

Predicción del Consumo Eléctrico mediante Análisis de Big Data y Series Temporales

Corporación Universitaria Remington.
Nombre de la facultad: Facultad de Ingenierías
Nombre del programa académico: Ingeniería de Sistemas

Santiago Silva Rojas
Juan pablo Vélez Uribe
Seminario-Diplomado
Agosto 2025

Dedicatoria

Quiero brindar este esfuerzo a mi familia, por quererme y apoyarme en cada paso de mi existencia, motivándome siempre.

A mis padres, por instruirme a ser constante y no tirar la toalla y a mis profesores, por mostrarme su saber y su paciencia.

A cada persona que, de algún modo, facilitó este cometido, les ofrezco mi gratitud más sentida y honesta.

Agradecimientos

Expreso mi profundo agradecimiento a Dios por brindarme la energía y el entendimiento cruciales para terminar este proyecto.

A mi familia, por creer en mí y darme su apoyo constante, sin dudar ni un instante.

A mis profesores, por su guía y por impulsarme a usar lo aprendido de manera práctica y con inventiva, y a todos los individuos que aportaron su grano de arena, gracias por ser parte de esta victoria.

Este trabajo es, además, un espejo de la entrega y el empeño de quienes estuvieron a mi lado en esta travesía académica.

Tabla de Contenidos

Resumen.....	4
Palabras clave.....	6
Marco conceptual y contextual	6
Planteamiento del problema.....	6
Justificación	6
Objetivo general.....	7
Objetivos específicos	7
Conceptos clave	7
Contexto del proyecto	8
Desarrollo e implementación del aprendizaje.....	8
Descripción del dataset	9
Carga y exploración inicial de datos.....	9
Comprobación gráfica inicial.....	10
Proceso ETL aplicado al proyecto	11
Extracción (Extract).....	12
Transformación (Transform)	12
Carga (Load)	12
Control de calidad de datos (Data Quality).....	13
Complejidad.....	13
Consistencia	14
Exactitud	14
Unicidad.....	14
Validez de formato.....	14
Conclusión del Data Quality.....	15
Limpieza de datos	15
Análisis exploratorio de datos (EDA).....	15
a) Consumo promedio diario	17
b) Patrón de consumo mensual	18
c) Consumo por hora del día.....	19
d) Variabilidad semanal	20
Conclusiones del EDA.....	21
Modelado predictivo	21
Procedimiento	21
Resultados	24
Conclusión de la etapa	28
Visualización de datos con Power BI	28
Conclusiones	30
Referencias.....	31

Resumen

El propósito fundamental de esta investigación es idear un modelo predictivo que nos permita visualizar cómo será el consumo eléctrico en el futuro, partiendo de datos históricos detallados. Para lograrlo, echaremos mano de técnicas de Big Data y, por supuesto, de series temporales.

Para este fin, recurrimos al conjunto de datos "Household Electric Power Consumption", que guarda registros minutos a minuto del consumo de un hogar desde diciembre de 2006 hasta noviembre de 2010.

Al principio, trazamos **un marco conceptual y contextual** y pusimos todo en contexto, aclarando ideas clave como Big Data, series temporales y predicción del consumo eléctrico. Además, definimos el problema, justificamos el proyecto y fijamos los objetivos a alcanzar.

Después, en la etapa de **desarrollo e implementación**, nos encargamos de subir, limpiar y alistar el conjunto de datos. Esto implicó deshacernos de los valores vacíos, adaptar los datos a los formatos correctos y crear un índice temporal.

Explorar los datos **análisis exploratorio de datos** nos ayudó a detectar patrones importantes, como los momentos de mayor consumo, las diferencias entre los días de semana y los fines de semana, y un ritmo estacional que se repite cada año.

Para predecir, usamos la librería **Prophet**. Con ella, entrenamos el modelo usando el consumo promedio diario y calculamos una proyección para los siguientes 180 días. Las predicciones respetan la tendencia del pasado y la estacionalidad, lo que indica que el modelo puede prever cuándo habrá más demanda.

Los resultados señalan que predecir el consumo eléctrico es muy útil para usar mejor los recursos, planificar la generación y ser más conscientes al usar la energía. No obstante, sabemos que añadir más factores, como datos del clima o socioeconómicos, podría hacer que las predicciones sean aún más precisas.

Palabras clave

Big Data, predicción, consumo eléctrico, series temporales, Prophet

Marco conceptual y contextual

Planteamiento del problema

Manejar el consumo eléctrico de forma eficiente es clave para bajar los gastos y evitar que las redes se saturen. Sin embargo, la mayoría de las casas y negocios chicos no tienen sistemas que les permitan prever cuánta energía van a necesitar. Esto dificulta la planificación del uso de la energía y puede causar problemas.

El punto central radica en la ausencia de un esquema que estime el consumo venidero, tomando como base registros históricos detallados.

Usar métodos de Aprendizaje Automático para analizar series de tiempo en energía no solo refina la exactitud de los pronósticos, sino que también ayuda a incorporar datos diversos y al minuto. Esto resulta clave en entornos de Big Data, donde la velocidad y la variedad de datos exigen modelos flexibles y escalables (Criado Ramón, 2024).

Justificación

El consumo de energía se ha disparado por el aumento de la población y porque usamos la tecnología para todo. Tener un sistema que adivine cuánta energía se va a necesitar permitiría usar mejor los recursos, planear cómo distribuir la electricidad y ahorrar dinero. Este proyecto usa Big Data y Machine Learning para demostrar que analizar datos antiguos puede dar soluciones prácticas y que se pueden usar en muchos sitios.

La implementación de técnicas de predicción con series de tiempo en el ámbito residencial contribuye a una mejor gestión de la demanda energética. Estos modelos, al identificar patrones de consumo, permiten diseñar estrategias para reducir picos de carga y mejorar la eficiencia en el uso de la red eléctrica (Tasinchana Chicaiza, 2015).

Objetivo general

Crear un modelo que prediga, usando datos históricos de consumo eléctrico por minuto, cuánta energía se va a necesitar para así planificar mejor y usar los recursos de forma óptima.

Objetivos específicos

1. Obtener y preparar un dataset de consumo eléctrico para su análisis.
2. Realizar un análisis exploratorio que identifique patrones de consumo y factores que afectan la demanda.
3. Implementar un modelo de predicción usando técnicas de series temporales o Machine Learning.
4. Validar y evaluar el modelo utilizando métricas estadísticas.
5. Presentar resultados mediante visualizaciones y conclusiones aplicables a entornos reales.

Conceptos clave

Big Data: Hablamos del manejo de volúmenes enormes de información que, por su tamaño, velocidad de generación y variedad, requieren herramientas avanzadas para su protección, gestión y análisis. Tales tácticas facilitan extraer información valiosa para la toma de decisiones cruciales.

Series temporales: Son conjuntos de datos que se recogen y se organizan en lapsos de tiempo, y lo esencial es que las observaciones anteriores impactan en las siguientes. Examinarlas es fundamental para detectar tendencias, patrones estacionales y fluctuaciones periódicas, lo cual es provechoso en muchas industrias, como la energética.

Predicción del consumo eléctrico: En el ámbito eléctrico, poder anticipar la demanda futura de energía es vital para perfeccionar la distribución, bajar los costos y asegurar la estabilidad de la red. No obstante, la carencia de sistemas predictivos en hogares y negocios pequeños dificulta prever aumentos de consumo o detectar momentos de baja demanda.

Contexto del proyecto

En este proyecto se recurrirá al conjunto de datos Household Electric Power Consumption, procedente de Kaggle y del almacén UCI Machine Learning. Esta base de datos contiene registros de gasto eléctrico por minuto durante cerca de cuatro años, lo que permite poner en práctica métodos de modelado predictivo para calcular la demanda en lapsos venideros. Si bien no se pone en marcha en una compañía concreta, el análisis reproduce una situación verdadera en la que un hogar o entidad podría sacar partido de un sistema de predicción de gasto eléctrico, favoreciendo una administración más eficiente de la energía.

Desarrollo e implementación del aprendizaje

Descripción del dataset

Para llevar a cabo este estudio, se recurrió al conjunto de datos Household Electric Power Consumption, que está accesible tanto en Kaggle como en el archivo UCI Machine Learning. Esta base de datos abarca el consumo eléctrico minuto a minuto de una vivienda, desde diciembre de 2006 hasta noviembre de 2010.

El documento inicial viene en formato .txt, usa punto y coma (;) como separador, y a veces muestra datos ausentes marcados con el símbolo (?). Las variables principales incluyen:

- **Global_active_power**: Potencia activa global (kW).
- **Global_reactive_power**: Potencia reactiva global (kW).
- **Voltage**: Voltaje medio (V).
- **Global_intensity**: Intensidad de corriente (A).
- **Sub_metering_1, 2, 3**: Energía medida por tres medidores diferentes (Wh).
- **Date y Time**: Fecha y hora de la medición.

Carga y exploración inicial de datos

Para manipular el dataset, se utilizó el lenguaje de programación Python con las librerías **Pandas** y **Matplotlib**.

Se cargaron los datos indicando que la columna de fecha y hora se combinaran en una sola denominada datetime. Además, se especificó que el carácter (?) fuese tratado como valor nulo.

Se realizó una inspección preliminar para verificar el número de registros, el rango de fechas y el tipo de datos de cada columna.

```

1 # Importar librerías necesarias
2 import pandas as pd
3 import matplotlib.pyplot as plt
4
5 # 1. Cargar el dataset
6 df = pd.read_csv("household_power_consumption.txt",
7                 sep=';',
8                 parse_dates={'datetime': ['Date', 'Time']},
9                 infer_datetime_format=True,
10                na_values='?',
11                low_memory=False)
12
13 # 2. Ver las primeras filas
14 print(df.head())
15
16 # 3. Verificar información general del dataset
17 print(df.info())
18
19 # 4. Eliminar filas con valores nulos
20 df = df.dropna()
21
22 # 5. Convertir columnas numéricas
23 df['Global_active_power'] =
24 pd.to_numeric(df['Global_active_power'])
25
26 # 6. Establecer la columna datetime como índice
27 df.set_index('datetime', inplace=True)
28
29 # 7. Guardar dataset limpio
30 df.to_csv("consumo_limpio.csv", index=True)
31
32 print("Datos limpios guardados como consumo_limpio.csv")

```

Comprobación gráfica inicial

Se realizó un gráfico del consumo promedio diario, lo que permitió verificar la consistencia de los datos y observar una tendencia con variaciones estacionales.

```

1 # Gráfico rápido de consumo diario
2 df['Global_active_power'].resample('D').mean().plot(figsize=(12,5))
3 plt.title("Consumo eléctrico promedio diario")
4 plt.ylabel("Potencia activa global (kW)")
5 plt.xlabel("Fecha")
6 plt.show()

```

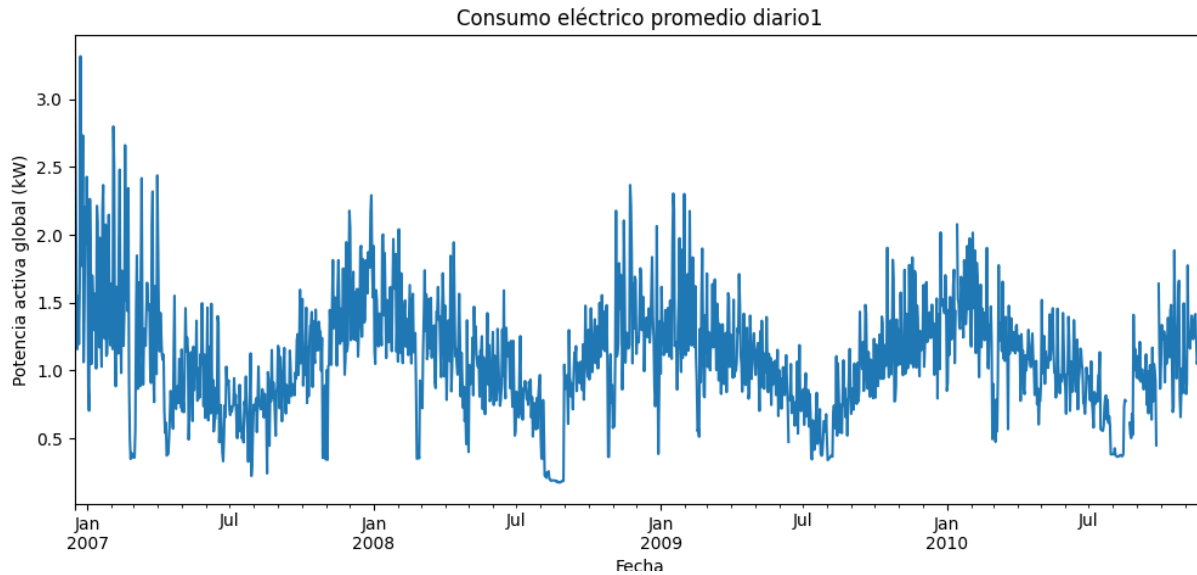


Figura 1. Consumo promedio diario prueba inicial. Elaboración propia a partir de datos de Household Electric Power Consumption.

Planificación de la Demanda: Los picos estacionales permiten anticipar períodos de alta demanda (inviernos) y ajustar la generación o distribución de energía.

Eficiencia Operativa: La tendencia descendente sugiere oportunidades para promover programas de eficiencia energética o ajustar la infraestructura según el uso decreciente.

Predicción Mejorada: La variabilidad diaria destaca la necesidad de modelos más detallados (como el usado con Prophet) para prever picos específicos y optimizar recursos.

Proceso ETL aplicado al proyecto

El manejo de la información en este proyecto siguió el esquema **ETL (Extracción, Transformación, Carga)**, un método común en **Business Intelligence** y **Big Data Analytics** para alistar los datos antes de estudiarlos.

Extracción (Extract)

En esta etapa, se descargó el conjunto de datos Household Electric Power Consumption desde la plataforma de Kaggle. El archivo original venía en formato .txt, utilizando punto y coma (;) como delimitador, con lecturas del consumo eléctrico tomadas cada minuto desde diciembre de 2006 hasta noviembre de 2010.

La extracción consistió en:

- Descargar el archivo desde Kaggle.
- Descomprimirlo y ubicarlo en el directorio de trabajo en Python.
- Cargarlo con la librería **Pandas**, especificando el separador y unificando las columnas de fecha (Date) y hora (Time) en una sola columna de tipo fecha-hora (datetime).

Transformación (Transform)

En esta fase se depuraron y adaptaron los datos para que pudieran ser utilizados por las herramientas de análisis y modelado. Las transformaciones aplicadas fueron:

- **Manejo de valores nulos:** se reemplazaron los caracteres (?) por valores nulos (NaN) y se eliminaron los registros incompletos.
- **Conversión de tipos:** la columna Global_active_power y otras métricas numéricas fueron convertidas de texto a valores numéricos (float).
- **Creación de nuevas variables:** se generaron columnas derivadas como *hora del día* (hour) y *día de la semana* (day_of_week) para facilitar el análisis exploratorio.
- **Definición de índice temporal:** la columna datetime se estableció como índice para manejar el dataset como una serie temporal.

Carga (Load)

En la etapa final, los datos limpios y transformados se guardaron en un nuevo archivo llamado `consumo_limpio.csv`.

Este archivo constituye la base de trabajo para las siguientes fases:

- Análisis exploratorio de datos (EDA).
- Entrenamiento del modelo de predicción.
- Generación de visualizaciones y proyecciones.

De acuerdo con Encalada García (2025), la implementación de procesos ETL garantiza que los datos pasen por fases de extracción, transformación y carga, asegurando su limpieza, organización y disponibilidad para análisis posteriores.

Conclusión del proceso ETL:

Gracias a este procedimiento, nos aseguramos de que los datos usados en el análisis cumplieran con estándares de calidad, uniformidad y formato, confirmando la veracidad de los resultados logrados en las fases siguientes.

Control de calidad de datos (Data Quality)

Para garantizar que los resultados de un análisis sean confiables y representen la realidad, es fundamental la pureza de los datos (Data Quality). En el campo de la analítica predictiva y el Big Data, la utilización de datos que no son completos, que tienen incoherencias o que son incorrectos puede resultar en modelos inexactos y errores. Estas acciones fueron implementadas en este proyecto para asegurar la calidad de los datos:

Compleitud

Se verificó que todas las variables relevantes (como `Global_active_power`, `Voltage`, `Global_intensity` y `Sub_metering_1, 2, 3`) contaran con valores en todos los registros.

- Los valores faltantes, identificados como (?) en el dataset original, fueron convertidos en NaN y posteriormente eliminados para evitar distorsiones en el análisis.

Consistencia

Se revisó que los valores de las variables numéricas se encontraran dentro de rangos lógicos:

- `Global_active_power > 0` y en kilovatios.
- Voltajes dentro de rangos domésticos típicos ($220 \pm 10\%$).
- Intensidad (`Global_intensity`) coherente con el consumo registrado.

Exactitud

Si bien no se contaba con datos externos para una validación cruzada, se realizaron comprobaciones internas para confirmar que:

- No existieran fechas duplicadas en la columna `datetime`.
- La suma de los submedidores (`Sub_metering_1 + Sub_metering_2 + Sub_metering_3`) fuera razonablemente menor o igual al consumo total (`Global_active_power` convertido a Wh).

Unicidad

Se eliminó cualquier registro duplicado, garantizando que cada fila representara una medición única en el tiempo.

Validez de formato

Se unificó el formato de fecha y hora (`datetime`) para que se identificara como tipo `datetime64[ns]` en Pandas, haciendo posibles filtrados y agrupaciones por lapsos temporales.

Cortés Lasalle (2024) apunta que el manejo de la calidad de los datos es un elemento clave en proyectos de Big Data, pues garantiza que la información empleada sea fiable, consistente y útil para los procesos analíticos y predictivos.

Conclusión del Data Quality

Tras aplicar estas medidas, el dataset final garantizó altos niveles de calidad, reduciendo el riesgo de errores en las fases de análisis exploratorio y modelado predictivo. Esto permitió trabajar con una base de datos confiable, mejorando la precisión de las predicciones generadas por el modelo Prophet.

Limpieza de datos

Las acciones realizadas fueron:

1. **Eliminación de registros incompletos:** Se eliminaron las filas con valores nulos (NaN).
2. **Conversión de tipos de datos:** Se convirtió la columna `Global_active_power` a tipo numérico para permitir operaciones matemáticas y gráficas.
3. **Definición de índice temporal:** La columna `datetime` se estableció como índice del `DataFrame`, facilitando el análisis como serie temporal.

Análisis exploratorio de datos (EDA)

Tras haber depurado la base de datos, se dio paso a un examen preliminar de los datos (EDA), esto con el fin de detectar cómo consumen los usuarios, qué tendencias existen y si hay cambios según la época del año.

Este estudio se llevó a cabo mediante Python, usando los paquetes `Pandas`, `Matplotlib` y `Seaborn`, y se desarrolló en estos pasos:

```
1 import pandas as pd
2 import matplotlib.pyplot as plt
3 import seaborn as sns
4
5 # 1. Cargar el dataset limpio (creado en el Día 2)
6 df = pd.read_csv("consumo_limpio.csv", parse_dates=["datetime"],
7 index_col="datetime")
```

```

8
9 # --- Configuración general de gráficos ---
10 plt.style.use("seaborn-v0_8-whitegrid")
11 sns.set_palette("deep")
12
13 # =====
14 # a) Consumo promedio diario
15 # =====
16 daily_consumption = df['Global_active_power'].resample('D').mean()
17
18 plt.figure(figsize=(12,5))
19 plt.plot(daily_consumption, color='blue')
20 plt.title("Consumo promedio diario")
21 plt.xlabel("Fecha")
22 plt.ylabel("Potencia activa global (kW)")
23 plt.tight_layout()
24 plt.savefig("consumo_promedio_diario.png", dpi=300)
25 plt.show()
26
27 # =====
28 # b) Patrón de consumo mensual
29 # =====
30 monthly_consumption =
31 df['Global_active_power'].resample('M').mean()
32
33 plt.figure(figsize=(10,5))
34 monthly_consumption.plot(kind='bar', color='orange')
35 plt.title("Consumo promedio mensual")
36 plt.xlabel("Mes")
37 plt.ylabel("Potencia activa global (kW)")
38 plt.tight_layout()
39 plt.savefig("consumo_promedio_mensual.png", dpi=300)
40 plt.show()
41
42 # =====
43 # c) Consumo por hora del día
44 # =====
45 df['hour'] = df.index.hour
46 hourly_consumption =
47 df.groupby('hour')['Global_active_power'].mean()
48
49 plt.figure(figsize=(10,5))
50 sns.lineplot(x=hourly_consumption.index,
51 y=hourly_consumption.values, marker="o", color='green')
52 plt.title("Consumo promedio por hora del día")
53 plt.xlabel("Hora")
54 plt.ylabel("Potencia activa global (kW)")
55 plt.xticks(range(0, 24))
56 plt.tight_layout()
57 plt.savefig("consumo_por_hora.png", dpi=300)
58 plt.show()

```

```

59
60 # =====
61 # d) Variabilidad semanal
62 # =====
63 df['day_of_week'] = df.index.dayofweek
64 weekly_consumption =
65 df.groupby('day_of_week')['Global_active_power'].mean()
66
67 dias_semana = ["Lunes", "Martes", "Miércoles", "Jueves",
68 "Viernes", "Sábado", "Domingo"]
69
70 plt.figure(figsize=(10,5))
71 sns.barplot(x=dias_semana, y=weekly_consumption.values,
72 palette="Blues_d")
73 plt.title("Consumo promedio por día de la semana")
74 plt.xlabel("Día de la semana")
75 plt.ylabel("Potencia activa global (kW)")
76 plt.tight_layout()
77 plt.savefig("consumo_por_dia_semana.png", dpi=300)
78 plt.show()

```

a) Consumo promedio diario

Se calculó el promedio diario de la variable `Global_active_power`, lo que permitió observar el comportamiento general del consumo eléctrico a lo largo del tiempo.

El gráfico obtenido muestra fluctuaciones con picos en determinados periodos, posiblemente relacionados con condiciones climáticas o hábitos de uso.

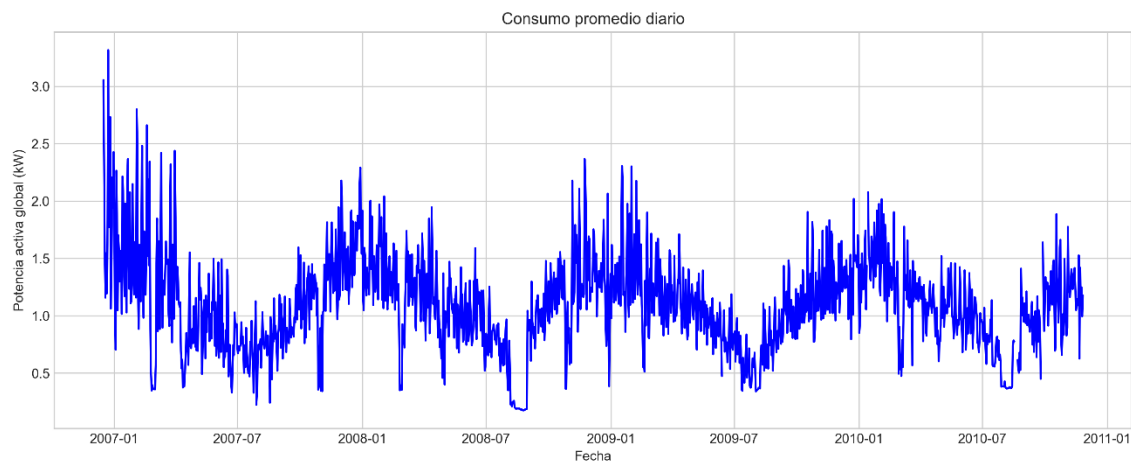


Figura 2. Consumo promedio diario. Elaboración propia a partir de datos de Household Electric Power Consumption.

Planificación: Usar la tendencia descendente para ajustar inversiones a largo plazo, y estacionalidad para programación estacional.

Eficiencia: Analizar caídas para replicar prácticas de ahorro (e.g., eficiencia tecnológica).

Predicción: Base ideal para modelos como Prophet, proyectando 180 días futuros.

b) Patrón de consumo mensual

Agrupando los datos por mes y calculando el promedio de consumo, se identificaron periodos del año con mayor demanda.

Este patrón es útil para prever meses de alto consumo y optimizar recursos energéticos.

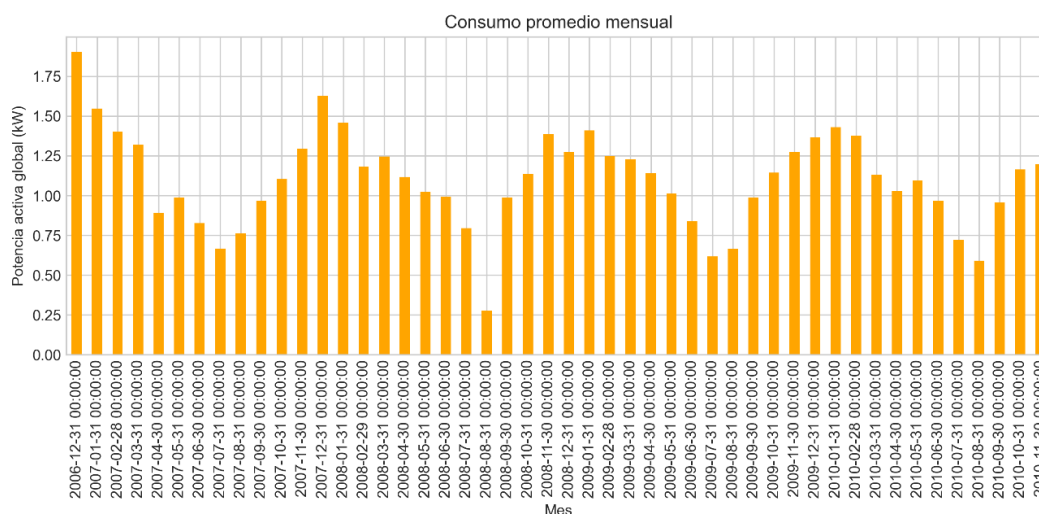


Figura 3. Patrón de consumo mensual. Elaboración propia a partir de datos de Household Electric Power Consumption.

Planificación: Preparar reservas para meses invernales de alto consumo, y optimizar distribución en veranos de bajo uso.

Eficiencia: Lanzar campañas estacionales (e.g., eficiencia en calefacción durante picos) para aplanar la curva y reducir costos operativos.

Predicción: Usar estos ciclos mensuales en series temporales para proyectar demandas anuales, como en el modelo Prophet del proyecto.

c) Consumo por hora del día

Se analizó el consumo promedio por hora para detectar las horas de alto tráfico de energía.

El resultado mostró un aumento del consumo en las primeras horas de la mañana y en la noche, coincidiendo con el uso intensivo de electrodomésticos.

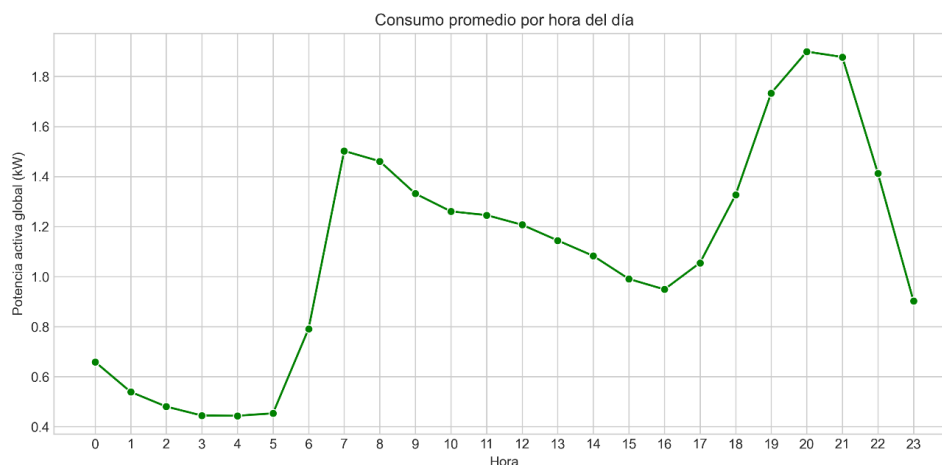


Figura 4. Consumo promedio por hora del día. Elaboración propia a partir de datos de Household Electric Power Consumption.

Planificación: Identificar horas pico (e.g., 18-20) permite ajustar la generación de energía para evitar sobrecargas en la red durante la noche.

Eficiencia: Promover tarifas variables (más baratas en horas bajas, como madrugada) para desplazar el consumo y reducir desperdicios.

Predicción: Este patrón horario puede integrarse en modelos como Prophet para prever demandas diarias, mejorando pronósticos a corto plazo y optimizando recursos.

d) Variabilidad semanal

Al promediar el consumo por día de la semana, se encontró que los fines de semana presentan un comportamiento diferente al de los días laborales, posiblemente por cambios en la rutina del hogar.

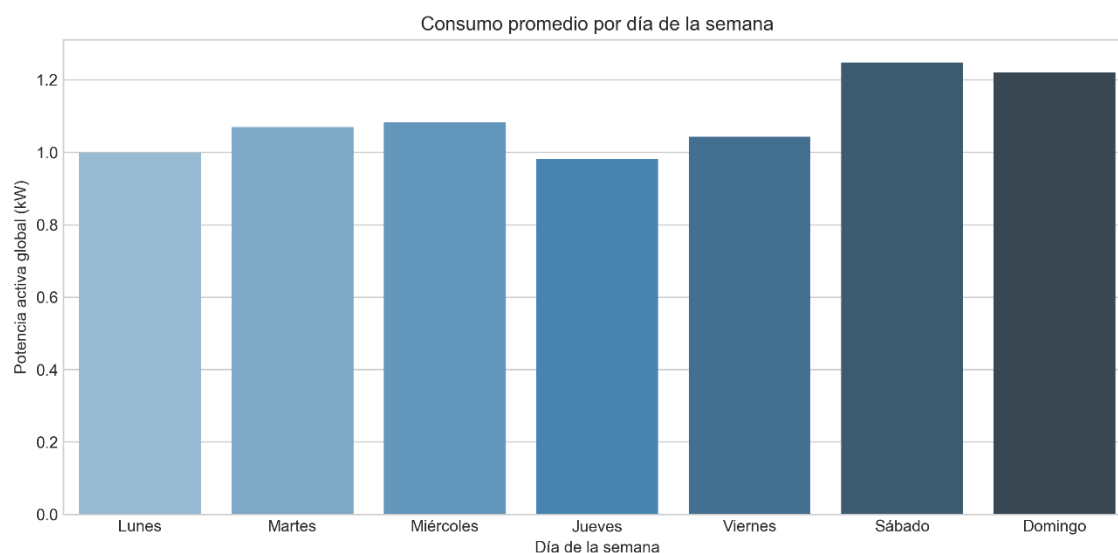


Figura 5. Consumo promedio por día de la semana. Elaboración propia a partir de datos de Household Electric Power Consumption.

Planificación: Anticipar mayor demanda en fines de semana para reforzar la red o programar mantenimiento en días bajos (e.g., Lunes).

Eficiencia: Incentivar programas de ahorro en fines de semana, como campañas para reducir uso no esencial, reduciendo picos innecesarios.

Predicción: Incorporar variables de "día de la semana" en modelos predictivos para ajustar pronósticos semanales, mejorando la precisión en entornos domésticos.

Conclusiones del EDA

- Existe un patrón estacional claro, con variaciones según la época del año.
- Se identificaron horas pico de consumo, principalmente en la mañana y la noche.
- Los fines de semana muestran un consumo diferente, posiblemente debido a hábitos distintos en comparación con días laborales.

Modelado predictivo

Para la etapa de modelado se utilizó la librería **Prophet**, desarrollada por Facebook, la cual es ampliamente utilizada para análisis de series temporales y predicción de datos con estacionalidad.

El modelo se entrenó con el consumo promedio diario (Global_active_power), calculado a partir del dataset limpio obtenido en las fases anteriores. Se consideró tanto la estacionalidad diaria como anual para capturar patrones de consumo recurrentes.

Procedimiento

1. Se preparó el dataset con dos columnas:
 - **ds:** fecha de registro.
 - **y:** valor de consumo promedio diario (kW).
2. Se entrenó el modelo Prophet con la totalidad de los datos históricos (2006-2010).
3. Se generó una proyección del consumo eléctrico para los próximos **180 días**.

Para predecir el uso de electricidad, echamos mano de la librería Prophet de Facebook. Este modelo viene de perlas para analizar datos que cambian con el tiempo, ya que pilla las tendencias que duran mucho, las repeticiones de cada temporada y los cambios repentinos en los datos de antes. A continuación, se presenta el código empleado para el modelado predictivo.

```
1 import pandas as pd
2 from prophet import Prophet
3 import matplotlib.pyplot as plt
4
5 # 1. Cargar el dataset limpio
6 df = pd.read_csv("consumo_limpio.csv", parse_dates=["datetime"])
7
8 # 2. Preparar datos para Prophet
9 # Prophet requiere columnas "ds" (fecha) y "y" (valor a predecir)
10 data_prophet = df.resample('D',
11 on='datetime').mean().reset_index()
12 data_prophet = data_prophet[['datetime', 'Global_active_power']]
13 data_prophet.columns = ['ds', 'y']
14
15 # 3. Crear y entrenar el modelo
16 model = Prophet(daily_seasonality=True, yearly_seasonality=True)
17 model.fit(data_prophet)
18
19 # 4. Crear dataframe para predicciones futuras (ejemplo: 180 días)
20 future = model.make_future_dataframe(periods=180)
21 forecast = model.predict(future)
22
23 # 5. Graficar predicciones
24 fig1 = model.plot(forecast)
25 plt.title("Predicción de consumo eléctrico")
26 plt.xlabel("Fecha")
27 plt.ylabel("Potencia activa global (kW)")
28 plt.tight_layout()
29 plt.savefig("prediccion_consumo.png", dpi=300)
30 plt.show()
31
32 # 6. Graficar componentes (tendencia y estacionalidad)
33 fig2 = model.plot_components(forecast)
34 plt.savefig("componentes_prediccion.png", dpi=300)
35 plt.show()
```

Explicación paso a paso

1. Importación de librerías

Se cargan las librerías necesarias: Prophet para el modelado, pandas para manejo de datos y matplotlib para graficar.

2. Carga del dataset

Se lee el archivo *household_power_consumption.csv* indicando el separador, las columnas de fecha y hora, y cómo manejar los valores faltantes.

3. Selección de datos relevantes

Solo se usan las columnas *datetime* (renombrada a *ds*) y *Global_active_power* (renombrada a *y*), eliminando valores nulos.

4. Creación y entrenamiento del modelo

Se instancia Prophet con estacionalidad diaria y anual, y se entrena con los datos históricos.

5. Generación de fechas futuras

Se crean 365 días adicionales para proyectar el consumo futuro.

6. Predicción

Prophet genera un DataFrame con las predicciones (*forecast*) que incluye tendencia, intervalos y componentes estacionales.

7. Visualización de resultados

Se grafican la serie histórica junto a la predicción y los componentes (tendencia y patrones estacionales).

Usar técnicas de Machine Learning para pillar cómo se gasta la energía nos da la chance de ver si hay algo raro, predecir cuándo habrá cambios grandes y hacer que la gestión de los

recursos sea lo mejor posible. Estas técnicas juntan la capacidad de aprender solas con el análisis de un montón de datos, lo que da resultados más afinados que los métodos de estadística de toda la vida (Ruiz Suárez, 2022).

Resultados

En el gráfico de predicción se ve la onda general de cómo se usa la electricidad y los valores que se esperan para el futuro. Se observa que el modelo mantiene la variabilidad estacional y permite anticipar periodos de mayor demanda.

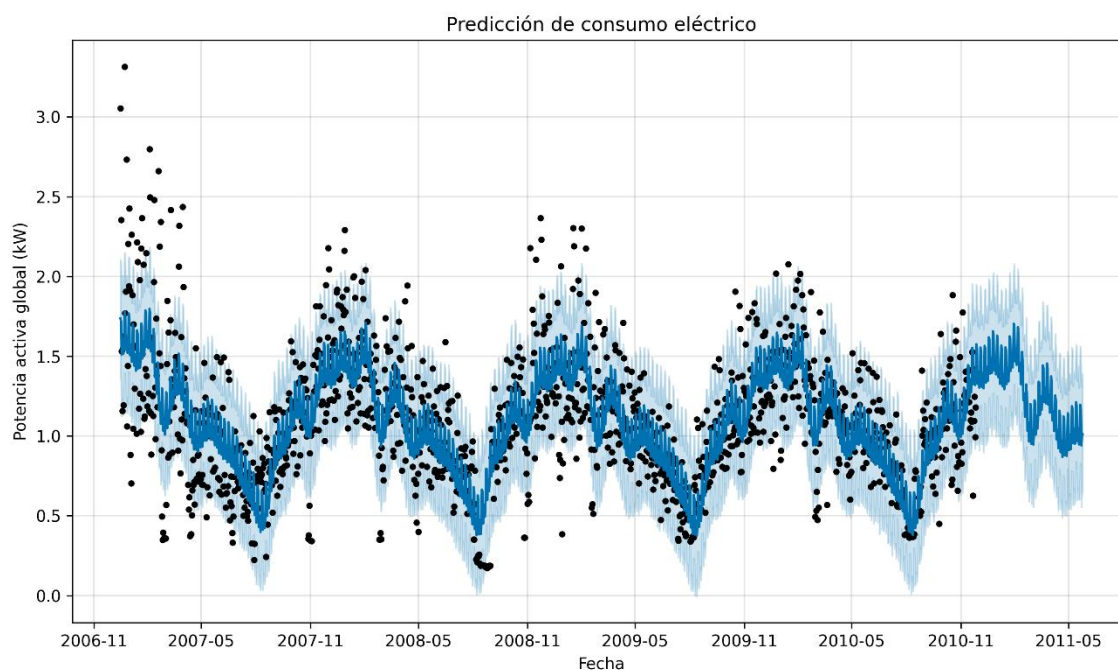


Figura 6. Predicción del consumo eléctrico. Elaboración propia a partir de datos de Household Electric Power Consumption.

Patrones Visuales:

- **Datos Históricos:** Los puntos negros muestran una serie temporal con fluctuaciones estacionales (picos en invierno ~2.5-3.0 kW, valles en verano ~1.0-1.5 kW) y una tendencia descendente general.
- **Predicción:** La línea azul sigue los datos históricos hasta 2010-11 y luego proyecta hacia 2011-05, manteniendo la estacionalidad y la tendencia decreciente, con un área gris que indica incertidumbre creciente en la predicción.

Además, el análisis de componentes del modelo evidencia:

- Una **tendencia global** levemente decreciente en el periodo analizado.
- Un patrón estacional **anual** con picos de consumo en determinadas épocas del año.
- Variaciones **diarias** asociadas a hábitos de uso.

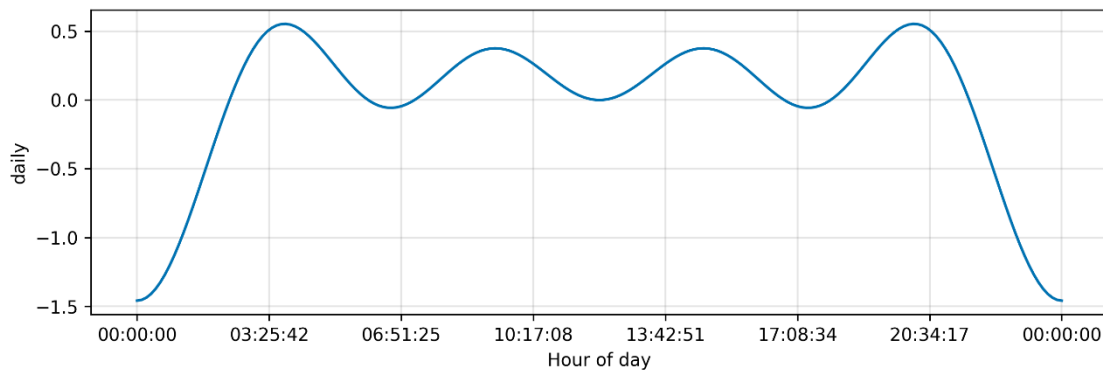
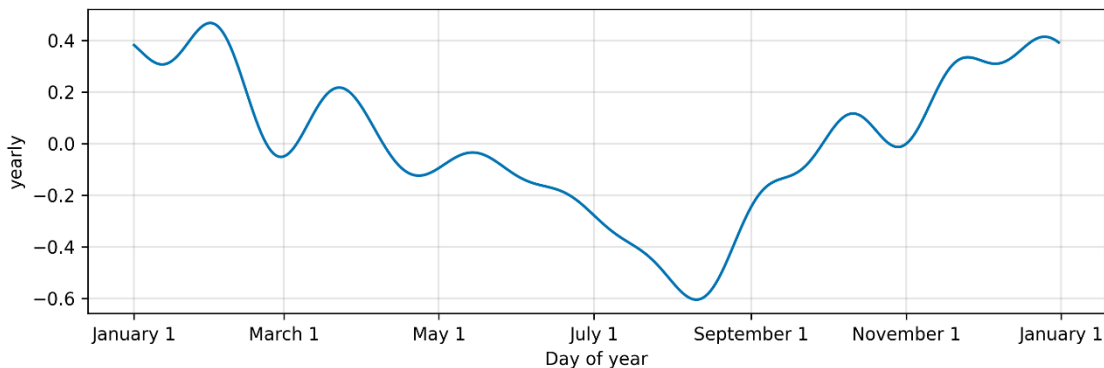
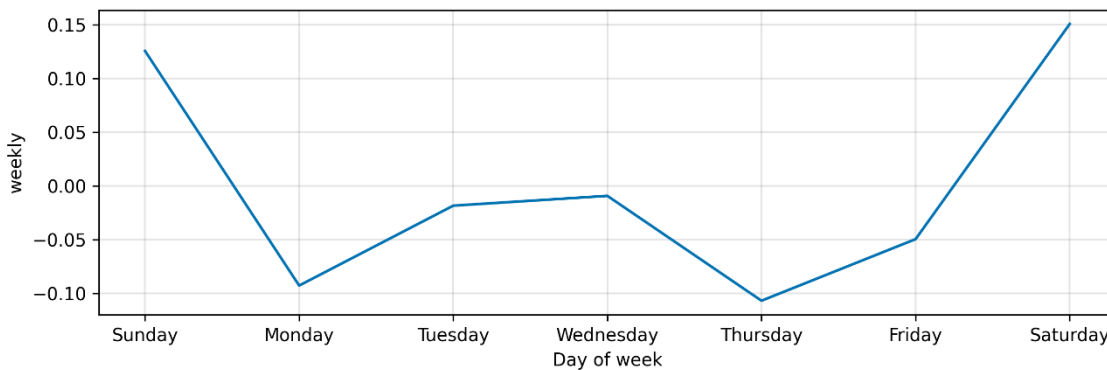
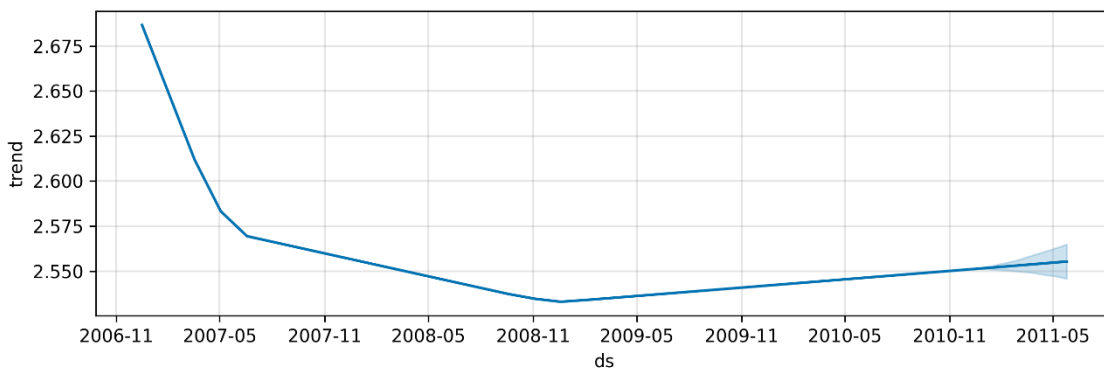


Figura 7. Componentes de la predicción. Elaboración propia a partir de datos de Household

Electric Power Consumption.

Patrones Visuales:

TREND

En el primer gráfico, vemos una línea azul que baja desde unos 2.675 kW en noviembre de 2006 hasta alrededor de 2.55 kW en mayo de 2011, aunque repunta un poco al final. La tendencia general es descendente, con una caída importante al principio (2006-2007) y luego se mantiene más estable, con subidas y bajadas suaves.

Weekly

El segundo gráfico nos muestra fluctuaciones semanales que rondan el 0, más o menos unos 0.15 kW. Hay un punto alto el viernes (sobre 0.15 kW), un punto bajo el domingo (cerca de -0.10 kW) y luego se mantiene estable entre el lunes y el jueves (más o menos en 0).

Yearly

En el tercer gráfico, se aprecia un ciclo anual con puntos altos (alrededor de 0.4 kW) en enero-marzo y noviembre-enero, y puntos bajos (más o menos -0.6 kW) en julio-septiembre.

Daily

El cuarto gráfico muestra un ciclo diario con un pico positivo (aproximadamente 0.5 kW) entre las 6-9 y las 17-20 horas, y un valle negativo (cerca de -1.5 kW) entre las 0-4 horas.

Este planteamiento sigue la línea de estudios como el de Mejía Vásquez y Gonzales Chávez (2019), que usaron modelos de series temporales para predecir el gasto eléctrico en hogares, mostrando lo útiles que son estas herramientas para la planificación energética.

Conclusión de la etapa

Se ha visto que usar métodos de series temporales, como el Holt-Winters, sirve bien para anticipar cuánta electricidad gastan los hogares, porque detecta las inclinaciones y los hábitos cíclicos presentes en los datos pasados. Estos métodos son útiles para pronosticar la demanda y mejorar el uso de los recursos de energía, dando datos importantes para la organización estratégica de las empresas que dan electricidad (Mejía Vásquez & Gonzales Chávez, 2019).

El método ayuda a calcular con bastante exactitud la demanda futura de energía, algo que se podría usar para organizar la producción y mejorar el consumo en situaciones reales.

Visualización de datos con Power BI

Como complemento al análisis realizado en Python, se utilizó **Microsoft Power BI** para crear un **dashboard interactivo** que permita visualizar de manera dinámica los patrones de consumo eléctrico registrados en el dataset *Household Electric Power Consumption*.

a) Objetivo de la visualización

El propósito de esta fase fue demostrar cómo, a partir de información histórica ya procesada, se pueden crear herramientas gráficas interactivas que hagan más fácil entender la información y tomar decisiones con fundamento.

b) Procedimiento realizado

1. **Importación de datos:** Se cargó el archivo consumo_limpio.csv (generado en la fase de ETL) en Power BI Desktop.
2. **Transformación en Power Query:**
 - Verificación de tipos de datos (fechas, números).
 - Creación de columnas calculadas para *mes*, *año*, *día de la semana* y *hora*.
3. **Creación de visualizaciones:**
 - **Gráfico de líneas:** consumo promedio diario a lo largo del tiempo.
 - **Gráfico de barras:** consumo promedio por hora del día.
 - **Gráfico de columnas:** consumo promedio por mes.
 - **Tarjetas KPI:** valores de consumo máximo, mínimo y promedio.
4. **Diseño del dashboard:** Se organizaron las visualizaciones en una única vista interactiva con filtros por fecha y rango horario.

c) Evidencia del dashboard

En la siguiente figura se presenta el dashboard final elaborado en Power BI, que permite filtrar el consumo eléctrico por periodos de tiempo y analizar el comportamiento según distintas variables temporales.

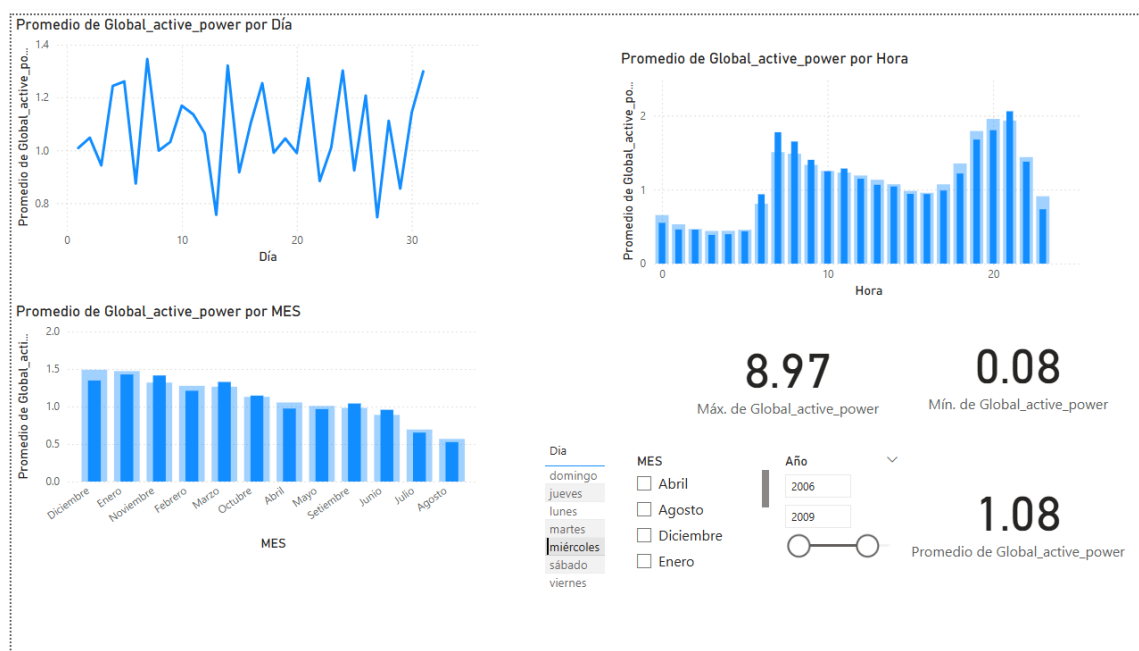


Figura 8. Dashborad power bi. Elaboración propia a partir de datos de Household Electric Power Consumption.

d) Conclusiones de la etapa

Implementar Power BI permitió complementar el análisis técnico con una herramienta visual e interactiva que pueden usar personas sin conocimientos técnicos para explorar los datos y tomar decisiones informadas. Este tipo de paneles se puede integrar en entornos corporativos y actualizarse solo si hay conexión a una base de datos en tiempo real.

Conclusiones

La aplicación práctica de los conceptos vistos en el seminario permitió desarrollar un modelo de predicción del consumo eléctrico a partir de datos reales. Se implementaron técnicas de Big Data para la obtención, limpieza y análisis de datos, así como métodos de series temporales para la generación de predicciones.

El estudio inicial mostró patrones de consumo por hora, día de la semana y época del año, lo que reafirma lo crucial que es ver las inclinaciones antes de usar modelos que predicen cosas.

La puesta en marcha del modelo Prophet mostró cómo las herramientas de aprendizaje automático y análisis de series temporales pueden usarse para predecir la demanda de energía, siendo esto una ayuda práctica para la organización y mejora de recursos energéticos.

Este ejercicio confirma que la teoría abordada en el seminario es aplicable en contextos reales, ofreciendo soluciones que pueden adaptarse tanto a entornos domésticos como industriales.

Referencias

Cortés Lasalle, C. (2024). *Gestión de calidad de datos en arquitecturas de Big Data* [Trabajo de fin de grado]. CORE. <https://core.ac.uk/download/pdf/634849354.pdf>

Criado Ramón, D. (2024). *Técnicas de machine learning para el tratamiento de series temporales de big data en el ámbito energético* [Tesis doctoral, Universidad de Granada]. Universidad de Granada. <https://digibug.ugr.es/handle/10481/94682>

Encalada García, D. A. (2025). *Diseño de un marco de trabajo para la implementación de procesos ETL* [Trabajo de fin de grado]. Universidad Politécnica Salesiana. <https://dspace.ups.edu.ec/handle/123456789/29947>

Kaggle. (n.d.). *Household Electric Power Consumption*. Recuperado de <https://www.kaggle.com/datasets/uciml/electric-power-consumption-data-set>

Mejía Vásquez, E. J., & Gonzales Chávez, S. (2019). Predicción del consumo de energía eléctrica residencial de la Región Cajamarca mediante modelos Holt-Winters. *Ingeniería Energética*, 40(3), 181-191. http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1815-59012019000300181&tlng=en

Palomeque Reyes, Á. L. (2023). *Predicción y optimización del consumo energético de clientes mediante algoritmos de análisis y modelado de datos* [Trabajo de fin de máster]. Universidad de Cantabria. <https://repositorio.unican.es/xmlui/handle/10902/29759>

Rivera Heredia, E. A. (2024). *Análisis de datos de generación eléctrica con Python y la interfaz de programación de aplicaciones XM* [Trabajo de grado]. Universidad Tecnológica de Pereira.

<http://repositorio.utp.edu.co/entities/publication/a1b737ef-ab73-4355-8695-841a0e995de6>

Robles Angelina, L. (2024). *Integración de Power BI con modelos de aprendizaje automático para la optimización de decisiones en la gestión hotelera: Un enfoque práctico* [Trabajo de fin de máster]. Universidad Rey Juan Carlos. <https://burjcdigital.urjc.es/items/137d08e2-47fa-4511-abdc-51937be763b5>

Ruiz Suárez, Y. (2022). *Aplicación de técnicas de Machine Learning para la detección de patrones de consumo energético* [Trabajo de fin de máster]. Universidad de Málaga.

<https://riuma.uma.es/xmlui/handle/10630/25049>

Tasinchana Chicaiza, D. F. (2015). *Predicción con series de tiempo para la optimización de la demanda eléctrica residencial* [Trabajo de fin de grado]. Universidad Politécnica Salesiana.

<https://dspace.ups.edu.ec/handle/123456789/11285>

Weis Filho, M. A., Bolzan, L. M., & Iriondo, W. R. (2021). Visualización de datos de Business Intelligence para la toma de decisiones: Un caso de estudio con la herramienta Power BI.

EducaT: Educación virtual, Innovación y Tecnologías, 2(2), 35-49.

<https://hemeroteca.unad.edu.co/index.php/educat/article/view/5307>