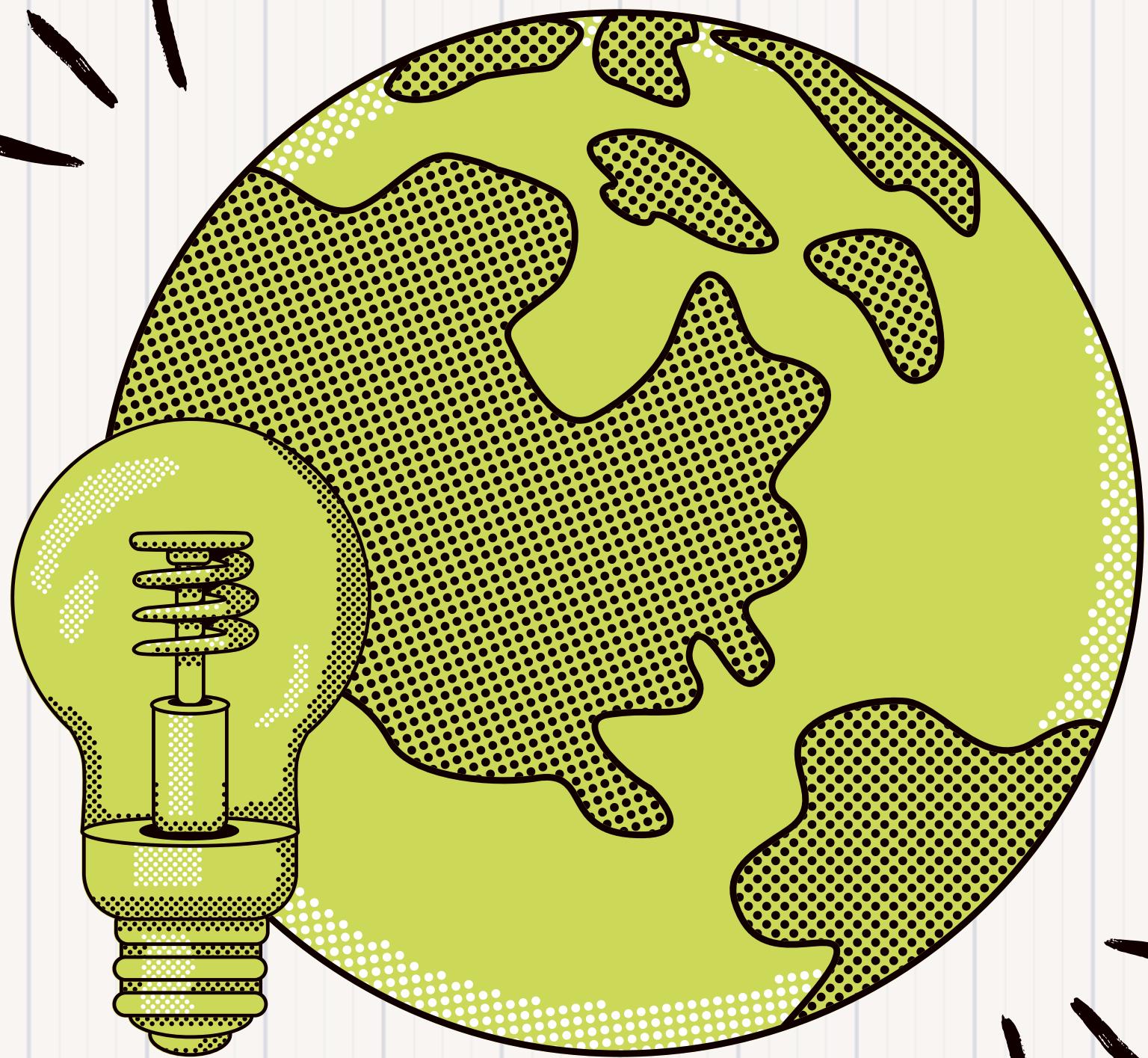


Grupo 02

Aplicación de Data Mining para el consumo energético

Data Mining



Contexto y Urgencia Energética

4/5

¹ De la energía primaria consumida:
Combustibles fósiles
(carbón, petróleo, gas).

²La IEA reportó **37.4 gigatoneladas** de CO₂ en 2023, impulsado por la recuperación económica.

CONTRADICCIÓN

Alta **demandas energéticas** proyectada al 2030 (especialmente en economías emergentes) y la **meta climática** de reducir las emisiones un **43%**.

“La energía es clave para el desarrollo económico, social y ambiental”.

1 Ritchie, H., & Rosado, P. (2017, 2 octubre). Fossil fuels. Our World In Data. <https://ourworldindata.org/fossil-fuels>

2 (International Energy Agency [IEA], 2024)

Objetivos y Metodología KDD

Identificar **patrones** y **factores** asociados al **consumo energético** y emisiones de CO₂ usando **Data Mining** y **Machine Learning**.

KDD:

selección, transformación, limpieza, minería de datos y evaluación de patrones.

PCA:

Técnica de reducción de dimensionalidad. Conserva las variables más informativas.

K-Means:

Algoritmo no supervisado. Agrupa datos en clústeres según similitud.

Holt-Winters:

Método de pronóstico para series temporales que asigna mayor peso a datos recientes.

- **Realizar limpieza** e imputación de valores faltantes en la base de datos de consumo energético.
- **Visualizar la evolución temporal** del consumo energético y de las emisiones de CO₂.
- **Explorar correlaciones** entre consumo energético, PBI, población y emisiones.
- **Identificar clústeres** de países con patrones energéticos similares.
- **Evaluar modelos** de predicción de consumo energético o emisiones a futuro.

Descripción del conjunto de datos

Medio de obtención:

kaggle

Dimensión inicial

- Número de filas: 22012
- Número de columnas: 114

DESAFÍOS

El preprocesamiento tiene el mayor impacto en la **precisión** de los modelos energéticos.

El dataset presentaba alta **dimensionalidad** y muchos **valores faltantes**.

The screenshot shows a Kaggle dataset page titled "World Energy Consumption". At the top, it displays the author's profile (PRALABH POUDEL), the date it was updated (2 years ago), the number of views (214), a download button, and a more options menu. Below the title, the description reads "Consumption of energy by different countries". There are four tabs: "Data Card" (selected), "Code (25)", "Discussion (3)", and "Suggestions (0)". Under "About Dataset", it says "Data on Energy by Our World in Data". To the right, there are sections for "Usability" (10.00) and "License" (Attribution 4.0 International (CC BY)).

Contenido:

- Países de todos los continentes
- Evolución temporal
- Patrones de comportamiento de energías (consumo, producción, etc.)



Descripción del conjunto de datos

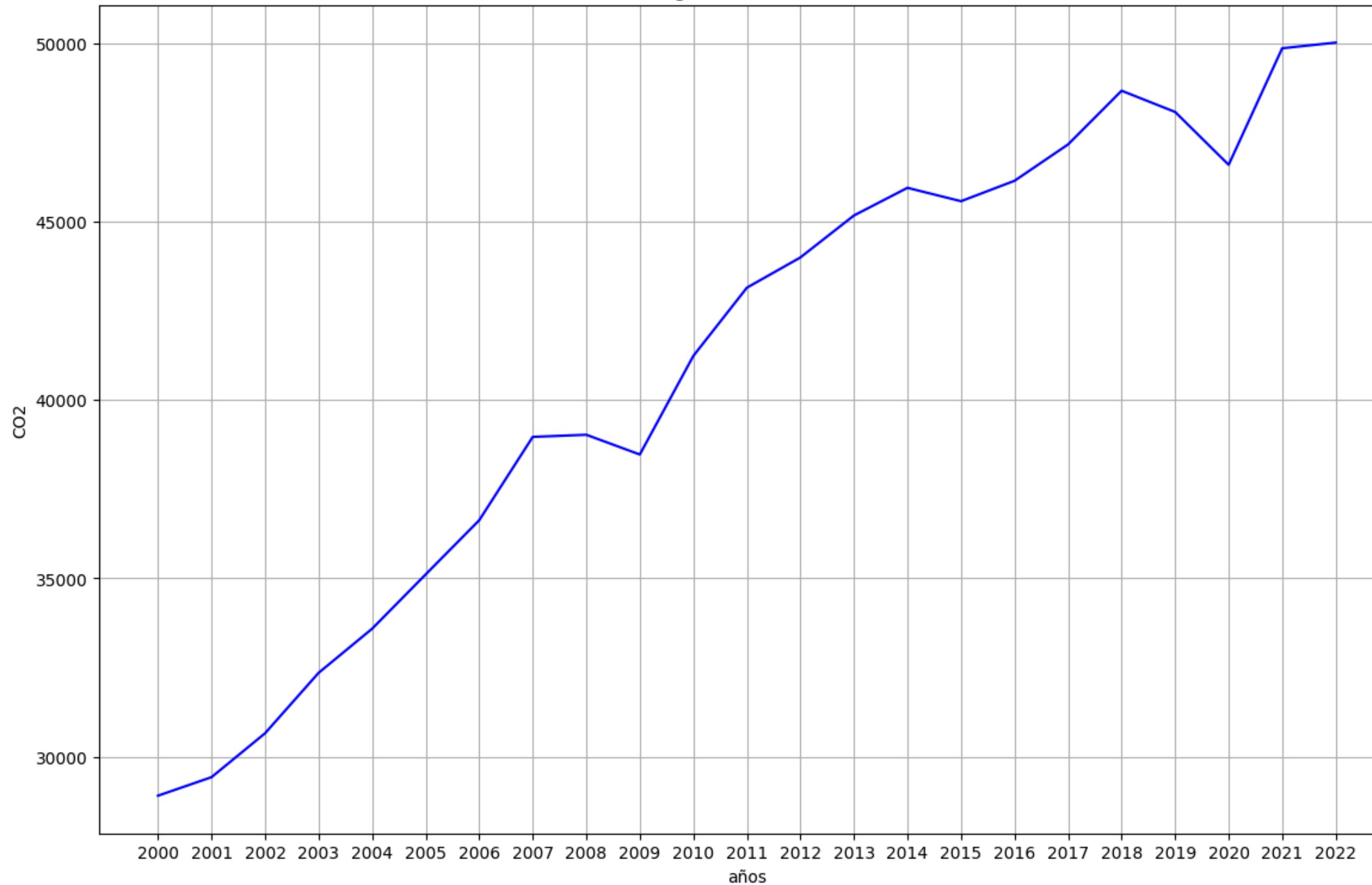
Diccionario de Datos

Variable	Descripción
country	país
year	Año de observación
iso_code	código ISO
population	población
gdp	Producto Bruto Interno
.....

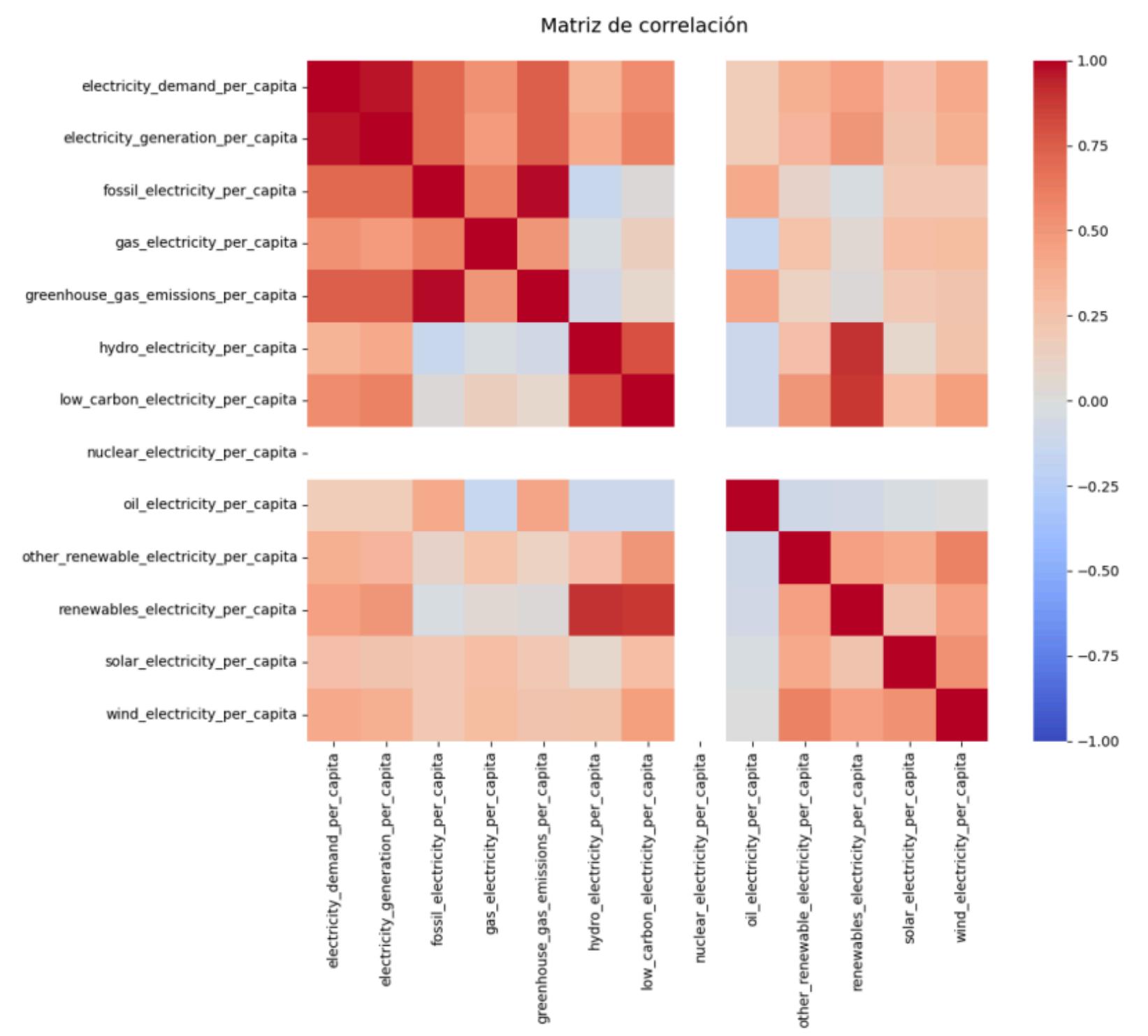
Variable	Descripción
fuente\cons\change\pct	Variación porcentual anual del consumo por tipo de energía.
fuente\cons\change\twh	Variación anual del consumo de por tipo energía.
fuente\elec\per\capita	Electricidad generada per capita por tipo de energía.
fuente\electricity	Electricidad total generada por tipo de energía.
fuente\share\elec	Participación del consumo de energía primaria medida como porcentaje del total de energía primaria por tipo de energía.

CO2 Generado

Emisiones de gases de efecto invernadero



Matriz de Correlación



	Variable_1	Variable_2	Correlacion
0	fossil_electricity_per_capita	greenhouse_gas_emissions_per_capita	0.98
1	electricity_demand_per_capita	electricity_generation_per_capita	0.96
2	hydro_electricity_per_capita	renewables_electricity_per_capita	0.91
3	low_carbon_electricity_per_capita	renewables_electricity_per_capita	0.89
4	hydro_electricity_per_capita	low_carbon_electricity_per_capita	0.79
5	electricity_generation_per_capita	greenhouse_gas_emissions_per_capita	0.75
6	electricity_demand_per_capita	greenhouse_gas_emissions_per_capita	0.74
7	electricity_demand_per_capita	fossil_electricity_per_capita	0.71
8	electricity_generation_per_capita	fossil_electricity_per_capita	0.71

Calidad y Preprocesamiento de Datos

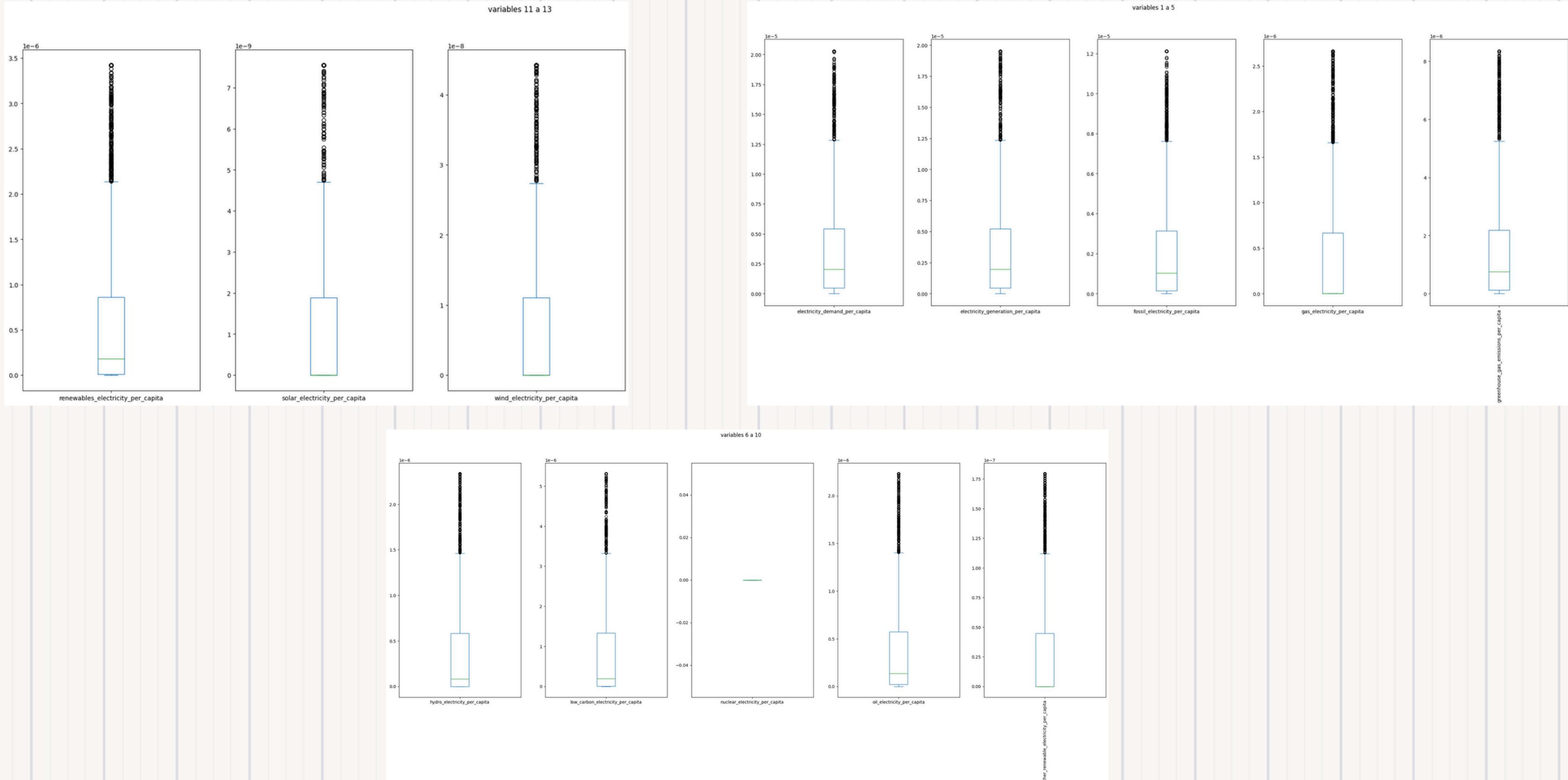
Imputación de Nulos

- **Restricción temporal:** Se trabajó únicamente con datos posteriores al año 2000
- **Interpolación lineal** (para variables energéticas con años consecutivos conocidos)
- **Imputación por propagación** para population, gdp y energía.
- **Eliminación** de columnas con más del **30% de nulos**.

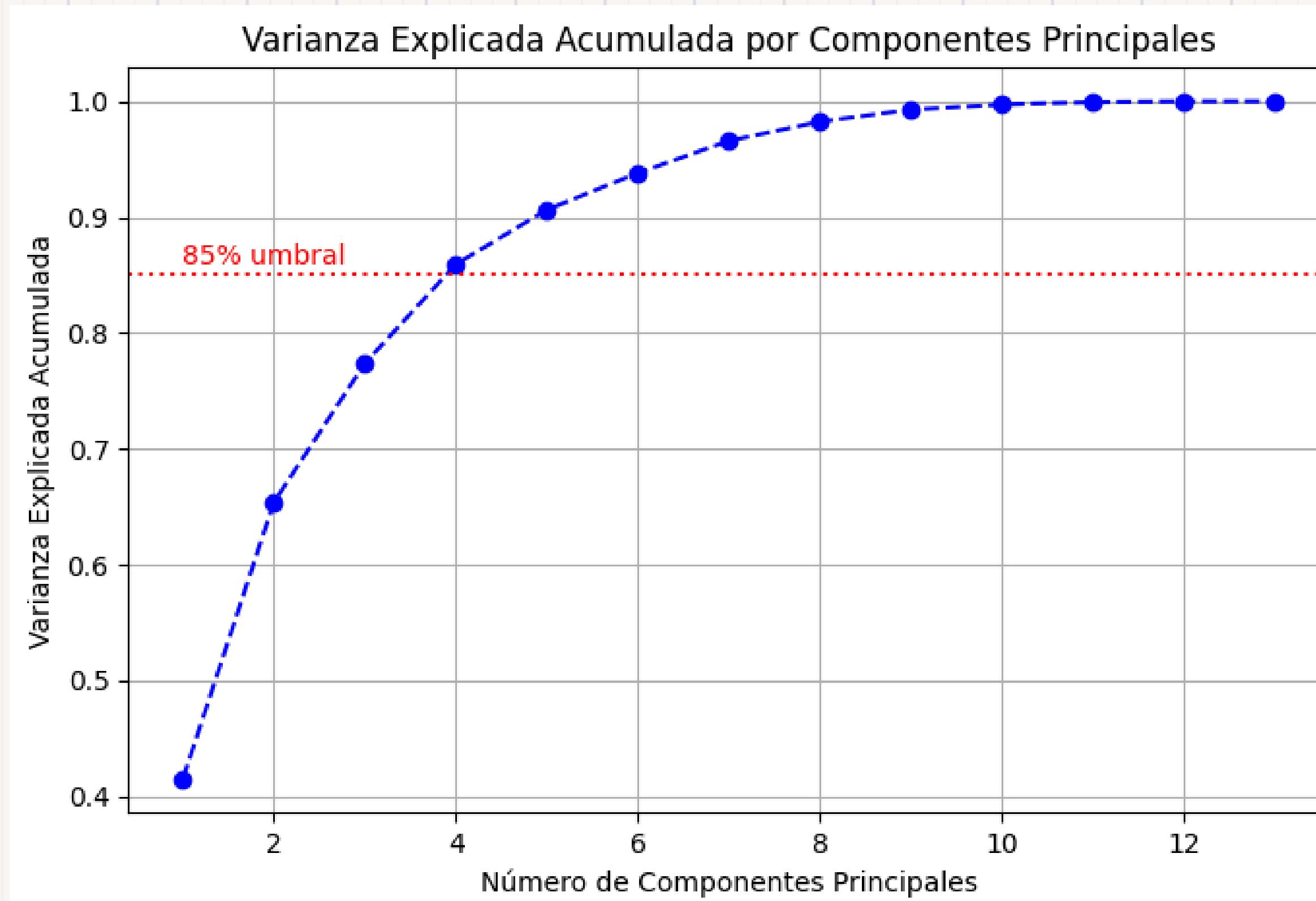
Transformación y Outliers

- **Transformación:** Estandarización (z-score) o reescalamiento (MinMax/Robusto) según la distribución.
- **Outliers:** Aplicación de **Winsorización** y métodos basados en el Rango Intercuartílico (**IQR**).

Distribución sin Outliers



Varianza explicada PCA



Clustering con K-Means (k=3)

Objetivo:

Condensar la información y **eliminar redundancias** para mejorar la eficiencia del clustering.

Criterio:

Mantener un umbral mínimo del **80%** de varianza explicada.

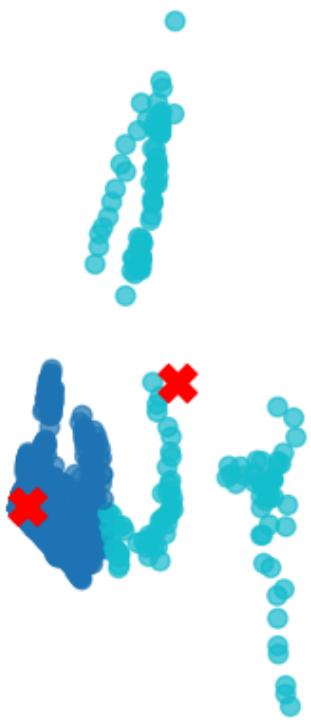
Resultado:

Se condensaron 44 variables en **4 Componentes Principales** (PCs), explicando el **82.6%** de la varianza acumulada.

Variables Dominantes:

PC1 (41.67%) dominado por *electricity demand* y *electricity generation*; **PC2 (23.02%)** dominado por *low carbon electricity*.

Clustering con K-Means ($k=3$)



Perfiles de Países Identificados

- **Clúster 1:** Países con alta producción y demanda eléctrica.
- **Clúster 2:** Países en transición con crecimiento moderado y mayor participación renovable.
- **Clúster 3:** Países con baja intensidad energética y menor producción eléctrica.

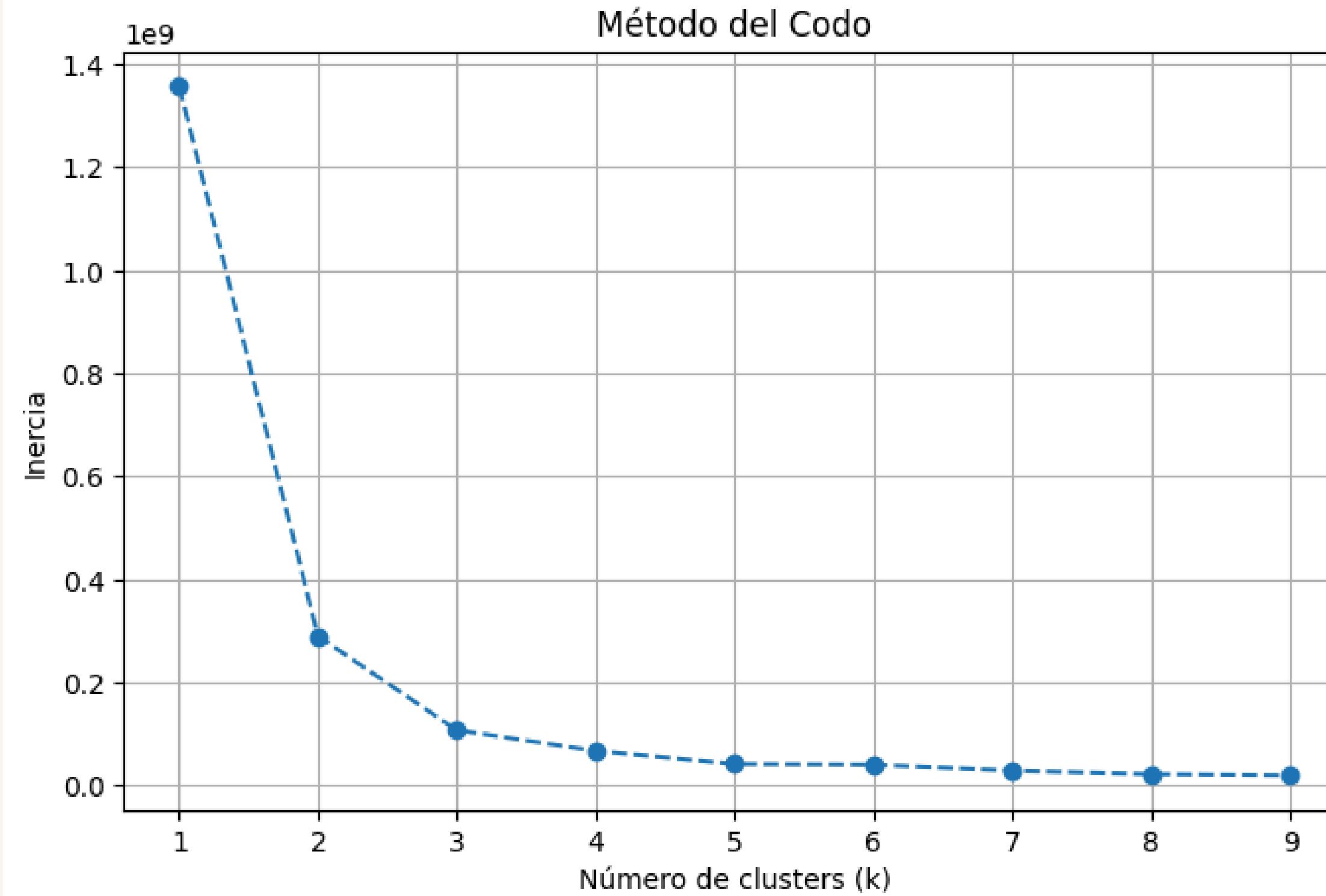
Objetivo

Agrupar **países** en perfiles **energéticos** diferenciados

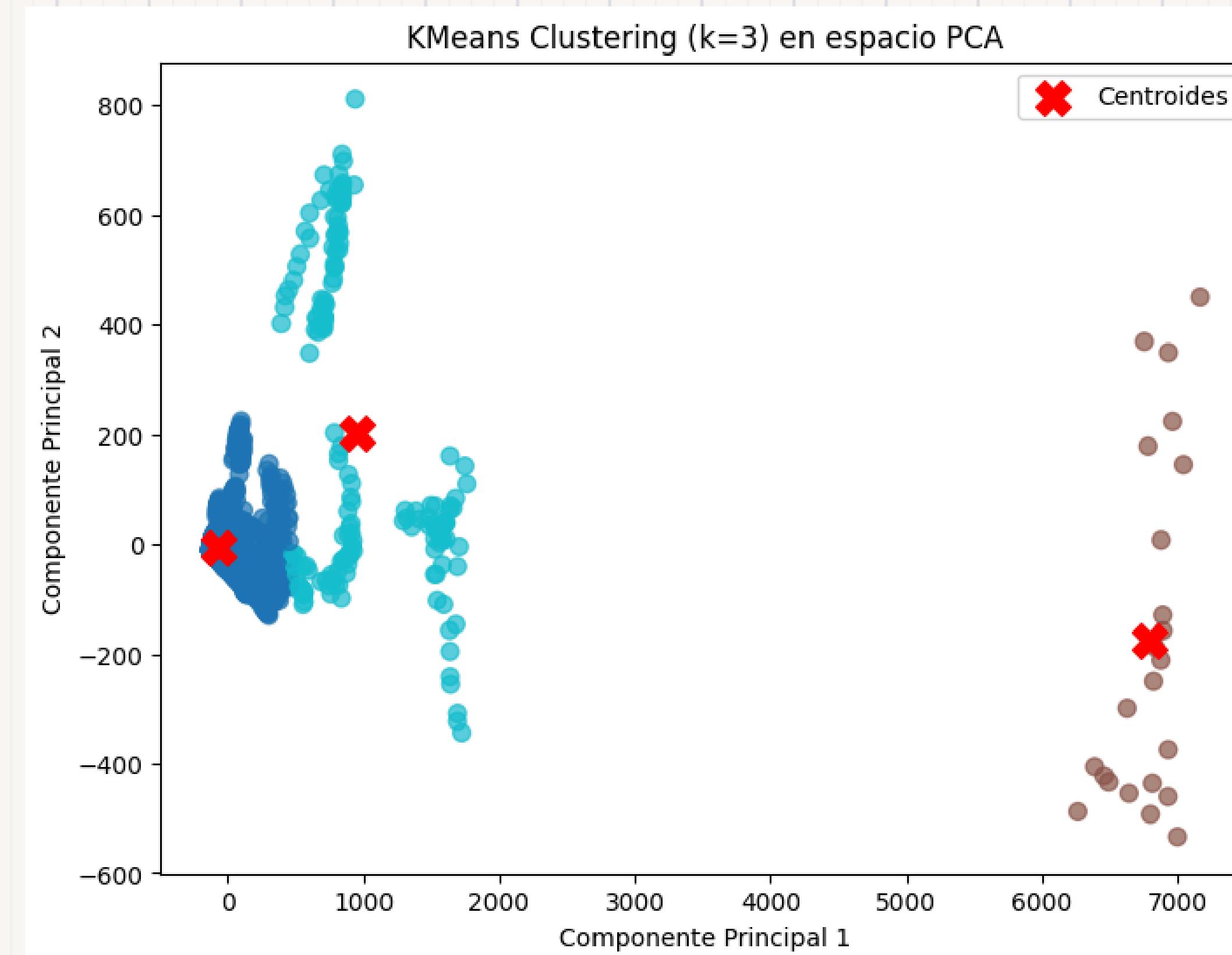
Determinación de k

El Método del Codo
(gráfica de Inercia vs. k)
indicó que el número óptimo de clústeres era
k=3

Clustering con K-Means (k=3)



Clustering con K-Means (k=3)



Pronóstico Temporal (Holt-Winters)

Modelo:

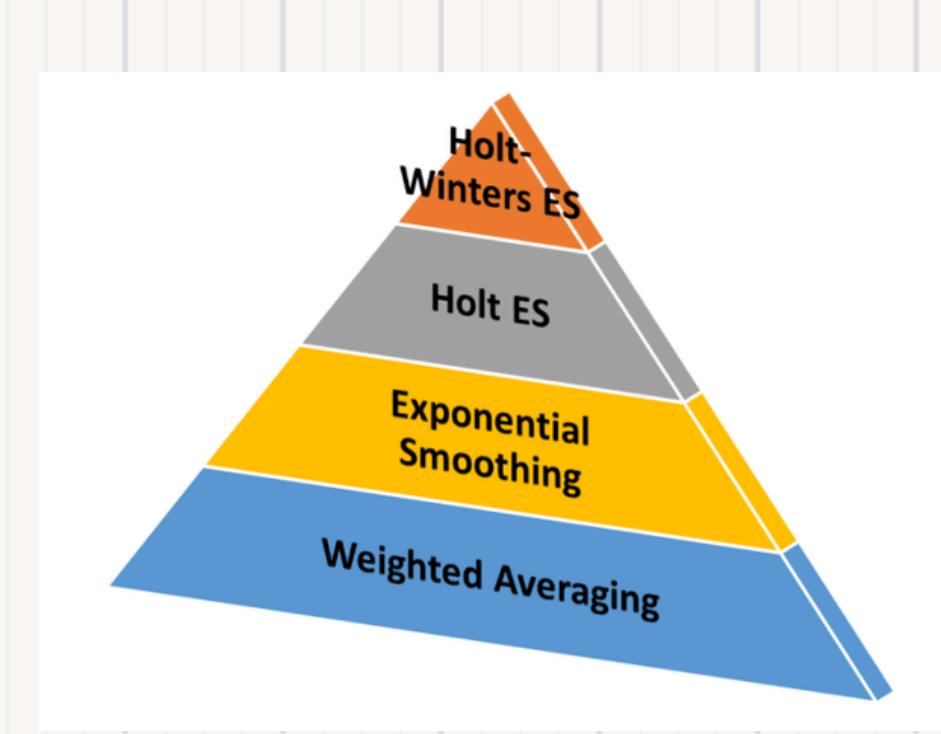
Suavizamiento Exponencial de **Holt-Winters** (sin estacionalidad, tendencia aditiva).

Uso:

Adecuado para capturar **tendencias suavizadas** en series de consumo energético.

Proyecciones:

Se generaron proyecciones para los **cinco años posteriores** al último dato disponible.



Consideraciones:

Se aplicaron **filtros** para evitar resultados artificiales.

Países con series completas (ej. China) mostraron **proyecciones coherentes**, mientras que otros sin evidencia histórica mostraron pronósticos cero.



Validación y Aportes Novedosos

Coincidencias (Validación de Tendencias)

Transición Energética: Se valida el incremento sostenido de la cuota de renovables (solar/eólica) a nivel mundial, alineado con informes de la IEA.

Patrones Regionales: Se confirma que la dinámica de consumo es heterogénea, con un desacoplamiento más rápido de fósiles en países desarrollados.

Aportes Novedosos del Proyecto

Cuantificación Granular: Mediante Data Mining, se cuantifica el impacto predictivo de variables menos estudiadas (ej. variaciones % en producción de gas o carbón).

Modelo Baseline País-Específico: Desarrollo de una herramienta de pronóstico simple, escalable y adaptable (proyecciones en GWh/TWh) para la planificación energética local.

Conclusiones Clave

Vínculo Económico-Energético:

Se confirma la fuerte relación entre crecimiento económico (PIB) y la demanda de energía.

Transición Irreversible:

La migración hacia fuentes renovables es una tendencia global en clara aceleración.

Heterogeneidad Regional:

El análisis segregado por país fue crucial para generar un modelo de pronóstico más útil para la toma de decisiones local.

Aporte Metodológico:

Se entrega un modelo predictivo de referencia (baseline) para la evaluación inicial de escenarios futuros de demanda eléctrica.

Limitaciones

Variables Incompletas:

La alta presencia de nulos limitó la inclusión de factores causales conceptualmente relevantes.

Sesgos Históricos:

El modelo simple puede no diferenciar entre anomalías temporales (ej. crisis, pandemia) y cambios de tendencia estructural.

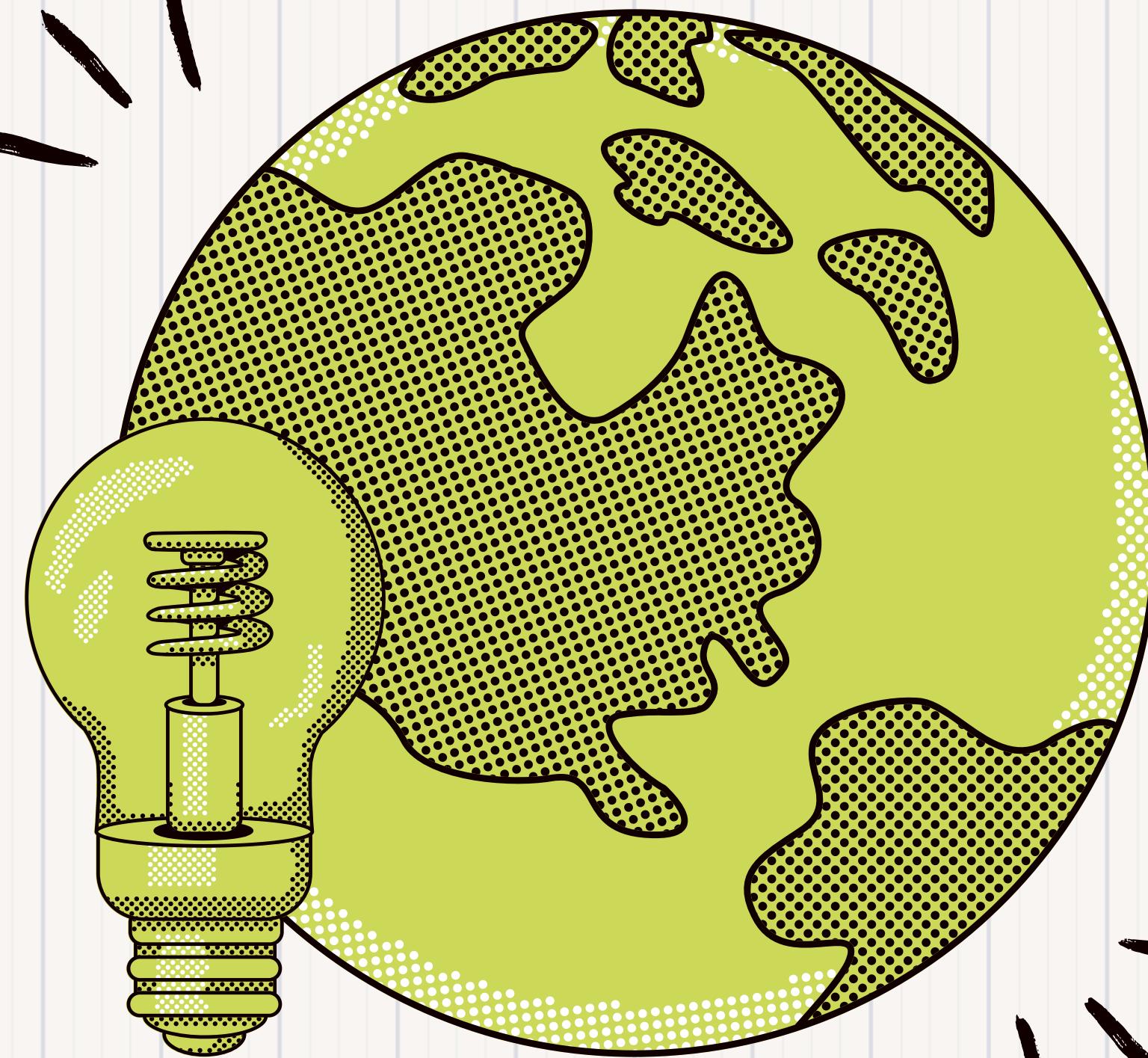
Recomendación:

Integrar variables de política regulatoria y usar modelos de ML más avanzados (Redes Neuronales) para mejorar la precisión a largo plazo.

Grupo 02

**¡Muchas
gracias!**

Data Mining



Bibliografía

1. Al-Essa, L. A., Ebrahim, E. A., & Mergiaw, Y. A. (2024). Bayesian regression modeling and inference of energy drivers. *Frontiers in Energy Research*.
2. Carpintero, O., & Frechoso, F. (2023). Energía, sostenibilidad y transición: nuevos desafíos y problemas pendientes. *Arbor*, 199(807), a687.
<https://doi.org/10.3989/arbor.2023.807001>
3. Duarte, O., Duarte, J., & Rosero-Garcia, J. (2024). Data Imputation in Electricity Consumption Profiles through Autoencoders. *Mathematics*.
4. Han, J., Kamber, M., & Pei, J. (2012). Data mining: Concepts and techniques (3rd ed.). Morgan Kaufmann.
5. International Energy Agency. (2023). World Energy Outlook 2023. <https://www.iea.org/reports/world-energy-outlook-2023>
6. International Energy Agency. (2024). Global Energy and \$CO₂\$ Status Report 2024. <https://www.iea.org/reports/global-energy-review-2025/co2-emissions>
7. International Renewable Energy Agency (IRENA). (2025, 26 de marzo). Record-Breaking Annual Growth in Renewable Power Capacity.
<https://www.irena.org/News/pressreleases/2025/Mar/Record-Breaking-Annual-Growth-in-Renewable-Power-Capacity>
8. IPCC. (2023). AR6 Synthesis Report: Climate Change 2023.
9. Jolliffe, I. (2011). Principal component analysis. In *International encyclopedia of statistical science* (pp. 1094-1096). Springer.
10. Li, K., Ma, Z., Robinson, D., Lin, W., & Li, Z. (2020). A data-driven strategy to forecast next-day electricity usage and peak electricity demand of a building portfolio using cluster analysis, Cubist regression models and Particle Swarm Optimization. *Journal of Cleaner Production*, 273, 123115. <https://doi.org/10.1016/j.jclepro.2020.123115>
11. Liu, X., Ding, Y., Tang, H., & Xiao, F. (2021). A data mining-based framework for the identification of daily electricity usage patterns and anomaly detection in building electricity consumption data. *Energy & Buildings*, 231, 110601. <https://doi.org/10.1016/j.enbuild.2020.110601>
12. Mallala, B., Ahmed, A. I. U., Pamidi, S. V., Faruque, M. O., & Reddy, R. (2025). Forecasting global sustainable energy from renewable sources using random forest algorithm. *Results in Engineering*, 25, 103789.
13. Mateos, G., & Giannakis, G. B. (2013). Load Curve Data Cleansing and Imputation via Sparsity and Low Rank.
14. Pan, H., Yin, Z., & Jiang, X. (2022). High-Dimensional Energy Consumption Anomaly Detection: A Deep-Learning Based Method. *Energies*.
15. Parhizkar, T., Rafieipour, E., & Parhizkar, A. (2020). Evaluation and improvement of energy consumption prediction models using principal component analysis based feature reduction. *Journal of Cleaner Production*, 279, 123866. <https://doi.org/10.1016/j.jclepro.2020.123866>
16. Ritchie, H., & Rosado, P. (2017, 2 de octubre). Fossil fuels. *Our World In Data*. <https://ourworldindata.org/fossil-fuels>
17. Tambini, K., & Vergara, V. (2024). El impacto del consumo de energía en el crecimiento económico: Un análisis con datos de panel. *Desafíos Economía y Empresa*, 004, 99–114.
<https://doi.org/10.26439/ddee2024.n04.6247>
18. Taylor, J. W. (2003). Short-term electricity demand forecasting using double seasonal exponential smoothing. *Journal of the Operational Research Society*, 54(8), 799–805.
<https://doi.org/10.1057/palgrave.jors.2601589>
19. World energy consumption. (2023, 26 de noviembre). World energy consumption [Dataset]. Kaggle. <https://www.kaggle.com/datasets/pralabhpoudel/world-energy-consumption>
20. Zhang, L., Ge, R., & Chai, J. (2019). Prediction of China's energy consumption based on robust principal component analysis and PSO-LSSVM optimized by the Tabu search algorithm. *Energies*, 12(1), 196. <https://doi.org/10.3390/en12010196>
21. Zhao, T., Sun, Y., Chai, Z., & Li, K. (2022). An outlier management framework for building performance data and its application to the power consumption data of building energy systems in non-residential buildings. *Journal of Building Engineering*, 65, 105688. <https://doi.org/10.1016/j.jobe.2022.105688>