionadas para cada hora dentro del período de tiempo seleccionado, el cual, para efectos del presente trabajo, se ha tomado como del 2018-01-01 hasta el 2023-09-30. Al conjunto de datos resultante se le aplicaron diferentes análisis descriptivos para profundizar en el entendimiento y comportamiento de los datos y posteriormente se construyeron los respectivos conjuntos de datos de entrenamiento y validación con el fin de ser empleados en los diferentes modelos de Machine Learning Seleccionados.

Repositorio Github:<https://github.com/jestebanrojas/seminario.git>

*Palabras clave*: Demanda de Energía Eléctrica, Pronóstico, Machine Learning, series de tiempo.

# Abstract

The present work has been developed with the purpose of applying Machine Learning models to forecast hourly electricity demand for a specific marketing market in Colombia, specifically in the department of Antioquia. The goal is to reduce the percentage deviation of the forecasts compared to the actual energy demand, either by improving the current forecasts or by keeping the deviation level below +/-4%. To achieve this, various open sources of information have been used to obtain historical data related to electricity demand, the number of users connected to the system in each municipality of the department, ambient temperature, precipitation, and other climatological variables for each municipality. Additionally, macroeconomic variables at the national level have been considered.

Given that the data comes from different sources and exhibits different temporal granularity, it becomes necessary to apply various methodologies for the preparation and standardization of the data to be included in a single dataset that provides information on all the mentioned variables for each hour within the selected time period. For the purposes of this work, the time period has been chosen from January 1, 2018, to September 30, 2023. Different descriptive analyses were applied to the resulting dataset to deepen the understanding of the data and its behavior. Subsequently, the respective training and validation datasets were constructed to be used in the different selected Machine Learning models.

Git Repository: <https://github.com/jestebanrojas/seminario.git>

*Keywords: Electricity Demand, Forecast, Machine Learning, Time Series*

# Descripción del problema

La Comisión de Regulación de Energía y Gas, CREG, en cumplimiento de sus atribuciones, establece el código de redes en la resolución 025 de 1995 y aquí establece las responsabilidades en el pronóstico horario de la demanda y por su parte el CNO - Consejo Nacional de Operación, en el acuerdo CNO 1303 de 2020 determina la mecánica para la presentación al CND - Centro Nacional de Despacho-, por parte de las empresas responsables, el respectivo pronóstico para su mercado de comercialización.  Los operadores de red, como agentes responsables del pronóstico para su respectivo mercado de comercialización, deben presentar al CND el pronóstico de demanda horaria según sus criterios e información disponible, o en su defecto pueden acogerse a los pronósticos de demanda estimados por el CND. Dicho pronóstico deberá estar lo más ajustado posible a la demanda real a fin de garantizar el adecuado despacho de la generación. Para tal fin, los operadores de red han de contar con herramientas técnicas y metodológicas para la estimación de la demanda y estar en la capacidad de suministrar al CND el pronóstico para los siguientes 7 días calendario con un detalle horario por cada día.

De acuerdo con los indicadores publicados por XM en su portal web, se tienen los siguientes porcentajes de desviación entre el pronóstico y la demanda real por cada uno de los mercados de comercialización en el país para los diferentes meses del año 2023.

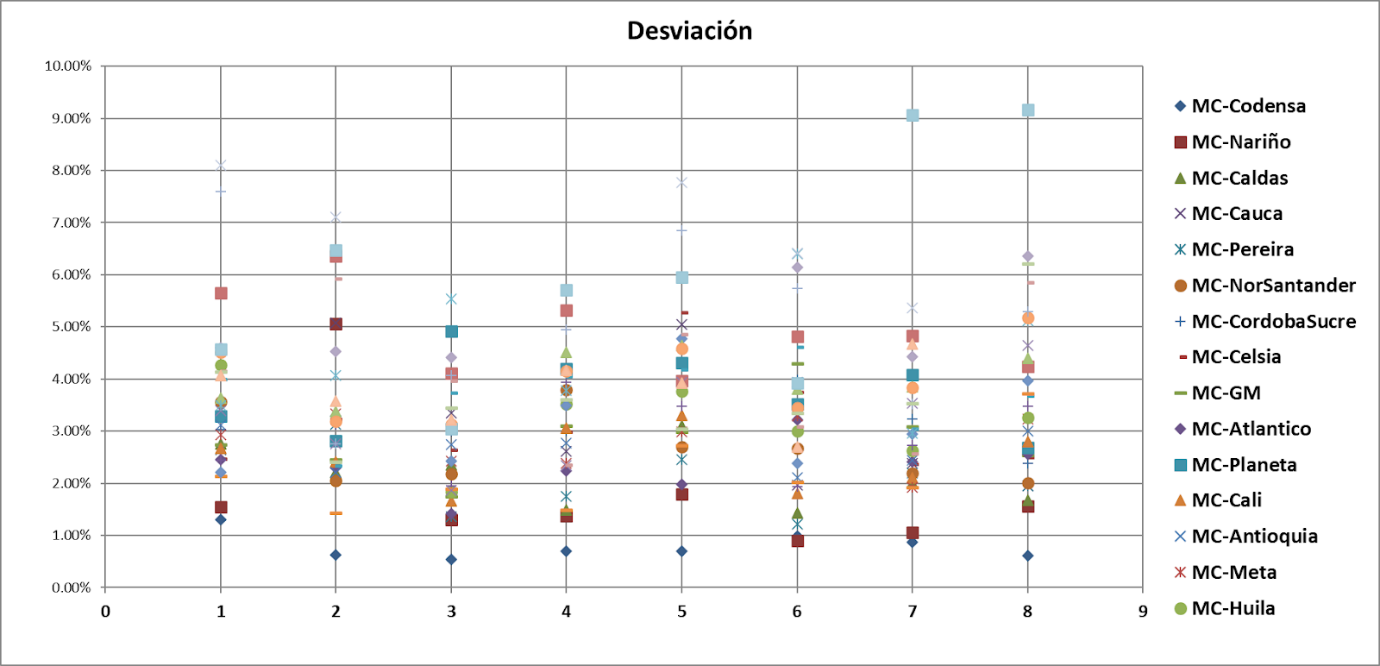


Figura 1. Desviación mensual por mercado de comercialización de la demanda real vs la demanda pronosticada. Elaboración propia a partir de información pública de XM.

El Acuerdo CNO 1303 (2020) considera que valores de desviación que estén por encima del 5%, pueden poner en riesgo la atención de la demanda y en caso de que para un mercado de comercialización se supere este límite por dos días consecutivos, el operador de red deberá realizar el respectivo análisis y presentar al CND las acciones de mejora a aplicar para evitar la futura ocurrencia de las desviaciones y hacer seguimiento a fin de garantizar la efectividad de estas.

Como se puede observar en la . Desviación mensual por mercado de comercialización de la demanda real vs la demanda pronosticada. Elaboración propia a partir de información pública de XM. Figura 1, es evidente que es frecuente que se presenten desviaciones que superen el umbral del 5% y también un mayor número de ocasiones en las que se supera el 4% estando cerca del umbral definido, dando cuenta de la necesidad de establecer mecanismos que faciliten la estimación de la demanda y disminuyan los niveles de desviación con respecto a la demanda real.

## Problema de negocio

Se requiere realizar el pronóstico de demandas operativas del Sistema Interconectado Nacional - SIN - por parte de un operador de red para su mercado de comercialización de forma horaria para un período de 7 días con un nivel de desviación que mejore las predicciones actuales y/o que en todo caso no supere el 4% con respecto a la demanda real, por lo que en el presente trabajo se abordará el mercado de comercialización correspondiente a Antioquia.

## Aproximación desde la analítica de datos

Se aplicarán varias técnicas propias de modelos de regresión lineal o series de tiempo, empleando modelos tales como lo son redes neuronales, máquinas de soporte vectorial o aplicando el método ARIMA. Luego de aplicar estos modelos y modificar sus parámetros para escoger los más óptimos, se tomará aquel que mejor nivel de precisión presente en el pronóstico. El método ARIMA (AutoRegressive Integrated Moving Average) se perfila como una posible alternativa a trabajar dado que puede ser empleado para encontrar patrones para una predicción a partir de datos del pasado y no por variables independientes (Ortuño, Ramos, & Senent, sin fecha). Las redes neuronales también podrán ser exploradas dado que se han empleado con éxito en ejercicios de predicción en series de tiempo con alto grado de precisión, con antecedentes en su aplicación en casos tales como generación de electricidad y consumo de gas natural (Escobar, Luis; Valdés, Julio; Zapata, Santiago, Sin fecha).

Ahora bien, teniendo en cuenta que se cuenta con información específica de temperatura ambiente, precipitaciones, usuarios y variables macroeconómicas, durante el desarrollo se evaluará también la posibilidad de emplear regresiones lineales para cada hora de cada tipo de día, lo que podría resultar en un modelo específico para cada combinación posible entre cada hora (desde las 0:00h hasta las 24:00h) para cada tipo de día, eliminando, bajo este enfoque, la concepción de un modelo dependiente meramente del tiempo sino que incluya posibles variables independientes que definan la demanda de energía. Esto a fin de evaluar si con este método se obtienen mejores precisiones en el pronóstico y se evidencia que bajo este enfoque se pueda inferir que no exista una dependencia exclusivamente del tiempo sino una mayor dependencia de otras variables predictoras.

Para la ejecución de los modelos mencionado se emplearán herramientas tales como Python y librerías disponibles tales como Scikit-learn, stats-models, statsForecast, entre otras.

## Origen de los datos

Teniendo en cuenta lo mencionado en el numeral 1.2, se requerirán datos históricos desde el 1 de enero de 2018 tales como la demanda de energía eléctrica histórica para el mercado de comercialización específico, así como información histórica de temperatura ambiente, precipitaciones, usuarios y variables macroeconómicas.

**Demanda histórica por mercado de comercialización con resolución horaria desde 2018.** Información disponible directamente de la fuente oficial (XM) con la información de la demanda real del Sistema Interconectado Nacional.

**Históricos de temperatura ambiente y precipitaciones con resolución horaria por municipio perteneciente al mercado de comercializador del Operador de Red específico.** Información disponible en bases de datos de la NASA, el National Renewable Energy Laboratory (NREL), El IDEAM, entre otros.

**Pronóstico de temperatura y precipitación para un período de 7 días de la semana siguiente con resolución horaria.** Información tomada de las publicaciones oficiales del IDEAM.

**Indicadores macroeconómicos históricos (mensuales / anuales) tales como TRM, IPP e IPC en Colombia desde el 2018**. Información disponible en fuentes tales como el DANE y Banco de la República.

**Usuarios totales conectados al Operador de Red por Municipio a final de cada mes.** Información disponible en el SUI - Sistema Único de Información de Servicios Públicos Domiciliarios.

Al tratarse de data oficial de entidades como XM y la superintendencia de servicios públicos domiciliarios (Demandas operativas de energía y cantidad de usuarios) que corresponde a reportes de carácter regulatorio y periódicos, se cuenta con información que cumple con criterios de completitud y sometida a diferentes procesos de calidad de datos por parte de las entidades encargadas de su reporte; con respecto a la información climática y macroeconómica se emplearán también fuentes oficiales o reconocidas en su materia, quienes publican de forma recurrente dicha información.

La información a emplear corresponde a información de dominio público dispuesta en los portales web específicos de cada entidad mencionada, la cual puede ser consultada y exportada en archivos tipo CSV y/o Excel para diferentes rangos de tiempo según la entidad. Para el caso de las demandas operativas históricas se puede acceder a la información por cada mercado de comercialización de un operador de red específico, con archivos mensuales desde el año 2018. Los pronósticos de temperatura pueden hallarse en el portal web del IDEAM para cada municipio para la siguiente semana; por su parte los valores de temperatura histórica por horas se pueden obtener de repositorios como el de NREL o la NASA, para lo cual se tomará como referencia la temperatura en las coordenadas del casco urbano de cada municipio. La información de los usuarios conectados en cada municipio se encuentra disponible en el portal del SUI, Aquí se puede descargar la información de cada mes con la cantidad de usuarios según su tipo discriminados por municipio para un departamento seleccionado.

Los indicadores macroeconómicos serán obtenidos a través de información disponible en el DANE y de ser necesario en el Banco de la República.

## Métricas de desempeño

Los modelos se evaluarán a partir de las métricas disponibles para cada modelo aplicado, tales como F1-Score, R^2, Accuracy, entre otros y adicionalmente se evaluarán los resultados a través de la comparación entre la demanda real registrada para el mercado de comercialización de un Operador de Red específico y el pronóstico del modelo; dicha comparación servirá para medir el nivel de desviación entre ambos.

Si bien se busca que el porcentaje de desviación se encuentre por debajo del rango de +/- 4%, se medirá también si existe un porcentaje de mejora en el pronóstico por parte del modelo aplicado vs el pronóstico realizado por parte del operador de red específico.

A continuación de describen las métricas de negocio mencionadas:

# Objetivos

## Objetivo general

Plantear una solución para el pronóstico de la demanda de energía eléctrica en un mercado de comercialización determinado en Colombia atendido por un operador de red específico en cumplimiento de las disposiciones emitidas por la CREG – Comisión de Regulación de Energía y Gas - y por el Consejo Nacional de Operación - CNO a través del planteamiento de un modelo basado en analítica y ciencia de datos minimizando el porcentaje de error existente entre el pronóstico y la demanda real.

## Objetivos específicos

* Obtener, procesar y analizar la información histórica proveniente de diferentes fuentes disponibles que pueda influir en el comportamiento de la demanda de energía eléctrica.
* Identificar aquellas variables predictoras que presentan mayor relación e influencia en la demanda de energía eléctrica.
* Desarrollar, entrenar y probar uno o varios modelos de machine learning para la predicción de la demanda de energía eléctrica para los siguientes 7 días a partir de la información procesada con las variables predictoras elegidas.
* Elegir el modelo que mejor precisión presente al evaluarlos con los datos de prueba.
* Aplicar el modelo elegido para la predicción de la demanda de energía eléctrica con resolución horaria para los siguientes 7 días.

# Datos

## Datos originales

Como se ha mencionado en el numeral 1, la información a ser considerada corresponde a: demanda de energía eléctrica histórica para el mercado de comercialización específico, información histórica de temperatura ambiente, precipitaciones, usuarios y variables macroeconómicas; a continuación, se hace una descripción de los datos:

## Histórico de demanda de Energía Eléctrica.

La información correspondiente a la demanda de energía eléctrica histórica para el mercado de comercialización - MC - seleccionado, el cual para el caso del presente trabajo se define que será Antioquia, se recurre al portal de XM en el cual se realiza el respectivo cargue de manera mensual para todos los MC.

En este caso, la información se encuentra en ficheros que responden a la siguiente estructura: año/mes/informe\_mc.xls o año/mes/informe\_mc.xlsx donde el archivo año/mes/informe\_mc.xls o año/mes/informe\_mc.xlsx corresponde al informe de los indicadores de pronósticos oficiales de demanda para cada mercado de comercialización, por lo que se procede con la descarga de los datos para el MC seleccionado (Antioquia). La extensión y nombre del archivo puede variar según la época y cambios en el uso de las herramientas de excel empleadas por el Centro nacional de Despacho - CND - para el respectivo reporte.

Una vez se han descargado los datos históricos entre el 2018-01-01 y el 30-09-2023, se deberá proceder con la unión de cada uno de los archivos para la conformación de un dataset único con todo el histórico de la demanda.

La estructura de los datos es tal y como se muestra a continuación:

Tabla . Estructura del dataset Histórico de demanda de Energía Eléctrica (demanda.csv)

|  |  |  |
| --- | --- | --- |
| **Nombre** | **Descripción** | **Tipo** |
| UCP | Unidad de Control de Pronóstico. Para efectos del presente trabajo corresponde a Antioquia. | Categórica |
| Variable | Tipo de variable del reporte, para efectos del presente trabajo siempre corresponderá a “Demanda Real”. | Categórica |
| FECHA | Fecha en formato Año-Mes-Día al cual corresponde la demanda real. | Datetime |
| TIPO\_DIA | Tipo de día al cual corresponde la demanda real, indicando si corresponde a un día entre semana, fin de semana, día festivo o alguna fecha especial. | Categórica |
| P1…P24 | Energía demandada para la UCP en cada una de las 24 horas del día indicado en el campo FECHA. | Numérica |
| Total | Energía total demandada para la fecha indicada en el campo FECHA. | Numérica |
| PO19, PO20, PO21 | Demanda de potencia máxima para las horas 19, 20 y 21 de la fecha indicada en el campo FECHA. | Numérica |

La siguiente imagen ilustra la disposición de los archivos en el repositorio web dispuesto por XM.

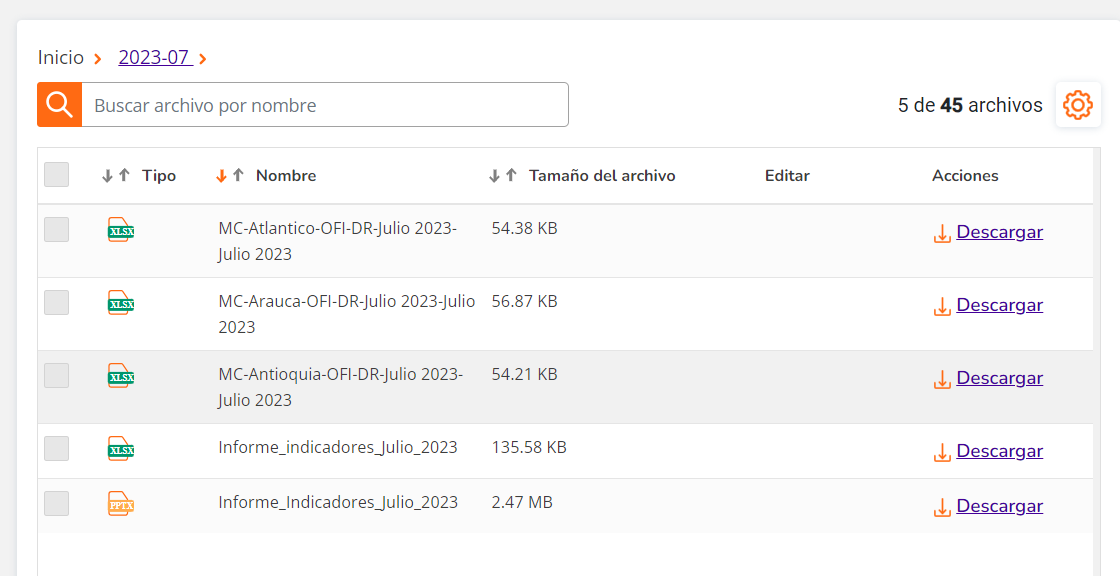


Figura . Estructura de archivos con información histórica de la demanda en repositorio de XM

La siguiente tabla muestra la descripción del dataset resultante:

Tabla . *Descripción del archivo demanda.csv*

|  |  |
| --- | --- |
| **Descripción** | **Características** |
| Nombre del dataset | demanda.csv |
| Tamaño | 633 kb |
| Cantidad Registros | 2099 |

## Histórico de variables climatológicas.

Para este caso, la información climatológica es la correspondiente a la información histórica con desagregación horaria de temperatura, precipitaciones, humedad relativa, radiación solar, claridad del cielo, entre otros, de los municipios del departamento del mercado de comercialización elegido. La información se toma del portal de la NASA dispuesto para tal fin y se emplea la API disponible en: // https://power.larc.nasa.gov/data-access-viewer/. Para descargar esta información se hace necesario contar con las coordenadas de los municipios respectivos, por lo que se toma la información disponible en el DANE y se dispone previamente para ser empleada por un script de Python que se encargará de correr la API para cada uno de los municipios del departamento.

La estructura de la data entregada por la API es de la siguiente manera:

Tabla . Estructura del dataset con el histórico de variables climatológicas (clima.csv)

|  |  |  |
| --- | --- | --- |
| **Nombre** | **Descripción** | **Tipo** |
| **YEAR** | Año | Numérica |
| **MO** | Mes | Numérica |
| **DY** | Día | Numérica |
| **HR** | Hora | Numérica |
| **ALLSKY\_SFC\_SW\_DWN** | Irradiancia de onda corta que llega a la superficie terrestre desde todas las direcciones del cielo, independientemente de las condiciones meteorológicas o la presencia de nubes medido en Wh/m^2. | Numérica |
| **ALLSKY\_KT** | Indicador de la claridad del cielo en relación con la radiación solar incidente. | Numérica |
| **T2M** | Temperatura a 2m en °C | Numérica |
| **RH2M** | Humedad relativa a 2 metros en % | Numérica |
| **PRECTOTCORR** | Precipitación Corregida en mm/hora | Numérica |
| **CLRSKY\_SFC\_SW\_DWN** | Cantidad de radiación solar de onda corta que llega a la superficie terrestre bajo condiciones de cielo despejado, es decir, sin nubes ni obstrucciones atmosféricas medido en Wh/m^2. | Numérica |
| **T2MWET** | Temperatura de bulbo húmedo a 2m en °C | Numérica |

Es de anotar que adicional a la información contenida en la estructura anterior, la información para cada coordenada viene acompañada de un “Header”, el cual debe ser eliminado para ser anexado a un dataset con la información para todas las coordenadas, esto último también se realiza en el script de Python mencionado previo a almacenar el dataset en un archivo csv.

La siguiente tabla muestra la descripción del dataset resultante:

Tabla . Descripción del archivo clima.csv

|  |  |
| --- | --- |
| **Descripción** | **Características** |
| Nombre del dataset | clima.csv |
| Tamaño | 376.2 Mb |
| Cantidad Registros | 2099 |

## Histórico de indicadores macroeconómicos.

**3.1.3.1. Histórico TRM**

El dataset TRM se toma de la página web del Banco de República de Colombia, en el siguiente enlace:<https://www.banrep.gov.co/es/estadisticas/trm>. Allí nos dirigimos al botón Descargar y se descarga toda la información histórica del precio de cierre de la TRM en Colombia desde el año 1991 hasta el día anterior que se descargue el archivo. Este archivo de la TRM consta de 2 columnas:

Tabla . Descripción del dataset TRM descargado del Banco de la Republica

|  |  |  |
| --- | --- | --- |
| **Nombre** | **Descripción** | **Tipo** |
| Fecha | Fecha al cierre de ese día de la TRM | Datetime |
| TRM | Valor de la TRM al cierre del mercado de divisas | Numérica |

Al seguir los pasos anteriores, se descarga un archivo con las siguientes características:

Tabla . Características del dataset TRM descargado del Banco de la Republica

|  |  |
| --- | --- |
| **Descripción** | **Características** |
| Nombre del dataset | TRM.xslx |
| Tamaño | 175 kb |
| Cantidad Registros | 11687 |

En la siguiente imagen se puede evidenciar el lugar en la página web donde se descarga la información para la TRM, adicionalmente se encuentra un gráfico del comportamiento de la TRM en Colombia:



Figura . Lugar en la página del Banco de la Republica en donde se puede Descargar la información histórica de la TRM. Adicional el grafico del comportamiento de la TRM ofrecida por la página.

**3.1.3.2. Histórico IPP e IPC**

El dataset IPC e IPC corresponde a la información histórica con resolución mensual desde el año 2018 a 2023 del comportamiento de estas dos variables macroeconómicas. Estos se pueden descargar de la página del DANE y la estructura del dataset es la siguiente:

Tabla . Descripción del Dataset descargado del DANE sobre el IPC e IPP

|  |  |  |
| --- | --- | --- |
| **Nombre** | **Descripción** | **Tipo** |
| Fecha | Mes y año de la información del IPP e IPC | Datetime |
| IPP | Unidades del IPP | Numérica |
| IPC | Unidades del IPC | Numérica |

Complementando la información de la , el dataset Evolución IPC E IPP base 2008 - Resolución\_ Mensual.xlsx, tiene un tamaño de 8 Kb y 71 filas.

## Histórico de usuarios conectados al sistema.

El dataset usuarios.csv, corresponde a la información histórica de la cantidad de usuarios conectados a la red de energía a nivel nacional desde enero del 2018 a septiembre del 2023.

Para conocer esta información se descargó de la página web del SUI (Sistema único de información de servicios públicos domiciliarios). En dicho aplicativo solo se permite descargar 1 mes a la vez por cada año de todos los municipios de Antioquia, por ello se descargaron 69 archivos .csv, los cuales posteriormente se unieron en un solo archivo usuarios.csv de forma ordenada usando el lenguaje de programación Pyhton. El enlace para descargar los archivos es el siguiente: http://reportes.sui.gov.co/fabricaReportes/frameSet.jsp?idreporte=ele\_com\_096.

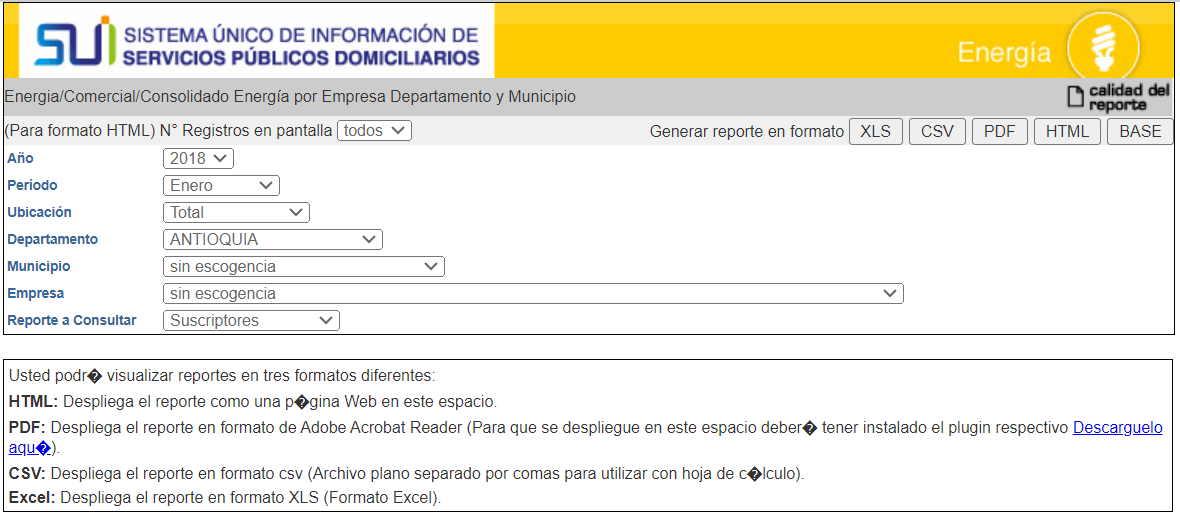


Figura . Estructura de la página del SUI donde se descarga la información de los usuarios conectados a la red energía.

Los parámetros tomados fueron los siguientes:

* Año: 2018 hasta 2023
* Periodo: Enero a Diciembre, para el año 2023 hasta septiembre
* Ubicación: Total (incluye Urbano, Rural, Centro Poblado)
* Departamento: ANTIOQUIA
* Municipio: Sin escogencia (Todos los municipios del departamento)
* Empresa: Sin escogencia (Todas las empresas)
* Reporte a consultar: Suscriptores (usuarios)

Luego de descargar y unir los archivos de cada mes, el resultado es el dataset usuarios.csv con un peso de 609 kb y una cantidad de registros de 8488. Cuenta con la siguiente estructura:

Tabla . Dataset resultante de usuarios después de unir todos los archivos del SUI

|  |  |  |
| --- | --- | --- |
| **Nombre** | **Descripción** | **Tipo** |
| MUNICIPIO | Nombre del municipio de Antioquia. | Categórica |
| AÑO | Año correspondiente al reporte de usuarios por municipio. | Numérica |
| ESTRATO | Cantidad de usuarios del estrato 1 al estrato 6 en Antioquia por municipio en determinado día. | Numérica |
| TOTAL RESIDENCIAL | Cantidad de usuarios totales (Suma del estrato 1 al estrato 6) en Antioquia por municipio en determinado día. | Numérica |
| INDUSTRIAL | Cantidad de usuarios industriales en Antioquia por municipio en determinado día. | Numérica |
| COMERCIAL | Cantidad de usuarios comerciales en Antioquia por municipio en determinado día. | Numérica |
| OFICIAL | Cantidad de usuarios oficiales en Antioquia por municipio en determinado día. | Numérica |
| OTROS | Cantidad de usuarios otros en Antioquia por municipio en determinado día. | Numérica |
| TOTAL NO RESIDENCIAL | Cantidad de usuarios totales no residenciales (Suma de industrial, comercial, oficial, otros) en Antioquia por municipio en determinado día | .Numérica |

## Datsets

Dadas las características propias de cada fuente de datos, se deberá realizar un proceso de preparación de estos, para garantizar que, para el rango de tiempo histórico definido, se cuente con información de las diferentes variables, esto es que para cada hora de cada día se cuente con:

* Demanda de energía real.
* Temperatura promedio de cada municipio, asumida como la temperatura percibida en las coordenadas del casco urbano del municipio.
* Cantidad de clientes conectados a la red en cada municipio, para lo cual se asumirá como un valor constante para cada día y hora de cada mes según los registros reportados en el SUI.
* Precipitaciones registradas en promedio en cada municipio.

Teniendo en cuenta lo anterior y una vez se ha consolidado cada uno de los datasets resultantes de los numerales 3.1.1 a 3.1.4, se procede a realizar las transformaciones necesarias para la construcción de un único dataset, de modo tal que todas las variables queden a nivel horario y que de este modo puedan ser empleadas como variables predictoras en la aplicación de los diferentes modelos de Machine Learning que así lo requieran. Dichas transformaciones se describen a continuación:

* Consolidar la información de usuarios, clima y datos macroeconómicos, en un dataset que contendrá información para el departamento de Antioquia en cada día/hora del periodo 01/01/2018 al 31/10/2023. Para esto se carga el dataset con las variables climatológicas y se une con el dataset de usuarios, teniendo como llaves las columnas municipio, año y mes. En este caso para cada día/hora en cada municipio se toma como cantidad de usuarios la correspondiente a la cantidad de usuarios del mes para dicho municipio reportado, dado que esta información no se tiene con la resolución día/hora.
* Se procede a realizar una ponderación de cada una de las variables climatológicas para cada municipio, con la cantidad de usuarios para cada período de tiempo y obtener posteriormente un valor promedio ponderado para Antioquia para cada variable climatológica.
* Posteriormente se une el dataset anterior con los dataset que contienen la información histórica de la TRM así como del IPP e IPC.
* Por otra parte, se debe adecuar la información histórica de la demanda, con la finalidad de transformar las columnas con la demanda de cada hora en registros individuales para luego unir el dataset con el resultado del paso anterior, obteniendo así el dataset que servirá para la aplicación de los diferentes modelos.

## Conjunto de datos de entrenamiento y de validación.

Para la separación del dataset en conjuntos de entrenamiento y validación, se trabajará desde 2 enfoques diferentes. El primero de ellos, y teniendo en cuenta que se aplicarán series de tiempo, será el de tomar como datos de entrenamiento los correspondientes al período comprendido entre las fechas 2018-01-01 al 2022-12-31 y como datos de validación los datos entre las fechas 2023-01-01 al 2023-09-30. Lo anterior debido a que es necesario conservar el orden de los datos, entendiendo que existe una dependencia entre una observación y la siguiente (Radečić, 2021).

El segundo enfoque consiste en la división aleatoria del dataset empleando la función train\_test\_split de la librería de Python sklearn, en este caso se toma el 70% de los datos para entrenamiento y el 30% restante para validación, este segundo enfoque, será empleado cuando se apliquen modelos de Machine Learning diferenciados por cada tipo de día y hora del día.

## Analítica descriptiva

La siguiente gráfica muestra una tendencia general por cada año en el crecimiento tanto en la demanda como en los usuarios residenciales, no residenciales y totales. Se aprecia que la demanda en el año 2020, derivado de la pandemia por COVID19 de tuvo una demanda horaria inferior a 2019, pero del 2021 en adelante se observa nuevamente una tendencia al alza.

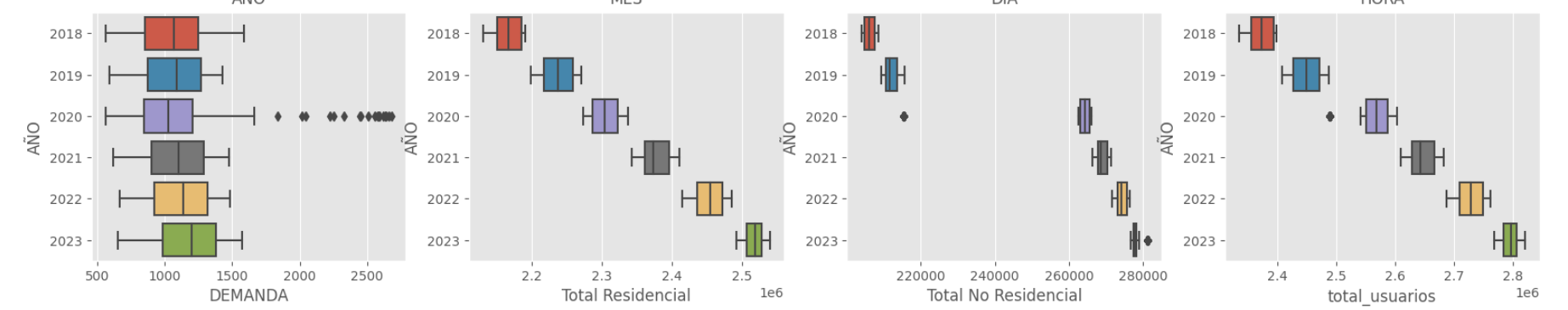


Figura 5. Diagrama de cajas y bigotes: i) Demanda de energía horaria por año, ii) Clientes Residenciales por año, iii) Clientes no Residenciales por año y iv) Total usuarios por año.

Ahora bien, tanto en la como en la se puede apreciar que existen algunos valores atípicos para las demandas horarias, los cuales se encuentran en magnitudes superiores a los 2.000 MWh, aspecto que se puede apreciar con mejor detalle en la , en la cual se evidencia que los valores atípicos corresponden a un día en específico, el cual se encuentra significativamente por encima de los valores típicos para cada una de las horas del día.

Gráfico, Histograma

Descripción generada automáticamente

Figura . Histograma de frecuencia demanda horaria de energía.

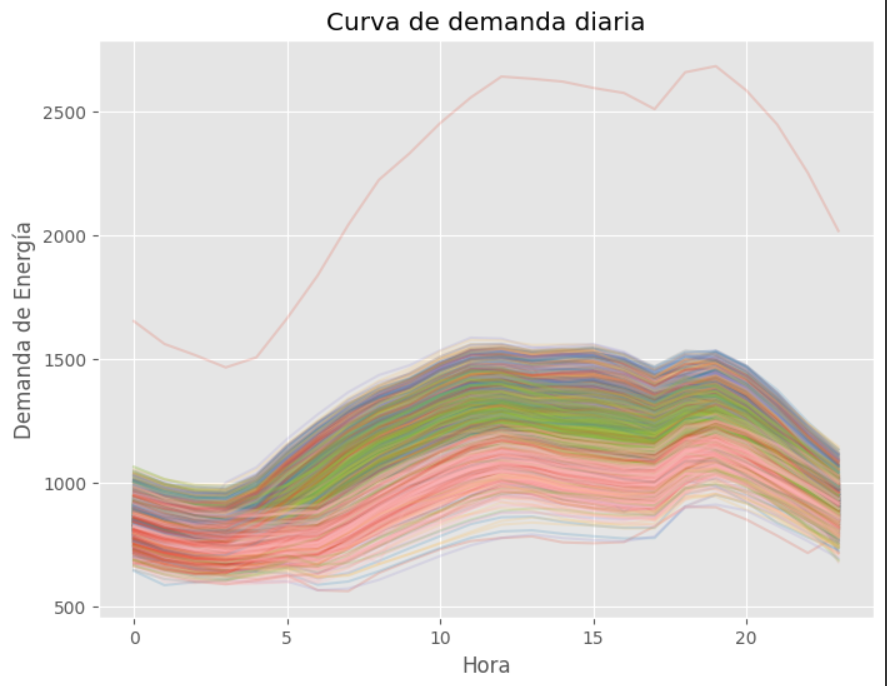


Figura . Curva horaria de la demanda de energía real.

De la , se puede apreciar que la demanda horaria presenta un patrón característico para cada hora del día, según su tipo, teniendo que, para las franjas comprendidas entre las 11:00 y las 12:00, así como a las 19:00, se presentan típicamente los momentos de máxima demanda del día. Como se puede apreciar en la , la demanda va creciendo en el tiempo conservando dicho patrón diario, salvo en 2020 en donde la demanda decreció con respecto al año inmediatamente anterior, pero en este caso también se preserva el patrón.

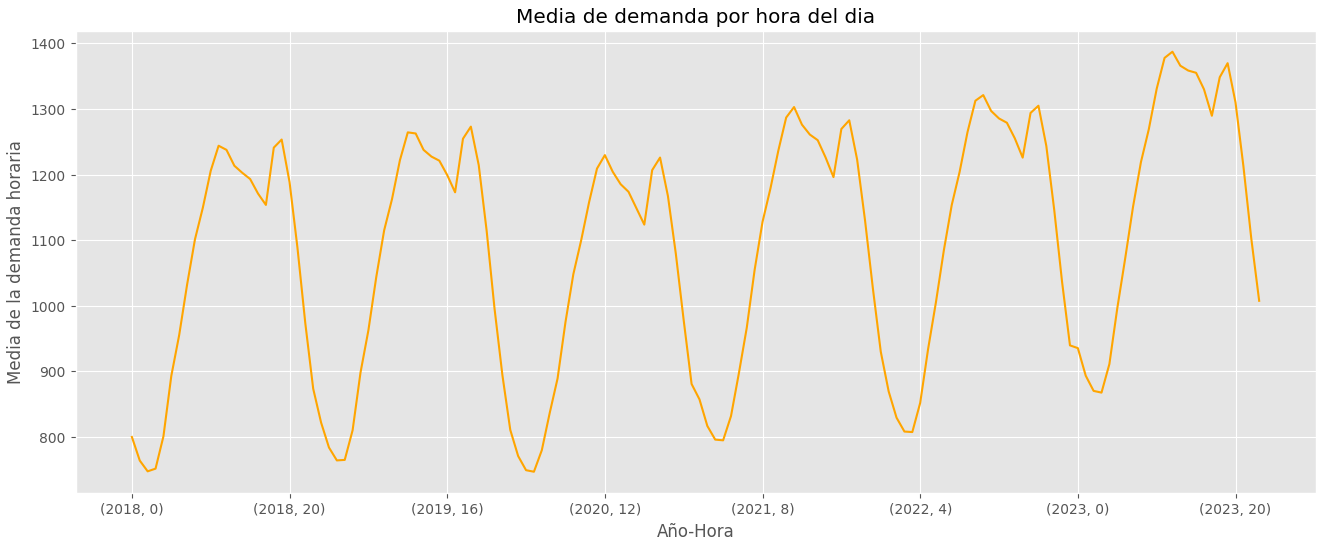


Figura . Curva de demanda media horaria y su evolución anual.

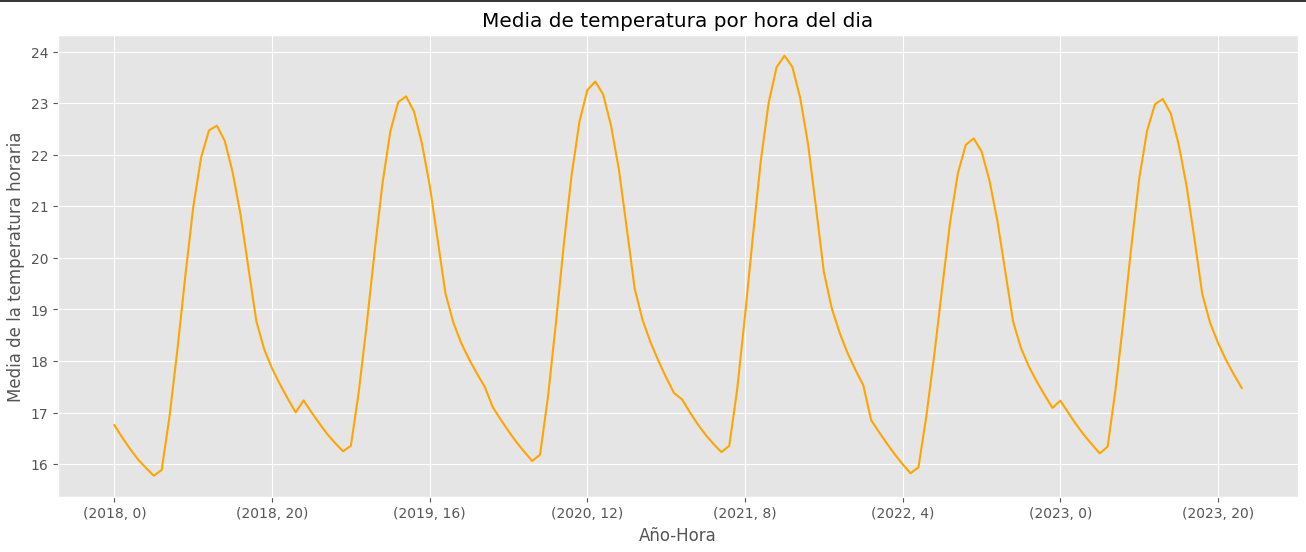


Figura . Curva de temperatura media horaria y su evolución anual.

Con miras a identificar si existe una relación clara entre la temperatura ambiente y la demanda de energía, se procede en primera instancia a evaluar si existe algún patrón específico en el comportamiento mensual. La , muestra la demanda de energía horaria promedio durante cada mes, así como también la temperatura promedio mensual, al realizar el análisis comparativo entre ambos se identifica que no necesariamente en los meses donde se presentan mayores temperaturas, la demanda horaria promedio sea superior, esto solo parece cumplirse para el mes de septiembre, pero no así para el mes de mayo, el cual presenta las mayores temperaturas, pero la demanda no se encuentra entre las máximas del mes.



Figura . i) Media de la demanda de energía horaria por mes del año, ii) Temperatura media por mes del año

Considerando el análisis anterior, se procede a realizar un análisis adicional, aplicando en este caso un diagrama de dispersión entre la temperatura horaria y demanda de energía horaria, cuyo resultado se aprecia en la . En este caso se puede apreciar una ligera tendencia a que en la medida que aumenta la temperatura, la demanda de energía es más elevada, lo que da cuenta de que el análisis a aplicar deberá contemplar aspectos como la hora del día a la que se presenta cierta temperatura, así como también el tipo de día, ya que, al retomar la curva diaria de la Figura 5, a las 19:00 se presenta uno de los picos de demanda, a pesar de que la temperatura promedio para esta hora no corresponde a una de las mayores temperaturas del día, sino que dicha hora coincide con aspectos relacionados con el desarrollo de las actividades cotidianas de los hogares en el país.

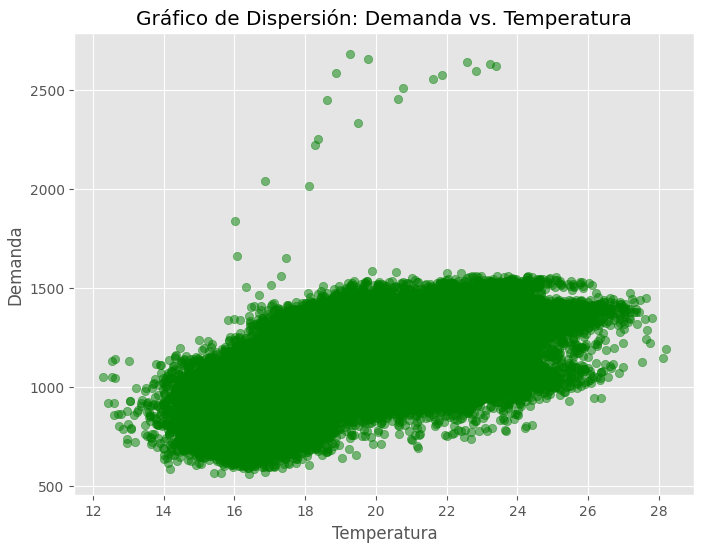


Figura . Diagrama de dispersión Demanda horaria Vs Temperatura.

# Referencias

Consejo Nacional de Operación [CNO]. (2020). Acuerdo 1303, Por el cual se actualizan los procedimientos para la gestión integral de la demanda.

Escobar, Luis; Valdés, Julio; Zapata, Santiago. (Sin fecha). *Redes Neuronales Artificiales en predicción de Series de Tiempo. Una aplicación a la Industria. .* Obtenido de Universidad de Palermo: https://www.palermo.edu/ingenieria/Pdf2010/CyT9/02.pdf

Ortuño, J. M., Ramos, J. A., & Senent, J. C. (sin fecha). *Modelo Arima. Análisis Estadístico de Series Económicas*. Obtenido de https://rstudio-pubs-static.s3.amazonaws.com/384039\_cc37e393f643455bb01ad4b392a081bd.html

Radečić, D. (26 de Julio de 2021). *Time Series From Scratch — Train/Test Splits and Evaluation Metrics*. Obtenido de Towards data science: https://towardsdatascience.com/time-series-from-scratch-train-test-splits-and-evaluation-metrics-4fd654de1b37