

Análisis Descriptivo de la Contribución Energética, Económica y Ambiental de los Proyectos FNCER en la Matriz Energética Colombiana

Universidad de Caldas – Utraining – MinTIC

Investigadores Principales:

*Karen Alejandra Jaramillo López
Juan Esteban Albornoz Gil
Óscar Andrés García Jurado
Juan Pablo Gonzalez Vásquez*

Docentes guía:

*Juan Andrés García Moreno
Ivonne Castaño Osorio*

Bootcamp Analítica de Datos Nivel Explorador

karenalejandraj10@gmail.com
juan.albornozgil@gmail.com
101oscar101andres@gmail.com
gonzalezvasquezj5@gmail.com
juandres8409@gmail.com
ivonnecastao@gmail.com

Febrero 2026

Pereira

Resumen—Este informe surge como refuerzo ante la poca disponibilidad de análisis de datos exploratorios y estadísticos referentes a los proyectos de Fuentes No Convencionales de Energía Renovable (FNCER) en Colombia.

Se analizaron 152 proyectos registrados hasta 2025 en estado de “operación” y “ejecución”, con el objetivo de evaluar su contribución energética, económica y ambiental. El análisis se realiza aplicando tecnologías de vanguardia como Python en Google Colab, utilizando librerías como Pandas, NumPy, Geopandas y Matplotlib.

La metodología aplicada se estructura en las siguientes etapas: Selección de datos, inspección y verificación de datos, preprocesamiento y limpieza, integración y transformación, análisis estadístico y análisis exploratorio de datos.

Los resultados evidencian que existe una relación lineal positiva entre la inversión estimada y la capacidad instalada en megavatio, por otra parte, se observa que la relación costo - beneficio energético es más estable en aquellos proyectos basados en geotermia en comparación a tecnologías de energía solar, eólica o biomasa. No obstante, la tecnología solar presenta la

mayor reducción total de emisiones de dióxido de carbono, seguida por la eólica. Aunque la biomasa y la geotermia tienen una participación considerablemente menor en reducción de emisiones, es importante señalar que los proyectos solares tienen un mayor número de implementaciones a nivel nacional

Palabras clave— Análisis exploratorio de datos, energía renovable, matriz energética, visualización de datos, impacto ambiental, inversión energética, capacidad instalada, emisiones de dióxido de carbono, análisis estadístico, sostenibilidad energética, geotermia, energía solar, energía eólica

Abstract—This report addresses the limited availability of exploratory and statistical data analysis regarding Non-Conventional Renewable Energy (NCRE) projects in Colombia.

We analyzed 152 projects registered up to 2025 in the “operation” and “execution” phases, with the objective of evaluating their energy, economic, and environmental contributions. The analysis was conducted using cutting-edge technologies such as Python in Google Colab, utilizing libraries like Pandas, NumPy, Geopandas and Matplotlib.

The methodology applied is structured in the following stages: data selection, data inspection and verification, preprocessing and cleaning, integration and transformation, statistical analysis, and exploratory data analysis.

The results show a positive linear relationship between the estimated investment and the installed capacity in megawatts. Furthermore, the energy cost-benefit ratio is more stable in geothermal projects compared to solar, wind, or biomass energy technologies. However, solar technology offers the greatest overall reduction in carbon dioxide emissions, followed by wind power. Although biomass and geothermal energy contribute considerably less to emissions reductions, it is important to note that solar projects have a greater number of implementations nationwide.

Keywords— *Exploratory data analysis, renewable energy, energy matrix, data visualization, environmental impact, energy investment, installed capacity, carbon dioxide emissions, statistical analysis, energy sustainability, geothermal, solar energy, wind energy*

I. INTRODUCCIÓN

La base de este trabajo reside en el uso de datos públicos oficiales sobre proyectos de Fuentes No Convencionales de Energía Renovable (FNCER) en Colombia. Dichos registros recopilan información relevante sobre capacidad instalada, inversión, generación de empleo y reducción de emisiones, constituyendo una oportunidad idónea para aplicar de manera integral las etapas fundamentales del análisis de datos, desde la selección y depuración de la base hasta su análisis e interpretación.

A través de este informe, se busca fortalecer la comprensión de los principios del análisis exploratorio de datos y desarrollar habilidades en el manejo de herramientas ampliamente utilizadas en entornos profesionales. Se pretende fomentar la capacidad de comunicar hallazgos mediante representaciones visuales claras, precisas y técnicamente fundamentadas.

Aunque el ámbito temático se acota al sector energético, el propósito central del proyecto trasciende el contenido específico y se enfoca en el desarrollo de competencias metodológicas en análisis de datos. En este sentido, el estudio se concibe como un ejercicio aplicado que integra técnicas estadísticas, procesos de limpieza y estrategias de visualización.

En última instancia, el proyecto pretende evidenciar cómo el tratamiento riguroso de datos abiertos puede contribuir tanto a la generación de conocimiento sectorial como al fortalecimiento de competencias analíticas transferibles a diversos contextos profesionales.

El informe se estructura de la siguiente manera: inicialmente se presentan los objetivos y la descripción del problema; posteriormente, se expone la metodología empleada, detallando las etapas de selección, limpieza y análisis de los datos; finalmente, se presentan los resultados obtenidos y las conclusiones derivadas del estudio. Tenga en cuenta que los resultados y las conclusiones se abordan en el mismo orden en el cual se describen en los objetivos.

A. Objetivo General

Aplicar un proceso completo de análisis de datos exploratorio utilizando Python y sus librerías especializadas (NumPy, Pandas, Matplotlib) en Google Colab, ejecutando las etapas de carga, limpieza, procesamiento y visualización de conjuntos de datos abiertos enfocado en energías renovables, con el fin de adquirir y consolidar habilidades prácticas fundamentales en ciencia de datos dentro del contexto del curso.

B. Objetivos Específicos:

1. Seleccionar, preparar, analizar, explorar un conjunto de datos de públicos de proyectos de FNCER en Colombia, mediante técnicas de limpieza y transformación con Python mediante uso de Pandas, Geopandas y NumPy.
2. Identificar la distribución geográfica de los proyectos FNCER, determinando los departamentos y municipios con mayor concentración de iniciativas de energías renovables y cuáles presentan la mayor inversión estimada.
3. Analizar la relación económica entre capacidad instalada e inversión estimada por tipo de tecnología (solar, eólico, biomasa, geotérmica) para evaluar patrones de costo-eficiencia.
4. Evaluar la relación entre capacidad instalada y empleos estimados generados, identificando qué tecnologías tienen mayor impacto en la generación de empleo por megavatio.
5. Identificar los proyectos con mayor reducción estimada de emisiones de CO₂ y analizar su distribución según tipo de energía y ubicación geográfica.
6. Realizar visualizaciones de los hallazgos y conclusiones obtenidas en base al análisis descriptivo de los datos con matplotlib

II. DESCRIPCIÓN DEL PROBLEMA

En los últimos años, Colombia ha asumido compromisos nacionales e internacionales orientados a la transición hacia una matriz energética más limpia y sostenible, promoviendo el desarrollo de Fuentes No Convencionales de Energía Renovable (FNCER) como la energía solar, eólica, biomasa y geotérmica. A partir de la Ley 1715 de 2014 [1] y políticas posteriores, se han impulsado múltiples proyectos energéticos distribuidos en diferentes regiones del país, los cuales buscan reducir las emisiones de CO₂ y fomentar el desarrollo económico regional mediante inversión y generación de empleo.

Como resultado de estas políticas, el país cuenta actualmente con registros públicos oficiales que consolidan información detallada sobre proyectos FNCER en estado de “operación” y “ejecución”, incluyendo variables como capacidad instalada, inversión estimada, empleos generados, tipo de tecnología, ubicación geográfica y reducción estimada de emisiones. Sin embargo, a pesar de la disponibilidad de estos datos abiertos, no se dispone de un análisis sistemático que permita comprender de manera clara cómo estos proyectos están contribuyendo realmente

a la matriz energética, económica y ambiental del país.

El problema central que motiva este proyecto es la poca disponibilidad de análisis exploratorios estructurados que permita identificar patrones, relaciones y comportamientos relevantes dentro de los datos de los proyectos FNCER. Aunque la información existe, esta no ha sido poco examinada desde una perspectiva analítica de datos que facilite su interpretación.

La falta de disponibilidad de este tipo de análisis limita la posibilidad de comprender el impacto real de las políticas energéticas implementadas y dificulta la generación de conocimiento que pueda servir de base para futuras decisiones estratégicas en materia de transición energética. Además, desde una perspectiva académica y formativa, estos datos representan una oportunidad valiosa para aplicar técnicas de análisis exploratorio de datos (EDA) que permitan desarrollar competencias prácticas en el manejo, limpieza, transformación y visualización de información utilizando herramientas profesionales como Python con Pandas, NumPy y Matplotlib.

Este trabajo se enfoca exclusivamente en la fase de Análisis Exploratorio de Datos (EDA). Se realizan procesos de carga, limpieza, validación, transformación y visualización de la información utilizando Python en Google Colab. No se incluirán modelos predictivos, pruebas estadísticas avanzadas ni algoritmos de aprendizaje automático, ya que el propósito en esta etapa es comprender la estructura del conjunto de datos, identificar relaciones iniciales entre variables y generar representaciones gráficas que faciliten la interpretación de la información.

Los resultados obtenidos pretenden demostrar el manejo adecuado de conceptos y herramientas de analítica de datos, además, servir como base para futuras etapas del proyecto donde se apliquen técnicas más avanzadas de analítica y modelado predictivo orientadas a la optimización y planeación energética en el contexto colombiano.

III. METODOLOGÍA

A. Selección de datos

El objetivo de esta etapa es seleccionar el conjunto de datos de interés para realizar el análisis exploratorio. Plantear preguntas orientativas según la capacidad de los datos y a partir de esto se debe identificar las variables relevantes involucradas en cada una de las preguntas.

Para finalizar esta etapa, se cargan los datos en el entorno Google Colab a partir de un archivo de valores separados por comas (CSV) obtenido del portal web de Datos abiertos Colombia denominado FNCER [2] para su posterior procesamiento y análisis.

B. Inspección y verificación inicial de los datos

En esta etapa se realiza una revisión inicial del conjunto de datos con el fin de comprender su estructura. Se analizan las primeras filas del dataset, el número de registros, los tipos de variables, el significado de los nombres de las variables según el diccionario de datos y se identifican la presencia de valores faltantes e inconsistencias que pudieran afectar el desarrollo del

análisis.

C. Preprocesamiento y limpieza

Para garantizar la calidad de la información obtenida a partir del estudio de los datos se debe realizar un preprocesamiento en el que se evalúe el impacto de los valores faltantes e inconsistentes encontrados al momento de responder las preguntas planteadas.

Si bien, el dataset original cuenta con una variedad amplia de estados en los que se encuentran los proyectos, dados los objetivos del análisis descriptivo, se debe acotar únicamente a los proyectos en estado de EN EJECUCIÓN y EN OPERACIÓN. Además, para facilitar el manejo se renombran las variables.

Se identifica que 55 de los registros poseen valor de NO APLICA en las variables DEPARTAMENTO y MUNICIPIO, por lo que es necesario realizar una búsqueda externa según la restante información, esto incluye búsquedas en la web y el uso de un segundo dataset, extraído de Datos abierto Colombia [3]. Se agregan las columnas DEPARTAMENTO_ACTUALIZADO y MUNICIPIO_ACTUALIZADO

Según un análisis del tipo de las variables, se determina que es necesario un casting del tipo de variable FECHA de int64 a datetime64, para mantener la integridad de los datos y facilitar su lectura y manejo. Al realizar dicho cambio, se encontró que 11 registros poseen valores irregulares de tipo NaT, al revisar en profundidad la información de estos registros se encontró que sus valores eran 0 en todos aquellos campos referentes al rendimiento, capacidad e inversión, se decide eliminar dichos datos

Se genera una nueva columna AÑO, con el cual facilitar la generación de gráficos y ordenar los datos en base estos valores de manera ascendente, esto incluye además el reinicio de los índices.

Al finalizar la fase de preprocesamiento y limpieza, se pasa de un conjunto de datos de 286 registros a uno de 152 registros, el cual se guarda como un csv distinto, con el que se procede a realizar las siguientes etapas

D. Integración y transformación

En esta fase los datos fueron organizados y transformados de acuerdo con las necesidades de cada análisis. Se realizaron procesos de ordenamiento, filtrado y preparación de las variables, con el fin de facilitar la generación de gráficos y tablas que permitieran responder las distintas preguntas planteadas en el proyecto.

Se reorganizan las categorías de la variable TIPO_ENERGIA, siendo el nuevo orden: EÓLICO, SOLAR, BIOMASA y GEOTERMIA, además de una codificación numérica de las variables categóricas: DEPARTAMENTO_ACTUALIZADO y MUNICIPIO_ACTUALIZADO, esto se efectúa para ordenar y favorecer el procesamiento de los cálculos posteriores del análisis estadístico y descriptivo.

E. Análisis estadístico

Para facilitar y apoyar el estudio exploratorio del conjunto de datos, es necesario realizar un análisis estadístico mediante las

funciones y métodos correspondientes. Se obtuvieron valores para las categorías: Medidas de tendencia central, Medidas de dispersión, medidas de forma, medidas de forma, medidas de posición. Tenga en cuenta que algunas variables no soportan los cálculos al no ser numéricas

TABLA I. MEDIDAS DE TENDENCIA CENTRAL: MEDIA ARITMÉTICA

CAPACIDAD_INSTALADA	2.740000e+01
CAPACIDAD_PICO	3.108000e+01
INVERSIÓN_ESTIMADA	1.047914e+11
EMPLEOS_ESTIMADOS	1.604700e+02
EMISIONES_CO2	4.861557e+04
AÑO	2.021930e+03

TABLA II. MEDIDAS DE TENDENCIA CENTRAL: MEDIANA

CAPACIDAD_INSTALADA	9.900000e+00
CAPACIDAD_PICO	1.139000e+01
INVERSIÓN_ESTIMADA	3.712500e+10
EMPLEOS_ESTIMADOS	6.900000e+01
EMISIONES_CO2	1.387584e+04
AÑO	2.023000e+03

TABLA III. MEDIDAS DE TENDENCIA CENTRAL: MODA

ESTADO	EN EJECUCIÓN
PROYECTO	ACACIAS 2 (CELSIA)
DEPARTAMENTO	NO APLICA
MUNICIPIO	NO APLICA
FECHA	2023-12-31
TIPO_ENERGÍA	SOLAR
CAPACIDAD_INSTALADA	9.9
CAPACIDAD_PICO	11.39
INVERSIÓN_ESTIMADA	3712500000.0
EMPLEOS_ESTIMADOS	69.0
EMISIONES_CO2	13875.84
DEPARTAMENTO_ACTUALIZADO	ATLÁNTICO
MUNICIPIO_ACTUALIZADO	PEREIRA
AÑO	2023.0

TABLA IV. MEDIDAS DE DISPERSIÓN: DESVIACIÓN ESTÁNDAR

CAPACIDAD_INSTALADA	4.649000e+01
CAPACIDAD_PICO	5.173000e+01
INVERSIÓN_ESTIMADA	1.751944e+11
EMPLEOS_ESTIMADOS	2.517500e+02
EMISIONES_CO2	1.043273e+05
AÑO	1.990000e+00

TABLA V. MEDIDAS DE DISPERSIÓN: VARIANZA

CAPACIDAD_INSTALADA	2.161580e+03
CAPACIDAD_PICO	2.676000e+03
INVERSIÓN_ESTIMADA	3.069310e+22
EMPLEOS_ESTIMADOS	6.338020e+04
EMISIONES_CO2	1.088419e+10
AÑO	3.960000e+00

TABLA VI. MEDIDAS DE FORMA: COEFICIENTE DE ASIMETRÍA

CAPACIDAD_INSTALADA	2.559
CAPACIDAD_PICO	2.428
INVERSIÓN_ESTIMADA	2.517
EMPLEOS_ESTIMADOS	2.412
EMISIONES_CO2	3.938
AÑO	-5.107

TABLA VII. MEDIDAS DE FORMA: CURTOSIS

CAPACIDAD_INSTALADA	6.886
CAPACIDAD_PICO	5.946
INVERSIÓN_ESTIMADA	6.632
EMPLEOS_ESTIMADOS	6.778
EMISIONES_CO2	17.798
AÑO	43.129

TABLA VIII. MEDIDAS DE POSICIÓN: PERCENTILES

CAPACIDAD_INSTALADA	1.680000e+00
CAPACIDAD_PICO	1.782000e+00
INVERSIÓN_ESTIMADA	6.675000e+09
EMPLEOS_ESTIMADOS	1.150000e+01
EMISIONES_CO2	2.354685e+03
AÑO	2.021000e+03

F. Análisis exploratorio de datos (EDA)

Finalmente, se realizó un análisis exploratorio de los datos mediante estadísticas descriptivas y visualizaciones elaboradas con la librería Matplotlib.

Para conocer la mayor concentración geográfica de los proyectos FNCR, en Colombia es necesario remitirse al uso de funciones estadísticas como la moda en dos escenarios distintos, el departamento y subsecuentemente los municipios. Véase la figura 1.

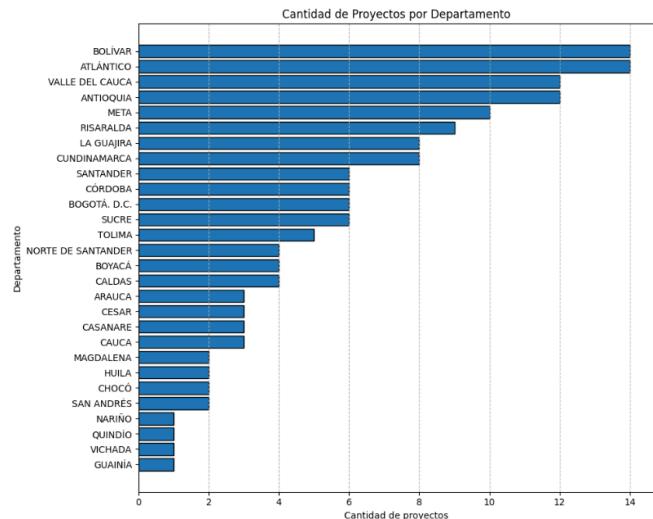


Figura 1. Cantidad de proyectos por departamento

Como se observa en la anterior figura, para los departamentos se obtiene un resultado bimodal con ATLÁNTICO y BOLÍVAR, por lo tanto, al momento de localizar la concentración en los municipios es necesario generar un gráfico por cada uno para favorecer su estudio individual. Véase figura 2 y 3.

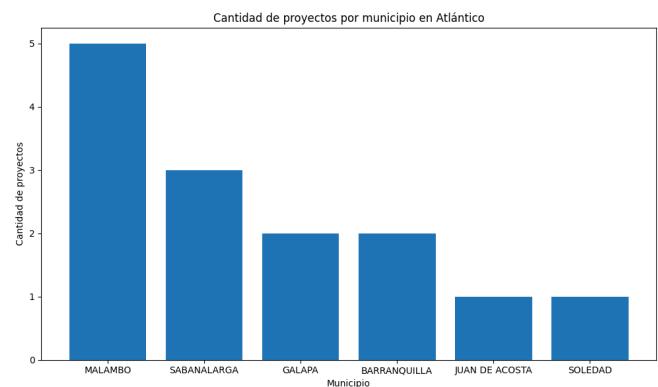


Figura 2. Cantidad de proyectos por municipio - ATLÁNTICO

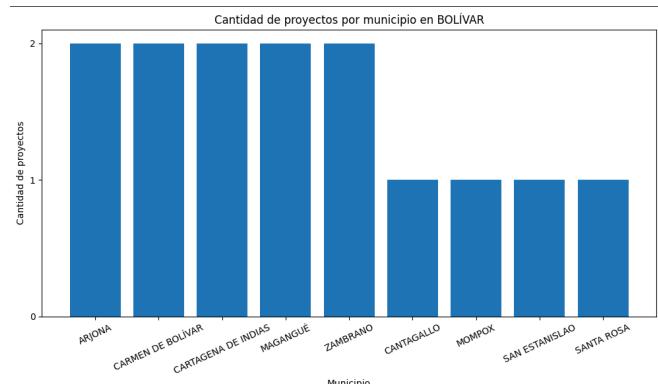


Figura 3. Cantidad de proyectos por municipio - BOLÍVAR

Adicionalmente, es preciso saber cuáles departamentos poseen una

mayor inversión estimada, por lo tanto, se realiza la suma total de inversión por departamento.

En este caso, teniendo en cuenta la naturaleza de los datos, para facilitar y mejorar la visualización de los mismos, se hace una copia del dataset original, para posteriormente modificar la escala de los valores de la variable INVERSION_ESTIMADA. La escala tomada fue de billón. Para mostrar los 16 departamentos se utiliza un gráfico lollipop. Véase la figura 4.

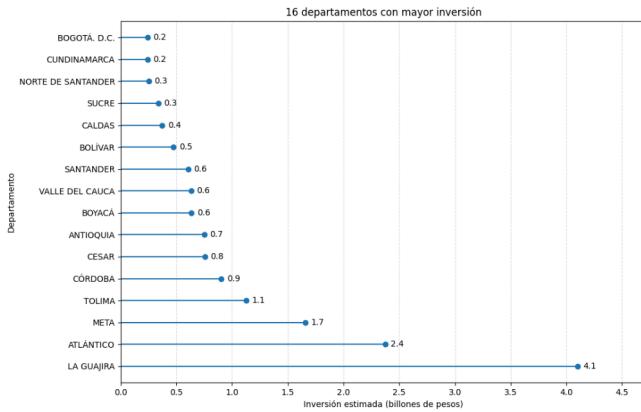


Figura 4. Inversión estimada por departamento

Debido que el departamento de LA GUAJIRA no presenta una cantidad de proyectos tan elevada como ATLÁNTICO y BOLÍVAR, para fortalecer los resultados de la metodología aplicada se realiza gráfico de correlación entre las variables de cantidad de proyecto e inversión estimada. véase la figura 5

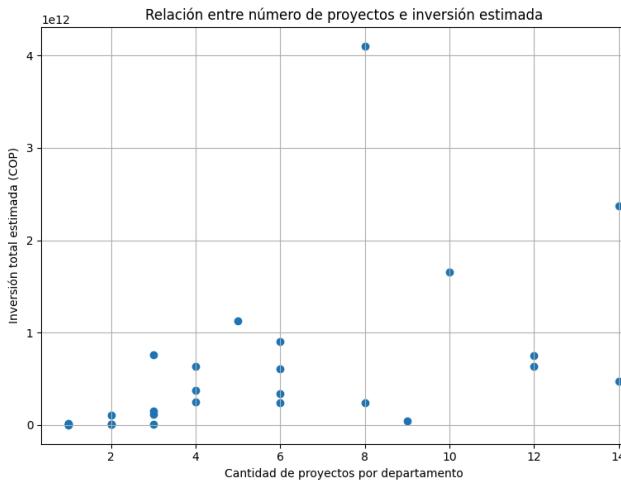


Figura 5. Capacidad vs inversión

Se obtiene un coeficiente de correlación de 0.5, siendo este un resultado moderado pero no determinante, indicando que la cantidad de proyectos por departamento no está directamente relacionada con la inversión por departamento, ya que allí pueden intervenir más factores como tamaño de proyecto y su capacidad instalada.

Para determinar la proporción existe entre la energía generada (capacidad) y la inversión realizada en los proyectos FNCER es

necesario establecer una correlación entre las variables mencionadas, a través de un gráfico de dispersión. Véase Figura 6

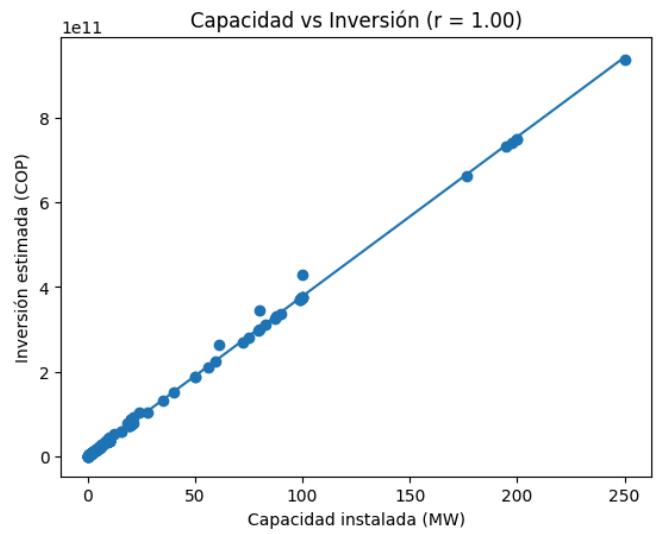


Figura 6. Capacidad vs inversión

En extensión, es necesario determinar la viabilidad económica de los tipos proyectos de proyecto, lo cual se ilustra mediante un gráfico de barras, Véase la figura 7

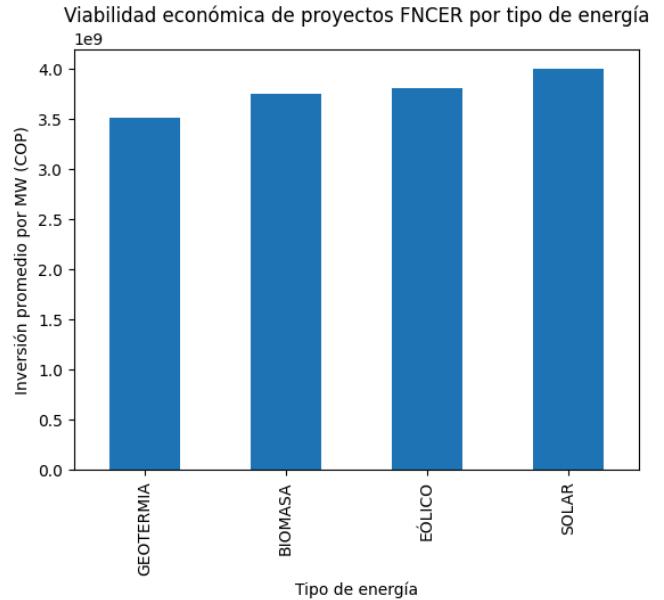


Figura 7. Viabilidad económica por tipo de proyecto

Adicional a lo anterior, se considera un estudio por medio de cajas para ilustrar la distribución de la inversión por megavatio según el tipo de proyecto. Véase Figura 8

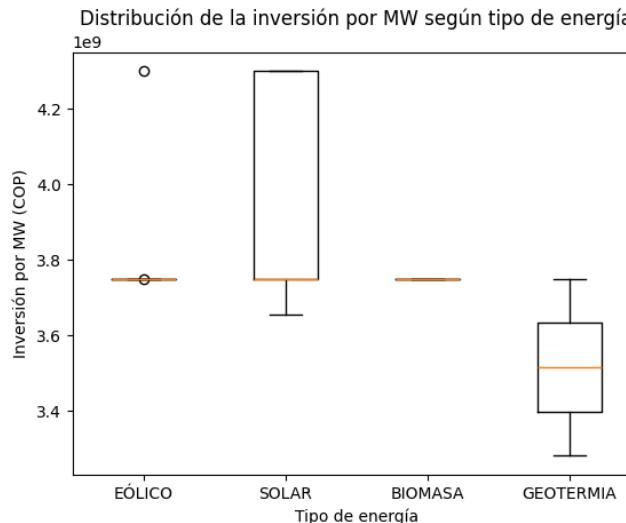


Figura 8. Distribución de la inversión según el tipo de energía

Para analizar la relación entre la capacidad instalada y los empleos estimados en los proyectos FNCER, inicialmente se seleccionaron del conjunto de datos las variables relevantes: CAPACIDAD_INSTALADA y EMPLEOS_ESTIMADOS. Se organiza la información de manera ascendente según la capacidad instalada, con el fin de visualizar de forma clara cómo varía el número de empleos a medida que aumenta el tamaño del proyecto.

Por último, se representa gráficamente la relación mediante un gráfico lineal, donde el eje de abscisas corresponde a la capacidad instalada (MW) y el eje de ordenadas a los empleos estimados. Esta visualización facilita la identificación de patrones y permite evaluar si existe una relación proporcional entre el crecimiento energético y la generación de empleo. Véase la figura 9

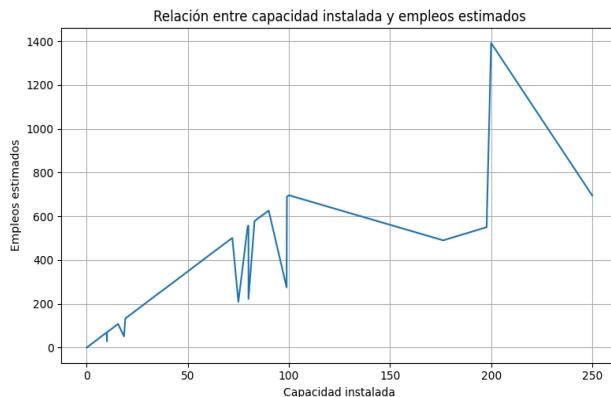


Figura 9

Para determinar qué proyectos FNCER presentan la mayor reducción estimada de emisiones de CO₂, inicialmente se seleccionaron del conjunto de datos únicamente las variables relevantes: PROYECTO, TIPO_ENERGIA y EMSIONES_CO2. Posteriormente, los proyectos se ordenaron de forma descendente según la reducción de emisiones, lo que permitió identificar aquellos con mayor impacto ambiental.

Debido a la gran cantidad de proyectos y a la longitud de sus nombres, se decide representar gráficamente sólo los diez proyectos con mayor reducción estimada de CO₂. Esta decisión se toma con el fin de mejorar la legibilidad del gráfico y facilitar la interpretación de los resultados, destacando los proyectos más representativos. Véase la figura 10.

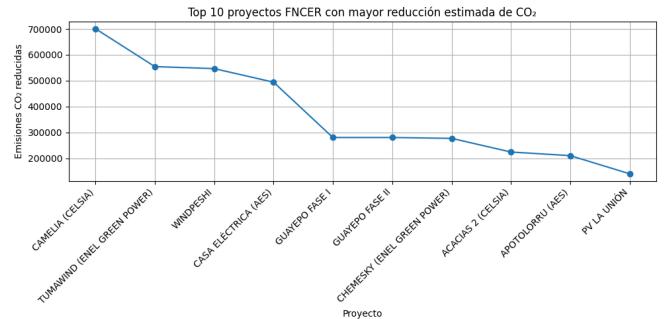


Figura 10. Proyectos con mayor reducción estimada de CO₂

Finalmente, para analizar cómo se distribuye la reducción de emisiones por tipo de energía, los datos se agrupan según el tipo de fuente energética y se sumaron las emisiones de CO₂ reducidas en cada categoría. Esta información se representa mediante un segundo gráfico de líneas, lo que permite comparar el aporte total de cada tipo de energía y concluir cuáles tecnologías contribuyen en mayor medida a la mitigación de emisiones. Véase la figura 11

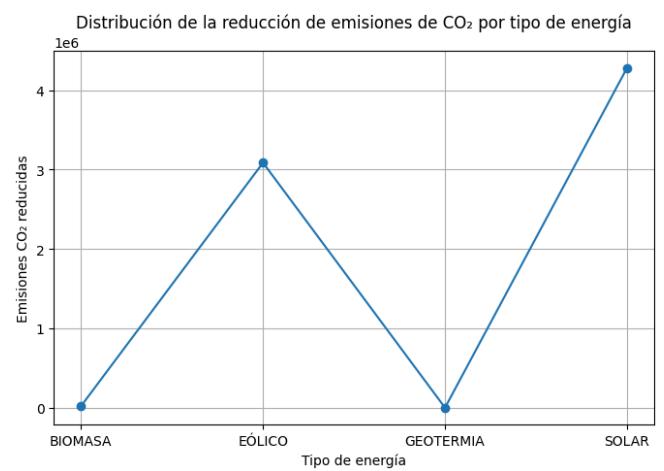


Figura 11. Distribución de emisiones de CO₂ por tipo de energía

IV. RESULTADOS

1. El análisis de correlación entre la capacidad instalada y la inversión estimada muestra una relación lineal positiva prácticamente perfecta ($r \approx 1$), lo cual indica que, dentro del conjunto de datos, la inversión crece de forma proporcional a la capacidad de los proyectos.

Este comportamiento sugiere que los costos de inversión por megavatio tienden a ser homogéneos entre los proyectos analizados, lo que es consistente con un esquema de estimación de inversión basado en capacidad instalada.

2. Al analizar la inversión promedio por megavatio según el tipo de energía, se observan diferencias en la viabilidad económica relativa de las tecnologías FNCER. La geotermia presenta el menor costo promedio por MW, seguida de la biomasa y la energía eólica, mientras que, la energía solar muestra el mayor costo relativo.

Estos resultados permiten identificar patrones de costo–eficiencia desde un enfoque descriptivo, sin asumir rentabilidad financiera directa.

3. El análisis mediante diagramas de caja evidencia diferencias significativas en la dispersión de la inversión por megavatio entre los distintos tipos de energía. La energía solar presenta la mayor variabilidad en los costos por MW, lo que sugiere heterogeneidad en el tamaño y características de los proyectos. En contraste, la energía eólica y la geotermia muestran distribuciones más concentradas, indicando mayor estabilidad en los costos relativos. La biomasa presenta una distribución limitada, asociada al reducido número de proyectos registrados.
 4. Se identificaron proyectos atípicos con valores de inversión por megavatio superiores al comportamiento general, como el proyecto eólico JEPIRACHI, cuyo alto costo relativo puede asociarse a su menor capacidad instalada y a condiciones específicas de implementación. La presencia de estos outliers refuerza la importancia de analizar la viabilidad económica considerando tanto el tipo de tecnología como la escala del proyecto.
 5. Existe una relación positiva entre la capacidad instalada y los empleos estimados generados por los proyectos FNCER. A medida que aumenta el tamaño del proyecto en megavatios, también tiende a incrementarse la cantidad de empleo asociado, lo que evidencia que el crecimiento de la infraestructura de energías renovables no solo impacta la producción energética, sino también el desarrollo económico y laboral. Sin embargo, la relación no es completamente proporcional, ya que se observan variaciones entre proyectos con capacidades similares que generan diferentes niveles de empleo.
- Esto sugiere que factores como el tipo de tecnología utilizada o la etapa del proyecto pueden influir en la cantidad de puestos de trabajo creados. En conjunto, los resultados permiten afirmar que los proyectos de mayor escala tienen un impacto más significativo en la generación de empleo, reforzando la importancia de las energías renovables como motor tanto energético como social en el país.
6. La mayor reducción estimada de CO₂ se concentra en un grupo reducido de proyectos de gran escala, destacándose especialmente uno como el de mayor impacto ambiental.

7. Al observar la distribución por tipo de energía, la tecnología solar presenta la mayor reducción total de emisiones, seguida por la eólica, mientras que la biomasa y la geotermia tienen una participación considerablemente menor.

V. CONCLUSIONES

1. Se realiza de manera satisfactoria la selección, preparación, análisis, exploración y visualización de un conjunto de datos públicos de FNCER en Colombia, mediante técnicas de limpieza y transformación con Python mediante uso de Pandas, NumPy y Matplotlib
2. El análisis de la distribución geográfica evidencia que los proyectos FNCER no se encuentran homogéneamente dispersos en el territorio nacional, sino que presentan concentraciones claras en determinados departamentos y municipios. Esta localización podría responder tanto a condiciones naturales favorables como a dinámicas de inversión y desarrollo regional. Se observa que las mayores inversiones estimadas tienden a asociarse con zonas donde se desarrollan proyectos de mayor escala, lo que sugiere una relación directa entre localización estratégica y magnitud de las iniciativas energéticas.
3. La relación lineal positiva prácticamente perfecta entre la capacidad instalada y la inversión estimada indica que, dentro del conjunto analizado, la inversión se estructura principalmente en función del tamaño del proyecto. Esto sugiere la existencia de criterios de estimación relativamente homogéneos en términos de costo por megavatio. Sin embargo, al desagregar por tecnología, emergen diferencias relevantes en los costos promedio por MW, lo que permite identificar variaciones en el costo – eficiencia relativa de las distintas fuentes energéticas. En este contexto, la geotermia muestra mayor estabilidad en costos, mientras que la energía solar presenta una mayor dispersión, reflejando heterogeneidad en los proyectos.
4. Los resultados confirman que existe una relación positiva entre la capacidad instalada y los empleos estimados generados, lo que permite afirmar que el crecimiento de la infraestructura FNCER contribuye no solo al sistema energético, sino también al desarrollo económico. No obstante, la ausencia de una proporcionalidad estricta revela que la generación de empleo depende de factores adicionales, como el tipo de tecnología, la complejidad técnica del proyecto o su fase de ejecución. En términos comparativos, ciertas tecnologías muestran mayor intensidad laboral relativa, lo que sugiere que el impacto social de los proyectos no está determinado únicamente por su escala energética.
5. La reducción estimada de emisiones se concentra principalmente en proyectos de gran escala, lo que indica que el impacto ambiental positivo está fuertemente influenciado por la magnitud de la capacidad instalada. A nivel tecnológico, la energía solar lidera la reducción total de emisiones, seguida por la eólica, resultado que se encuentra estrechamente vinculado con la mayor

cantidad de proyectos implementados bajo estas tecnologías. Aunque otras fuentes como la biomasa y la geotermia presentan una contribución menor en términos agregados, su participación sigue siendo relevante dentro de la diversificación de la matriz energética.

6. El uso de representaciones visuales permitió identificar patrones, relaciones y comportamientos atípicos que no resultan inmediatamente evidentes a partir de tablas o métricas aisladas. Las visualizaciones facilitaron la interpretación de correlaciones, dispersión de datos, diferencias entre tecnologías y detección de valores atípicos, consolidándose como una herramienta clave dentro del análisis exploratorio. En este sentido, el estudio evidencia que la visualización de datos no solo cumple una función ilustrativa, sino también analítica, contribuyendo significativamente a la comprensión integral del fenómeno estudiado.
7. En conjunto, el análisis evidencia que los proyectos FNCER en Colombia presentan patrones estructurales consistentes en términos de inversión, capacidad instalada, impacto laboral y contribución ambiental. La fuerte relación entre inversión y capacidad refleja esquemas de planificación relativamente estandarizados, mientras que las diferencias entre tecnologías revelan dinámicas económicas y operativas particulares. Asimismo, se confirma que las energías renovables no sólo desempeñan un papel estratégico en la transición energética, sino también en la generación de beneficios económicos y ambientales. Finalmente, el estudio demuestra la relevancia del análisis exploratorio y la visualización de datos como herramientas fundamentales para la interpretación rigurosa de datos abiertos en contextos sectoriales complejos.

VI. BIBLIOGRAFÍA

- [1] Congreso de la República de Colombia, «Función pública,» 13 Mayo 2014. [En línea]. Disponible en: <https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=57353>. [Último acceso: 5 Febrero 2026].
- [2] Gobierno de la República de Colombia, «Datos abiertos Colombia,» 19 Diciembre 2025. [En línea]. Disponible en: https://www.datos.gov.co/dataset/FNCER/nyyd-gdfx/about_data. [Último acceso: 7 Febrero 2026].
- [3] Gobierno de la República de Colombia, «Datos abiertos Colombia,» 19 Diciembre 2025. [En línea]. Disponible en: https://www.datos.gov.co/Minas-y-Energ-a/Meta-FNCER-Incorporar-en-la-matriz-energ-tica-nuev/vy9n-w6hc/about_data. [Último acceso: 7 Febrero 2026].