



# Energías renovables en la generación eléctrica en Colombia

Análisis exploratorio

---

Alba López    Ulises Linares    Luis Bertel

Inteligencia Artificial - Nivel Exploratorio

# Tabla de contenido

1. Contexto del problema

2. Materiales y métodos

3. Resultados

Análisis exploratorio

Análisis supervisado

Análisis no supervisado

4. Conclusiones

## Contexto del problema

---

# Retos de Colombia en cuanto a generación de energía

Colombia se encuentra en una etapa clave de transformación de su matriz energética, impulsada por la necesidad de diversificar las fuentes de generación eléctrica, reducir la dependencia de recursos hidroeléctricos y fósiles, y avanzar hacia un sistema energético más sostenible y resiliente frente al cambio climático. Históricamente, el país ha contado con una alta participación de la generación hidroeléctrica —superior al 60—, lo que ha permitido mantener una baja intensidad de emisiones de carbono en el sector eléctrico en comparación con otras economías de América Latina y el mundo. No obstante, esta dependencia ha expuesto al sistema a riesgos asociados con la variabilidad climática, especialmente durante los períodos de sequía causados por fenómenos como El Niño, que comprometen la estabilidad del suministro energético.

## Objetivo específico

Analizar la evolución y el aporte de las fuentes de energía renovable en la generación eléctrica en Colombia entre los años 2010 y 2025, con base en datos estadísticos comparativos, a fin de identificar tendencias, avances y desafíos en el marco de la transición energética del país.

¿Cómo ha evolucionado el aporte de las energías renovables en la generación eléctrica en Colombia entre 2014 y 2025, y qué factores explican su desempeño?

## Materiales y métodos

---

Este estudio se enmarca dentro del ciclo de vida de un proyecto de Machine Learning, específicamente en las etapas de adquisición de datos, limpieza, análisis exploratorio, visualización, análisis multidimensional y modelos predictivos. Se adoptan herramientas y metodologías propias de la ciencia de datos para estructurar un análisis riguroso, reproducible y orientado a la extracción de conocimiento a partir de datos energéticos. El enfoque se limita exclusivamente al caso de Colombia [García and Molina, 2018].



El análisis se basa en un conjunto de datos inicialmente obtenido desde el portal Kaggle, el cual contiene información mensual de generación eléctrica clasificada por tipo de fuente energética, país, año y mes, expresada en gigavatios- hora (GWh). Esta base cubre el período comprendido entre 2010 y 2022. Para extender el alcance temporal hasta 2025, se aplicaron técnicas de web scraping sobre fuentes oficiales y sitios especializados, lo que permitió actualizar el conjunto de datos con los registros más recientes disponibles. El dataset final fue consolidado y almacenado en formato CSV, lo cual facilitó su manipulación y análisis mediante herramientas de programación. Se realizó una curaduría manual y automatizada para garantizar la coherencia, completitud y calidad de los datos utilizados.

## Detalle del conjunto de datos

Variable	Descripción
COUNTRY	Nombre del país al que corresponde el registro.
CODE_TIME	Código temporal en formato abreviado que indica el mes y año (por ejemplo, JAN2010).
TIME	Representación en texto del mes y año (por ejemplo, January 2010).
YEAR	Año correspondiente al registro.
MONTH	Mes correspondiente al registro, en formato numérico (1–12).
MONTH_NAME	Nombre del mes correspondiente al registro.
PRODUCT	Tipo de fuente energética utilizada para la generación eléctrica (e.g., Hydro, Wind, Solar, etc.).
VALUE	Cantidad de electricidad generada en gigavatios-hora (GWh).

**Tabla 1:** Categorías del conjunto de datos

## Detalle del conjunto de datos - continuación

Variable	Descripción
DISPLAY_ORDER	Indicador numérico utilizado para el ordenamiento visual de los productos.
yearToDate	Acumulado de generación eléctrica en GWh desde el inicio del año hasta el mes actual.
previousYearToDate	Acumulado de generación eléctrica para el mismo periodo del año anterior.
share	Porcentaje de participación del producto en la generación total del país (en formato decimal, por ejemplo 0.12 = 12 %).

**Tabla 2:** Categorías del conjunto de datos

El procesamiento y análisis de los datos se llevaron a cabo en el lenguaje de programación Python, utilizando el entorno Jupyter Notebook por su versatilidad en la integración de código, visualizaciones y documentación. Las principales bibliotecas empleadas fueron:

- **Pandas:** para la manipulación y estructuración tabular de los datos.
- **NumPy:** para el manejo de operaciones numéricas.
- **Matplotlib y Seaborn:** para la creación de visualizaciones descriptivas y comparativas.
- **Scikit-learn:** para métodos de reducción de dimensiones, como PCA.

# Proceso metodológico (1)

## 1. Adquisición y consolidación de datos:

- Descarga inicial del dataset desde Kaggle.
- Actualización de datos mediante técnicas de web scraping para ampliar el rango temporal hasta 2025.
- Fusión de fuentes y estandarización del formato general.

## 2. Limpieza de datos:

- Revisión de datos faltantes, valores atípicos o inconsistentes.
- Conversión de tipos de datos y normalización de etiquetas (por ejemplo, estandarización de nombres de fuentes energéticas).
- Filtro de registros para conservar únicamente la información correspondiente a Colombia.

## 3. Análisis exploratorio de datos (EDA):

- Estadísticas descriptivas generales sobre la generación eléctrica total y por tipo de fuente.
- Evaluación de la evolución temporal de la participación de energías renovables frente a no renovables.

## Proceso metodológico (2)

### 4. Visualización de tendencias:

- Representación gráfica de series temporales, porcentajes de participación, acumulados anuales y estacionales.
- Comparación de patrones interanuales y detección de puntos de inflexión relevantes.

### 5. Análisis multidimensional:

- Aplicación de técnicas para explorar las relaciones entre variables como tipo de generación, volumen mensual, variabilidad estacional, y evolución anual.
- Uso de gráficos de calor, diagramas de dispersión múltiple o métodos de reducción de dimensiones (como PCA), en caso de ser necesarios.

### 6. Análisis supervisado y no supervisado:

- Análisis de tendencia usando regresión lineal.
- Uso de cluster para agrupar generación por año.

## Resultados

---

## Estructura de datos utilizada

```
<class 'pandas.core.frame.DataFrame'>  
Index: 3104 entries, 46557 to 212133  
Data columns (total 4 columns):  
#    Column    Non-Null Count  Dtype  
---  -  
0    YEAR       3104 non-null   int64  
1    MONTH      3104 non-null   int64  
2    PRODUCT    3104 non-null   object  
3    VALUE      3104 non-null   float64  
dtypes: float64(1), int64(2), object(1)  
memory usage: 121.2+ KB
```



# Cantidad de energía clasificada según tipo

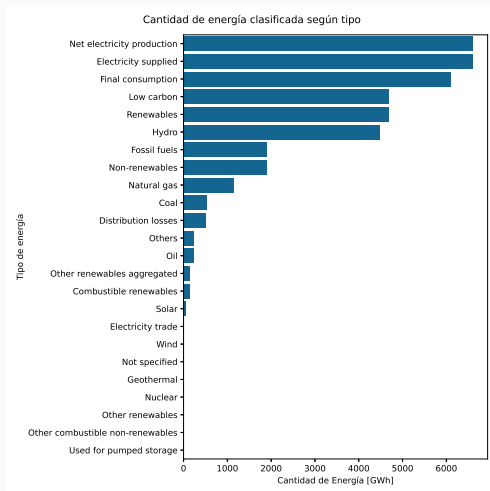


Figura 1: Cantidad de energía clasificada por tipo

# Evolución de la producción neta de energía en Colombia

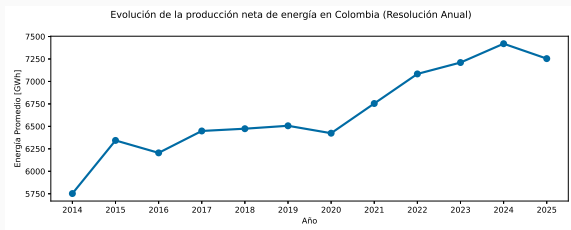


Figura 2: Evolución de la producción neta de energía en Colombia

# Producción neta mensual de energía en Colombia

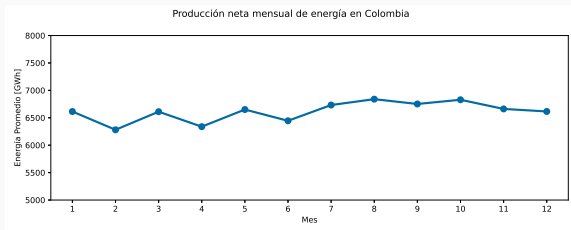


Figura 3: Producción neta mensual de energía en Colombia

# Matriz energética de Colombia

```
PRODUCT
Hydro          601415.683695
Natural gas    153592.481385
Coal           71158.873666
Others         31527.442435
Solar          7138.600132
Wind           629.372654
Nuclear        0.000000
Name: VALUE, dtype: float64
```

# Matriz energética de Colombia

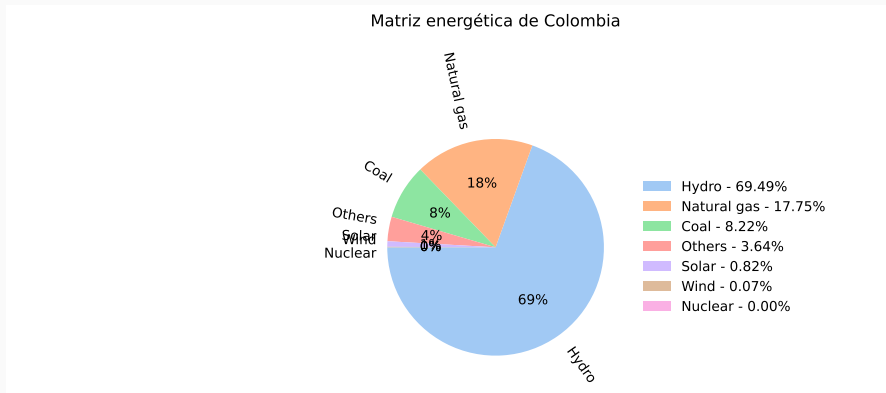


Figura 4: Matriz energética de Colombia

# Energía según tipo de generación

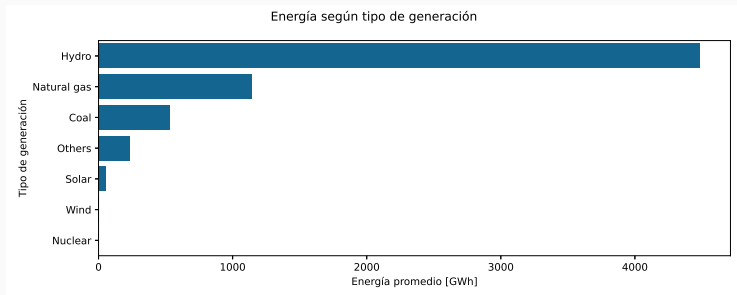
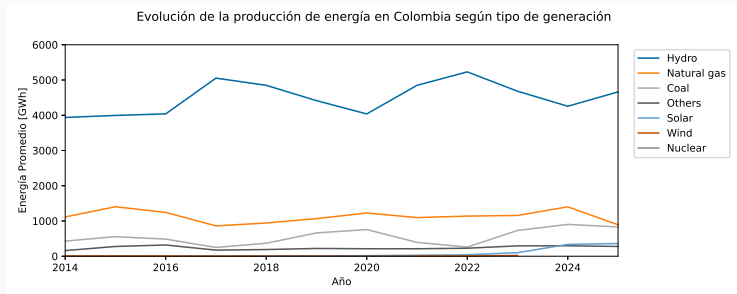


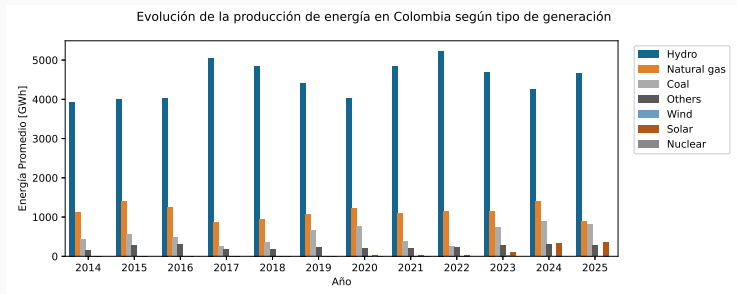
Figura 5: Energía según tipo de generación

# Evolución de la producción de energía en Colombia según tipo de generación



**Figura 6:** Evolución de la producción de energía en Colombia según tipo de generación

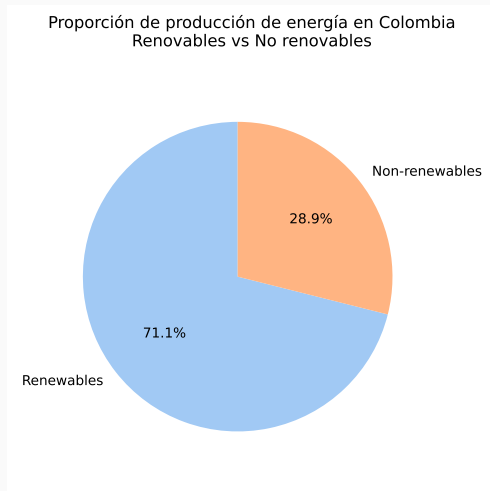
# Evolución de la producción de energía en Colombia según tipo de generación



**Figura 7:** Evolución de la producción de energía en Colombia según tipo de generación

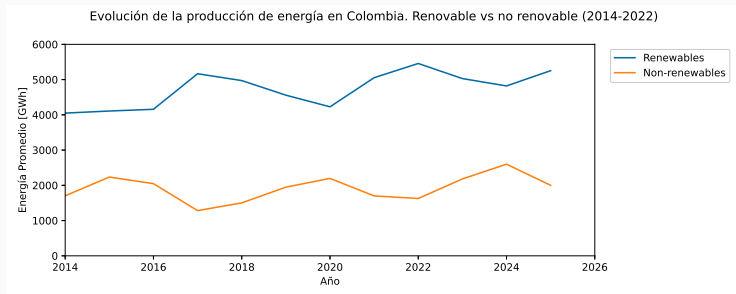


# Proporción energía renovables vs No renovables



**Figura 8:** Proporción de producción de energía en Colombia Renovables vs No renovables

# Evolución energía renovable vs no renovable (2014-2025)



**Figura 9:** Evolución de la producción de energía en Colombia. Renewable vs no renovable (2014-2025)

# Producción de energía de fuentes Renovables

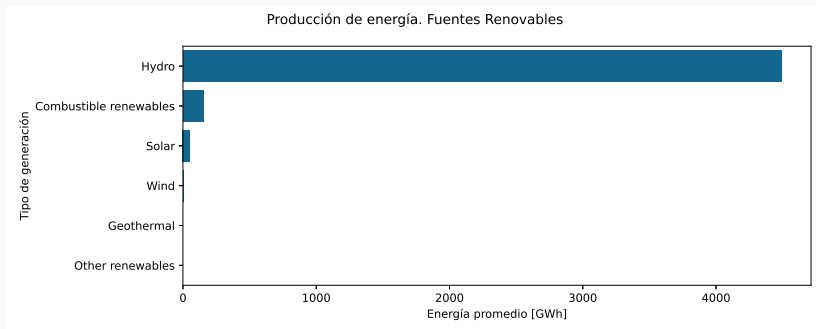
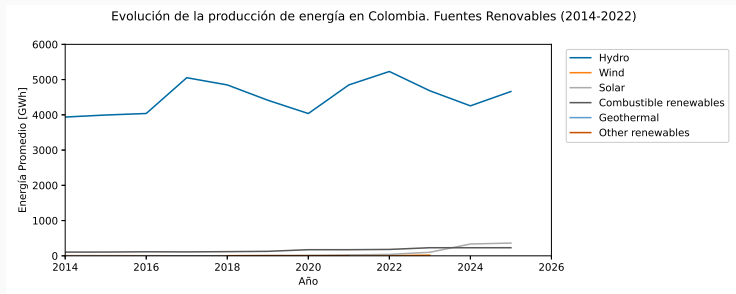


Figura 10: Producción de energía de fuentes Renovables

# Evolución energía de fuentes renovables (2014-2025)



**Figura 11:** Evolución de la producción de energía en Colombia. Fuentes Renovables (2014-2025)

# Producción de energía de fuentes Renovables No hidraulica

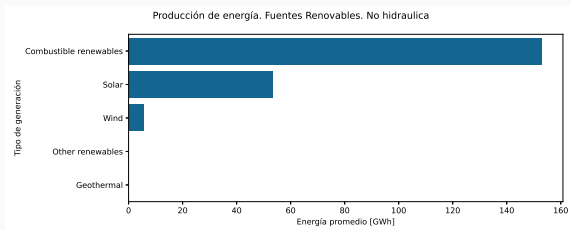
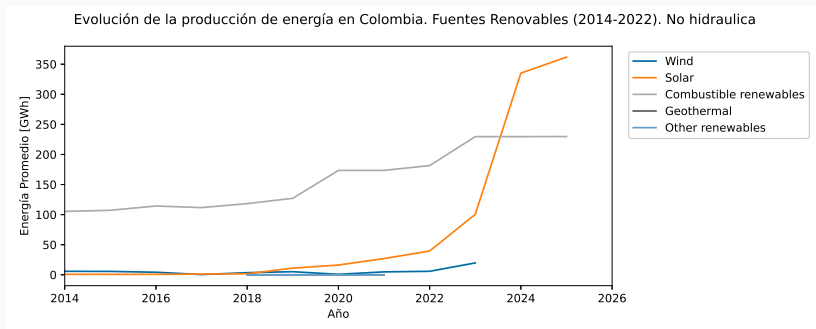


Figura 12: Producción de energía. Fuentes Renovables. No hidraulica

# Evolución energía de fuentes Renovables No hidraulica



**Figura 13:** Producción de energía. Fuentes Renovables. No hidraulica

# Producción de energía de fuentes no renovables

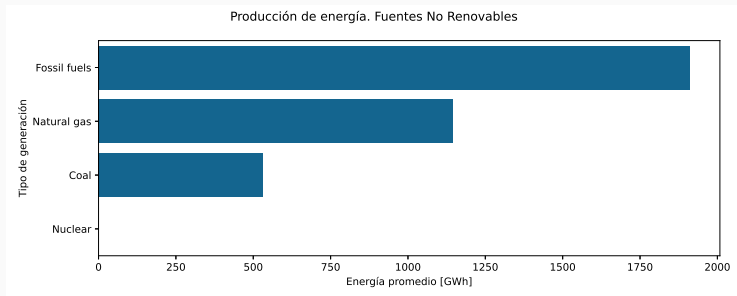
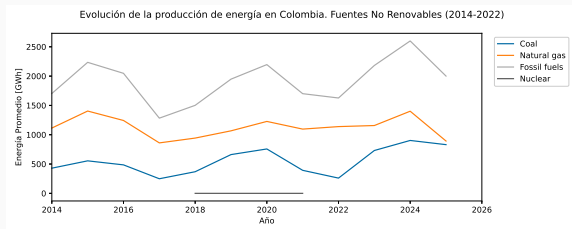


Figura 14: Producción de energía de fuentes no renovables

# Evolución energía en Colombia de fuentes no renovables (2014-2025)



**Figura 15:** Evolución de la producción de energía en Colombia de fuentes no renovables (2014-2025)



# Regresión lineal de la producción eléctrica en Colombia

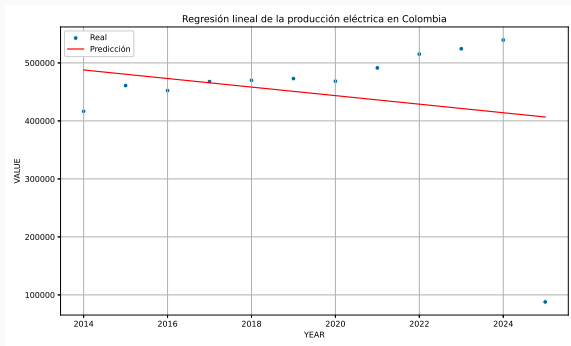
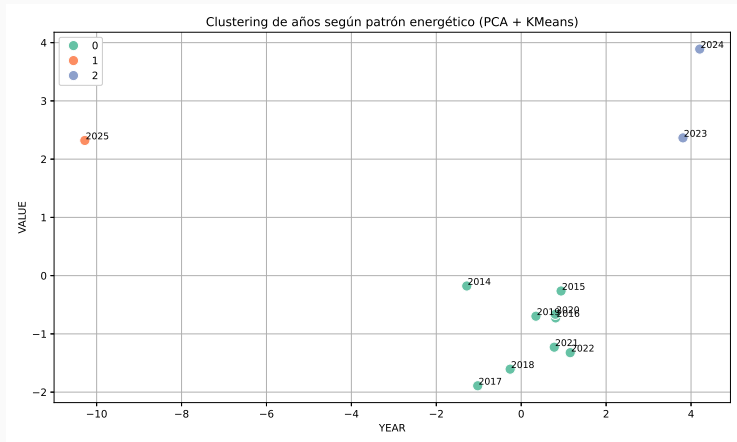


Figura 16: Regresión lineal de la producción eléctrica en Colombia

# Clustering de años según patrón energético (PCA + KMeans)



**Figura 17:** Clustering de años según patrón energético (PCA + KMeans)

# Conclusiones

---

- La generación hidroeléctrica domina ampliamente entre las fuentes renovables.
- Las fuentes como solar y eólica aún representan un porcentaje muy bajo, aunque con potencial de crecimiento.
- La energía basada en fósiles también tiene un peso relevante, especialmente el gas natural.
- Se observa un crecimiento sostenido de la producción entre 2014 y 2018.
- El análisis completo de 2019 a 2025 permitirá ver el impacto de políticas o proyectos recientes.

¿Preguntas?



García, J. and Molina, J. M. (2018).

*Ciencia de datos: técnicas analíticas y aprendizaje estadístico.*