PROYECTO FINAL DE ANÁLISIS DE DATOS NIVEL INTEGRADOR

ANÁLISIS DE DATOS DE LA TEMÁTICA DE CAPTURA DE CARBONO (CC) COMO HERRAMIENTA EMERGENTE PARA MITIGAR LAS EMISIONES DE DIÓXIDO DE CARBONO

ELABORADO POR

JHON FREDY RIOS HOYOS CC 71381452

MANUEL ARANGO FIGUEROA CC 1040753960

PAULA ANDREA PEREZ ESPITIA 43619863

MINISTERIO DE LAS TECNOLOGÍAS DE LA INFORMACIÓN Y LAS COMUNICACIONES – MINTIC

TALENTO TECH

MEDELLÍN

2025

Introducción

Este proyecto final tiene como objetivo evaluar los conocimientos adquiridos en el Nivel 2 del curso de análisis de datos. El equipo de campistas estuvo conformado por tres personas. En este proyecto se abordó una problemática real relacionada con la transición energética específicamente con las diferentes variables relacionadas de las tecnologías de captura de carbono a partir de dióxido de carbono de las emisiones. La fuente de los datos parte de tres datasets, los cuales son públicos y verídicos, dado que parten de instituciones científicas y reconocidas relacionadas con la temática. A partir de este conjunto de datos, se identificaron y a su vez se analizaron, las variables más relevantes del problema mencionado. Para lo cual se aplicaron las diferentes técnicas de análisis de datos vistas durante el curso *Bootcapms*.

ENLACES

Los enlaces del código (junto con el dashboard) se encuentra en el enlace de Github y del dashboard en se presentan a continuación:

Github
https://github.com/ManuAF-291196/Proyecto_TalentoTech/blob/data_analysis/Proyecto_CCS.ipynb

Todos los gráficos son presentados en el anexo a este trabajo escrito, correspondiente a la presentación realizada, la cual se encuentra al final del texto.

Tabla de contenido

Introducción 2

Objetivos del Proyecto 4

- Objetivo general 4
- Objetivos específicos 4

Desarrollo del Proyecto: 5

1. Identificación de la Problemática: 5
2. Recolección de Datos: 5
3. Exploración y Limpieza de Datos: 5
4. Análisis Descriptivo: 6
5. Interpretación de Resultados: 6

Conclusiones 7

Recomendaciones 8

Objetivos del Proyecto

Desarrollar un análisis exhaustivo de datos reales de la problemática identificada en el contexto de la transición energética. El proyecto busca que los campistas demuestren su capacidad para explorar, limpiar, manipular, y visualizar datos, interpretando patrones y tendencias para llegar a conclusiones significativas.

Objetivo general

Analizar los datos obtenidos a partir de datasets de procedencia pública y verídicas de las tecnologías asociadas en la captura de dióxido de carbono como herramienta emergente para mitigar las emisiones, éstas a su vez están asociadas en los diferentes proyectos de investigación, los cuales están en función de los países y de los estados de cada proyecto.

Objetivos específicos

1. Relacionar el estado de proyectos de las diferentes tecnologías de captura de carbono en función de los países donde se están desarrollando este tipo de proyectos.
2. Estimar las diferentes correlaciones de las variables de proyectos de investigación, tales como países, las tecnologías relacionadas con la captura de CO2, la capacidad de captura anunciada y la capacidad de captura estimada, en donde se están desarrollando estos proyectos de investigación.
3. Establecer una relación entre los datos de captura de CO2 y el costo de captura de CO2.

Desarrollo del Proyecto:

1. Identificación de la Problemática:

A pesar del creciente consenso científico sobre la urgencia de mitigar el cambio climático, las emisiones de dióxido de carbono (CO2) continúan siendo una preocupación central. La descarbonización de la economía global requiere una estrategia multifacética que incluye la reducción de la dependencia de combustibles fósiles, el aumento de la eficiencia energética y el desarrollo e implementación de tecnologías para capturar el CO2 emitido. En este contexto, la captura de carbono (CC) emerge como una herramienta potencialmente crucial para mitigar las emisiones de fuentes puntuales (como centrales eléctricas e industrias pesadas) y, en combinación con el almacenamiento geológico o la utilización, para avanzar hacia los objetivos de descarbonización.

A nivel global, diversos países e instituciones están invirtiendo en proyectos de investigación y desarrollo (I+D) para explorar y optimizar diferentes tecnologías de captura de CO2. Estas tecnologías varían en su madurez tecnológica, eficiencia energética, costos de implementación y aplicabilidad a diferentes fuentes de emisión. La diversidad de enfoques y la naturaleza incipiente de algunas de estas tecnologías plantean interrogantes importantes sobre su potencial real para contribuir a la descarbonización a gran escala.

2. Recolección de Datos

Para mayor detalle véase las páginas del Anexo: 8, 16 y 56. Los datos provienen de tres datasets, cuyas referencias son:

- Data set 1: https://netl.doe.gov/carbon-management/carbon-storage/worldwide-ccs-database
- Data set 2: https://www.iea.org/data-and-statistics/data-product/ccus-projects-database
- Data set 3: https://www.sciencedirect.com/science/article/pii/S0959652624017281?via%3Dihub#appsec1

Para el caso del Dataset 1, se extrajeron los datos del Laboratorio Nacional de las Tecnologías Energéticas (NETL por sus siglas en inglés), el cual está adscrito al Departamento de Energía de los Estados Unidos (DOE). Este Dataset está disponible para ser descargado de manera gratuita en formato xlsx en el sitio web del NETL. El diccionario corresponde a la tabla 1.

Tabla 1. Diccionario de los data sets.	
Captura	Captura de CO_2
CCS	De sus siglas en inglés (*Capture Carbon Storage)* Captura y almacenamiento de CO_2
Almacenamiento	Almacenamiento de CO_2 solamente.
T & S	De sus siglas en inglés (*Transport and Storage)* Transporte y almacenamiento de CO_2
Transporte	Transporte posterior del almacenamiento de CO_2
CCU	De sus siglas en inglés *Carbon, Capture, Utilisation*
Cadena completa	Captura, Almacenamiento, Transporte y utilización

El grueso de este trabajo corresponde al data set 2, del cual se extrajeron los datos de la agencia internacional de energía (IEA por sus siglas en inglés) acerca de la captura, transporte, almacenamiento y utilización de Dióxido de Carbono. Este Dataset está disponible de manera gratuita en el sitio web oficial de la IEA, y sus datos fueron recolectados y procesados por la agencia, y fueron producidos con el objetivo de monitorear el progreso de las tecnologías de captura de CO2. Los registros de este dataset contienen proyectos comisionados desde la década de los 70s con una capacidad anunciada mínima de 100000 toneladas por año (o 1000 toneladas por año para plantas de captura de CO2 directamente del aire) que han demostrado metas claras en la reducción de emisiones de CO2 y excluye aquellos proyectos cuyo uso del CO2 no conlleva a beneficios climáticos demostrables.

El Dataset 3 fue descargado de un artículo científico publicado en *Journal of Cleaner Production*, el cual relaciona las fuentes de emisiones en la industria siderúrgica de China con los posibles sumideros. El archivo xlsx. está disponible para descarga en los anexos del artículo y fue descargado gracias al acceso institucional a las bases de datos de la Universidad de Antioquia de la base de datos ScienceDirect.

3. Exploración y Limpieza de Datos:

Para el data set 1, la ruta de limpieza de datos fue la siguiente:

1. Se eliminaron columnas que no resultaban redundantes dentro del análisis planteado.
2. Se realizaron equivalencias de tipos de datos con los datos del Dataset referencia dataset 2
3. Conversión de tipos de datos para las columnas de análisis tipo de proyecto, País, Estado de proyecto.
4. Traducción de datos, a español, corrección de datos después de traducción.
5. Análisis de pertinencia de datos.

Para el dataset 2, fueron cargados a un Data Frame a través de la función read.xlsx de la librería de Pandas. Se realizó una exploración inicial de las columnas y los tipos de datos dentro del Data Frame usando la función *DataFrame.info()* para luego eliminar las columnas con información inútil o redundante. En este caso, eliminamos las columnas que contienen los sitios web de dónde se extrajo la información, y el nombre del conglomerado y de la región al que hacen parte los proyectos de captura de CO2 ya que todo nuestro análisis se centra en cuál es el comportamiento de los proyectos de captura de CO2 con relación a los países y al tipo de tecnología. Asimismo, se inspeccionaron los valores únicos dentro de las columnas "Country", "Project type", y "Project Status", para determinar si habían duplicados que debido a la sensibilidad de Python a las mayúsculas y minúsculas o al orden de las palabras se están tomando cómo registros distintos. Particularmente, se unificaron los valores "Full chain" y "Full Chain" en el campo "Project type", y "Australia-Japan" y "Japan-Australia" en el campo "Country". Adicionalmente, encontramos una entrada con el valor "Japan-Unknown" la cual para efectos prácticos llevamos a "Japan".

Por otro lado se llevaron los campos "Announcement", "FID", "Operation" y "Suspension/decommissioning" de float64 a datetime64 ya que estos datos corresponden a las fechas de anuncio, inversión de capital, inicio de operaciones y cierre del proyecto por lo que es necesario realizar la transformación para realizar series de tiempo.
Igualmente, se realizaron las traducciones de los valores en las columnas "Country", "Project type", "Project Status" y "Sector", usando la librería translate de Google para tener las etiquetas de nuestros datos en español en los gráficos. Después de ejecutar las traducciones, realizamos de nuevo la exploración de valores únicos y la limpieza de las inconsistencias usando la función *DataFrame.replace().*

4. Análisis Descriptivo:

En los tres datasets se utilizaron técnicas de análisis descriptivo para identificar patrones, tendencias, y distribuciones en los datos. También se calcularon las medidas de tendencia central, varianza, percentiles y otros estadísticos relevantes. Para los data sets 2 y 3 se calculó el análisis de componentes principales y la matriz de covarianza. Para el data set, dada la naturaleza de las variables numéricas se realizó la regresión lineal y a su vez, se entrenó un modelo de regresión lineal y polinomial, los cuales se validaron mediante el error del modelo, correspondiente a los valores predichos como los estimados. Los gráficos fueron obtenidos usando la librería Matplotlib.

4. Interpretación de Resultados:

Para el data set 1 se encontraron un total de 385 proyectos alrededor del mundo. Las variables categóricas son:

1. Porcentajes por tipo de proyectos
2. Porcentaje por estado de proyectos
3. Análisis de tipos de proyectos Vs estados de proyectos

4. Proyectos por tipos de estado vs Pais
5. Proyectos por estado vs Pais

A partir del data set 1 se obtuvo principalmente que, la mayoría de proyectos que se están trabajado a nivel mundial, son proyectos de Captura y almacenamiento (42.3%), seguidos de los proyectos de captura (27.8%). La gran mayoría de los proyectos se encuentra en estado activo (véase la página 11 del anexo).

En función del país, Estados unidos es el que tiene la mayor cantidad de proyectos de todo el conjunto de datos alrededor 43% (véase en la página 13 del anexo).

A partir de este análisis exploratorio obtenido del data set 1, es donde se adicionaron los data set 2 y 3, para complementar todo el análisis de datos de la temática abordada.

A partir del Dataset 2 se realizaron gráficos de proporción para analizar cómo se distribuyen los proyectos de captura de CO_2 de acuerdo con los países donde se presentan proyectos, las tecnologías usadas y el estado del proyecto.

Inicialmente, se puede apreciar que el 88% de los proyectos están en estado de planeación y que la tecnología más usada es la Captura de CO_2 (véase página 17 del anexo).

Para las variables numéricas "Capacidad Anunciada" y "Capacidad estimada" se calcularon las medidas de tendencia central y de dispersión con y sin los datos atípicos (páginas 27 y 33 del anexo, respectivamente). Para el primer caso se puede apreciar que, la desviación estándar para ambas variables es cuatro veces mayor que el promedio.

Al realizar los histogramas para ambas variables y los *boxplots* (páginas 28 a 31 del anexo) agrupados por industria y por país, se puede observar que, este comportamiento se debe a los valores atípicos presentados en la capacidad anunciada y estimada en proyectos de transporte de CO_2 en Estados Unidos, los cuales presentan valores de 250 Mt/año. Estos valores estaban afectando el promedio y la desviación estándar, ya que son 83 veces mayores al promedio de la muestra, por lo que se tomó la decisión de eliminarlos para analizar el comportamiento de las variables de análisis. Luego de eliminar estos valores, se puede percibir que, los valores de capacidad anunciada y estimada siguen estando bastante dispersos, ya que la desviación estándar es dos veces mayor que el promedio. Esto se puede deber a que, las emisiones de CO_2 de cada industria y de cada país presentan variaciones debido a diferentes factores tales como, la eficiencia de los procesos, las demandas energéticas y el volumen de producción (veáse páginas 34 a 37 del anexo).

Del data set 3, correspondiente a los datos de las siderúrgicas de China (Alto horno y horno con soplo enriquecido de oxígeno), después del análisis a partir de todos los estadísticos descriptivos, junto con los otros estadísticos relevantes, así como los de la modelación de regresión lineal, se encontrón una alta correlación lineal entre la captura de CO2 y el costo de captura. Dicha relación obtenida corresponde a un comportamiento inversamente proporcional, el dónde a valores bajos de captura el costo es mayor. También a partir del

estimativo de probabilidad de Kernel, el proceso comienza a ser rentable a valores superiores de 13 millones de toneladas de CO_2 capturadas. Cabe resaltar que el valor más alto de captura reportado corresponde a 104,2 millones de toneladas de CO_2 capturadas.

Conclusiones

Desde la pertinencia metodológica importante, a partir de la exploración preliminar del data set 1, de manera juiciosa, el análisis de los datos de la temática fue enriquecido con la inserción de los data set 2 y 3, con el fin de abordar de manera global la problemática planteada.

Desde los resultados obtenidos y del análisis de estos, algunas conclusiones importantes son:

Dada la gran dispersión presentada en la distribución de probabilidad de la variable capacidad de captura en la base de datos provenientes de la IEA, es crucial reconocer y cuantificar esta incertidumbre para tomar decisiones informadas sobre el potencial de la captura de CO_2 como tecnología de mitigación del cambio climático.

Finalmente, a partir de la ventana de observación de las fuentes publicas de los últimos dos años (2023 a 2024), puede concluirse que la alternativa emergente de la captura de CO_2 para mitigar las emisiones, es una posibilidad prometedora, que vale la pena, seguir investigando y mejorando continuamente.

Recomendaciones

Para futuros ejercicios, se puede analizar en qué medida los objetivos y resultados esperados de los proyectos de investigación en curso se alinean con las metas de descarbonización a nivel nacional e internacional (ej. Acuerdos de París).

Evaluar qué implicaciones tienen los diferentes enfoques tecnológicos y los avances en la investigación para la trayectoria futura de la descarbonización global, dado que una de las tecnologías asociada a la temática, correspondiente al almacenamiento, ha sido cuestionada por sus posibles impactos en la afectación de la calidad de aguas subterráneas.

ANEXO

Carbon Capture (CO_2)

4 de abril de 2025

TIC
TALENTO TECH
BOOTCAMPS GRATUITOS PARA TI
en habilidades digitales
Programación
Inteligencia Artificial
Blockchain
Análisis de Datos
Arquitectura en la Nube
Ciberseguridad
Refuerza, profundiza y certifica tus conocimientos en habilidades digitales con el Ministerio TIC para impulsar tu proyección laboral
¡Asegura tu cupo ahora!
Más información: talentotech2.com.co
UTraining
UNIVERSIDAD DE ANTIOQUIA
Universidad de Caldas
ubicua

ANÁLISIS DE DATOS DE LA TEMÁTICA DE CAPTURA DE CARBONO (CC) COMO HERRAMIENTA EMERGENTE PARA MITIGAR LAS EMISIONES DE DIÓXIDO DE CARBONO

- JHON FREDY RIOS HOYOS
- MANUEL ARANGO FIGUEROA
- PAULA ANDREA PEREZ ESPITIA

Fuentes de los Datasets

- Data set 1 : https://netl.doe.gov/carbon-management/carbon-storage/worldwide-ccs-database
- Data set 2: https://www.iea.org/data-and-statistics/data-product/ccus-projects-database
- Data set 3: https://www.sciencedirect.com/science/article/pii/S0959652624017281?via%3Dihub#appsec1

Objetivos específicos

1. **Relacionar el estado de proyectos** de las diferentes tecnologías de captura de carbono en función de los países donde se están desarrollando este tipo de proyectos.
2. **Estimar las diferentes correlaciones de las variables** de proyectos de investigación, tales como países, las tecnologías relacionadas con la captura de CO2, la capacidad de captura anunciada y la capacidad de captura estimada, en donde se están desarrollando estos proyectos de investigación.
3. **Establecer una relación** entre los datos de captura de CO2 y el costo de captura de CO2.

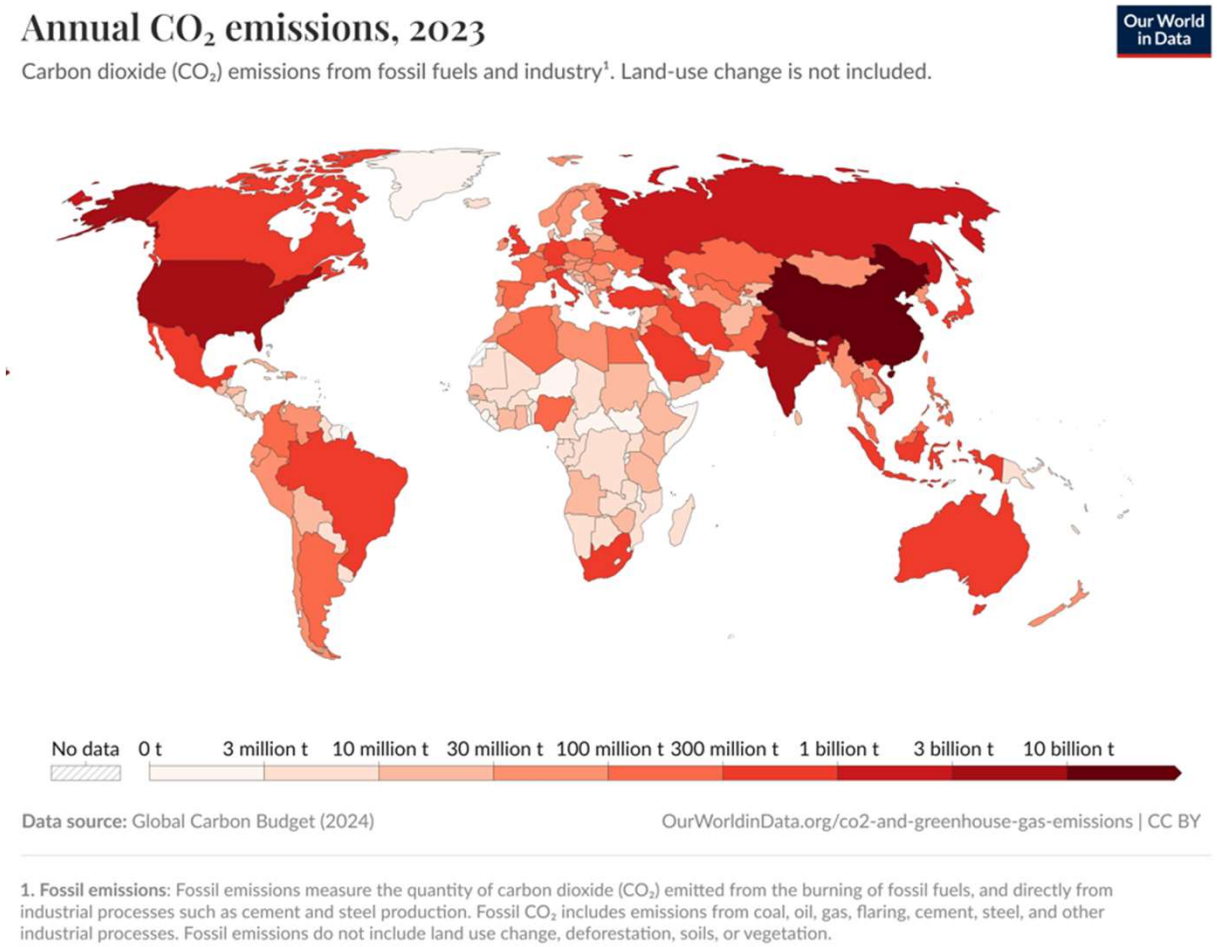
Análisis estadístico de datos de la temática de captura de CO_2 a partir de los tres data sets

	Data set 1	Data set 2	Data set 3
Objetivo 1 Visualización preliminar de datos / Análisis estadístico descriptivo	Diagrama de distribución, promedio	Diagrama de distribución, histograma, cajas y bigotes, serie de tiempos	Diagrama de distribución, histograma, cajas y bigotes,
Objetivo 2 (Misión 2) Conceptos estadísticos avanzados	No aplica	Evaluación de promedios y varianzas (No aplica PCA)	Análisis de componentes Principales (PCA) Modelado Regresión lineal Matriz de Covariabilidad
Objetivo 3 Dashboard (Misión 3)		Proyectos de todas las tecnologías a través del tiempo en Streamlit - Mapa	

Planteamiento del problema

- A pesar del creciente consenso científico sobre la urgencia de mitigar el cambio climático, las **emisiones de dióxido de carbono (CO2)** continúan siendo una preocupación central.

La **descarbonización de la economía global** requiere una estrategia multifacética que incluye la reducción de la dependencia de combustibles fósiles, el aumento de la eficiencia energética y el desarrollo e implementación de tecnologías para capturar el CO2 emitido.



https://ourworldindata.org/grapher/annual-co2-emissions-per-country?tab=map&country=USA~GBR~IND~CHN~FRA~DEU~BRA~COL

Fuente del Data set 1

NATIONAL ENERGY TECHNOLOGY LABORATORY

Research and Programs Business Library

Carbon Capture and Storage Database

Welcome to the National Energy Technology Laboratory's (NETL) Carbon Capture and Storage (CCS) Database, which includes information on active, proposed, and terminated CCS projects worldwide.

Publicly available information has been aggregated to provide a one-stop interactive tool that contains valuable data, including, but not limited to:

- Technologies being developed for capture.
- Evaluation of sites for carbon dioxide (CO_2) storage.
- An estimation of project costs.
- Project description and current status.
- Amount of CO_2 captured/stored.

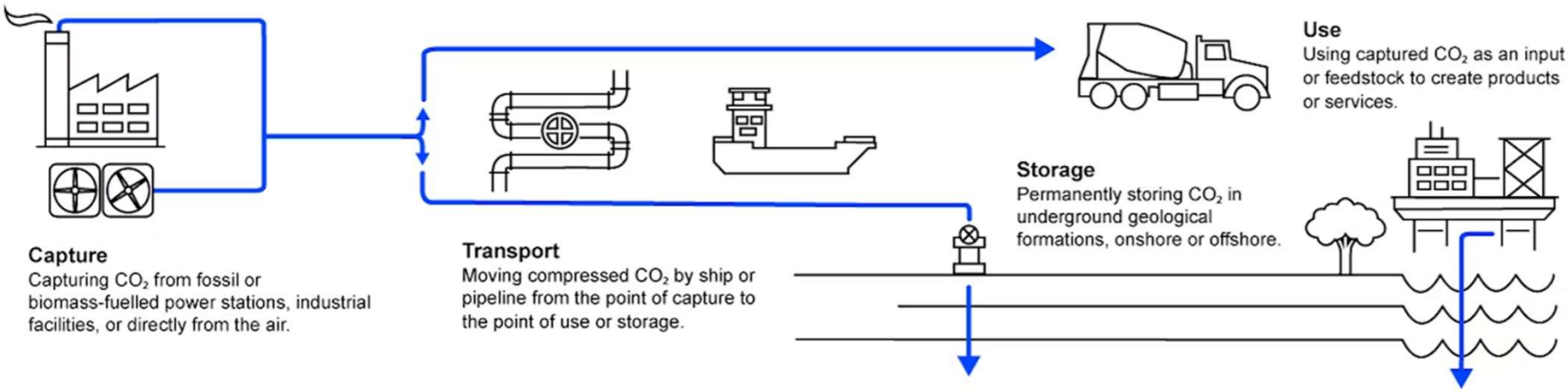
NETL's CCS Database provides the public with information regarding efforts by various industries, public groups, and governments that are being made towards development and eventual deployment of CCS technology. As of April 2018, the database contained 305 total CCS projects worldwide, with 299 sites identified. The 299 site-located projects include 76 capture, 76 storage, and 147 for capture and storage in more than 30 countries across 6 continents. While several of the projects are still in the planning and development stage, and many have been completed, 37 are actively capturing and/or injecting CO_2.

The CCS Database is presented using a Tableau Dashboard which is entirely interactive. Hovering, clicking, and/or dragging any of the icons will customize your view. Adjusting the toolbar on the left allows you to narrow your selection, pan, or zoom in/out on the map.

Diccionario

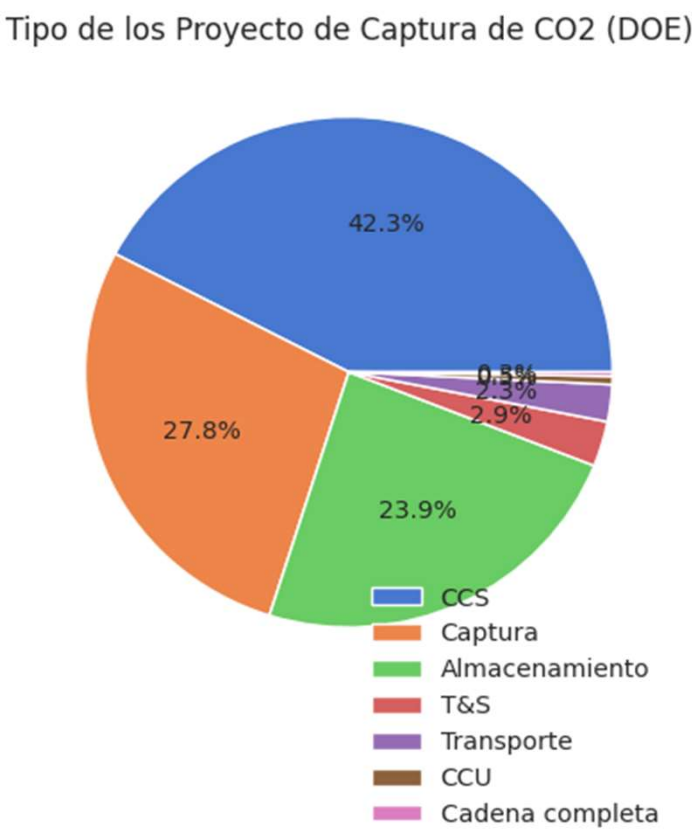
Captura	Captura de CO_2
CCS	De sus siglas en inglés (Capture Carbon Storage) Captura y almacenamiento de CO_2
Almacenamiento	Almacenamiento de CO_2 solamente.
T & S	De sus siglas en inglés (Transport and Storage) Transporte y almacenamiento de CO_2
Transporte	Transporte posterior del almacenamiento de CO_2
CCU	De sus siglas en inglés Carbon, Capture, Utilisation
Cadena completa	Captura, Almacenamiento, Transporte y utilización

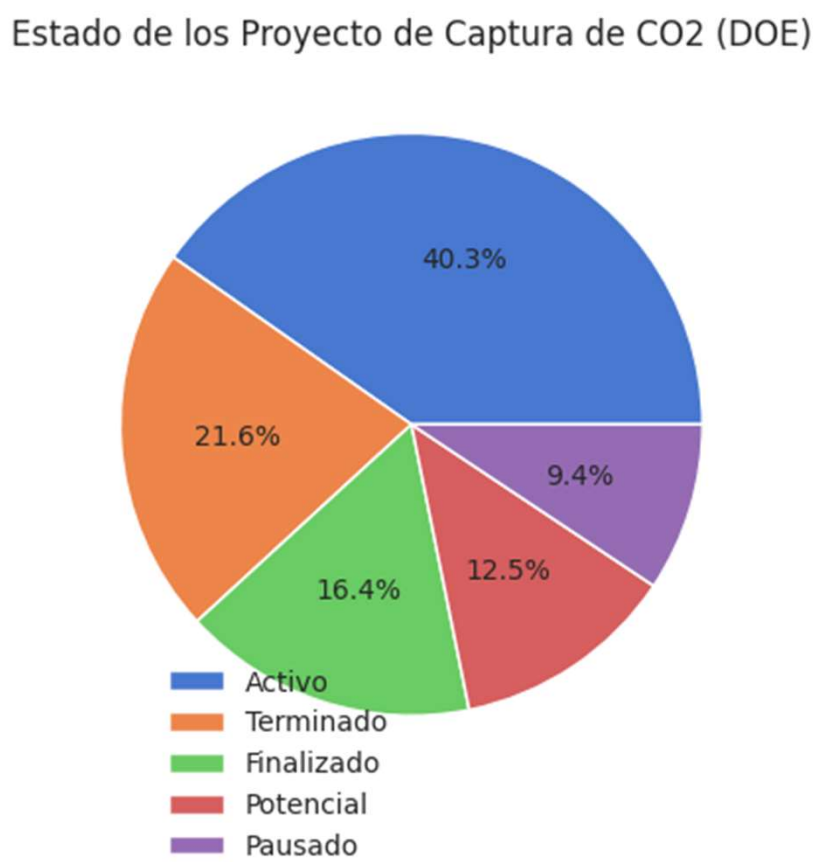
The CCUS Chain



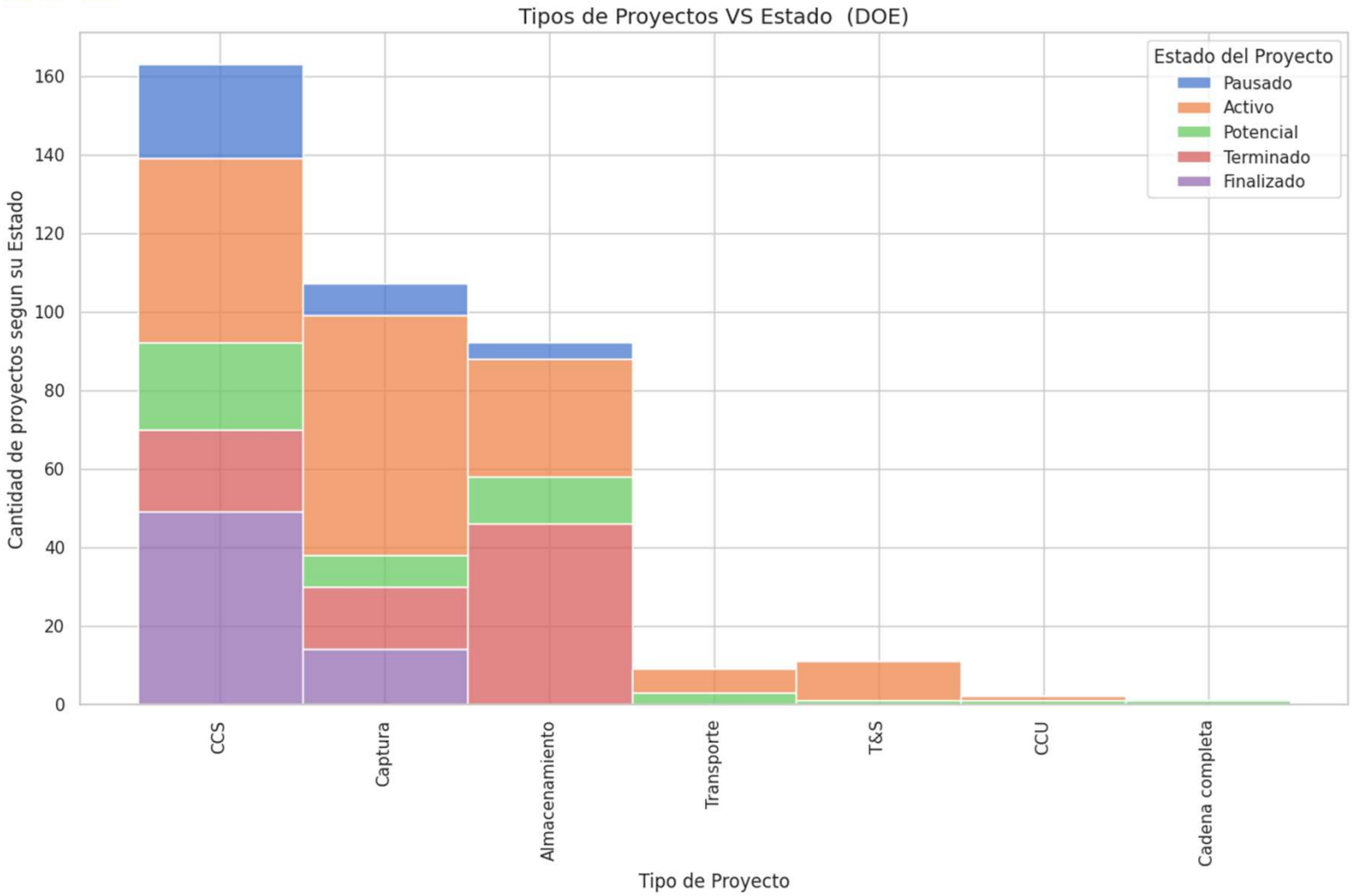
https://www.iea.org/energy-system/carbon-capture-utilisation-and-storage#how-does-ccus-work

Data set 1

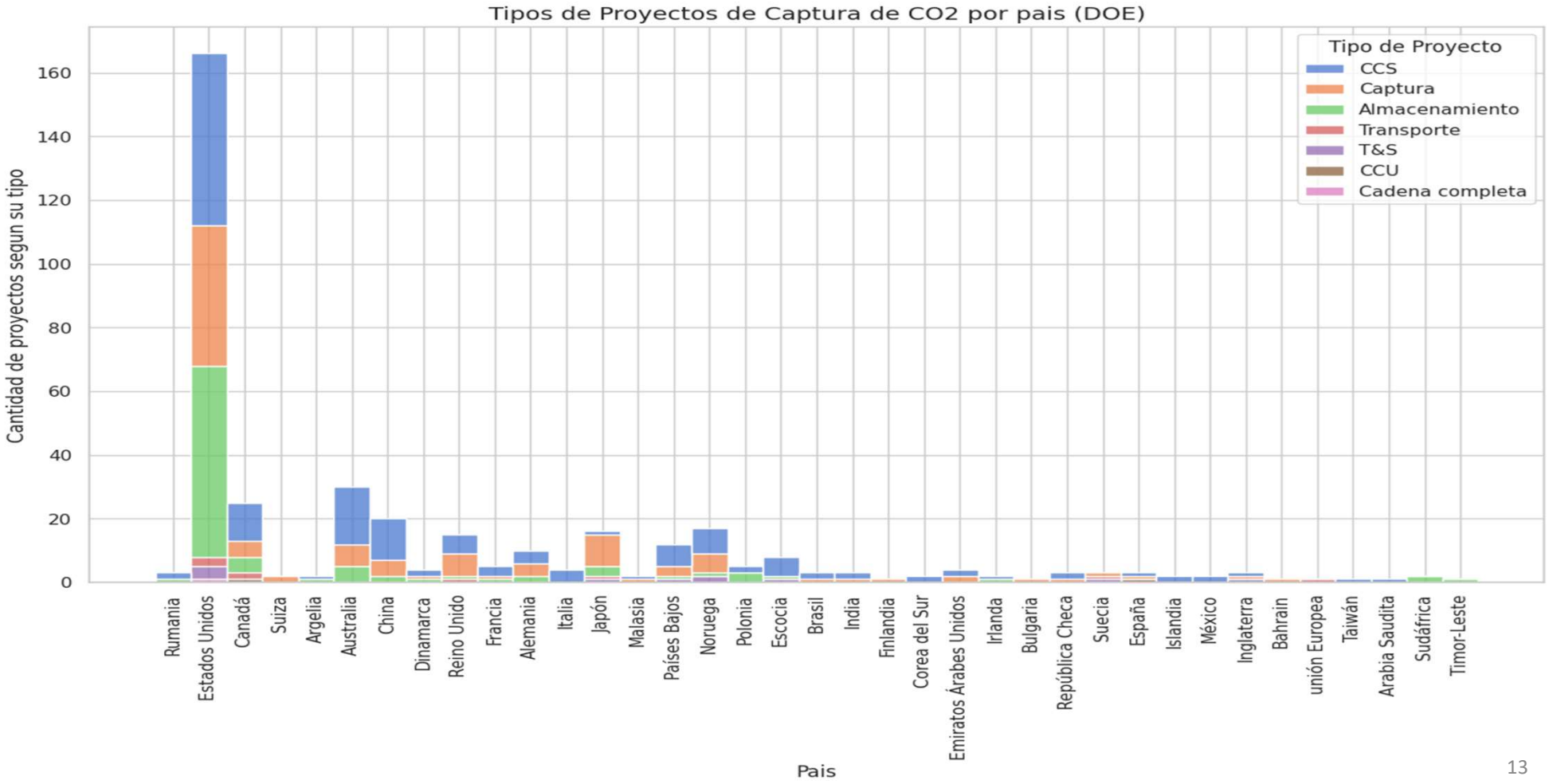

Tipo de los Proyecto de Captura de CO2 (DOE)
42.3%
27.8%
23.9%
2.9%
2.3%
CCS
Captura
Almacenamiento
T&S
Transporte
CCU
Cadena completa


Estado de los Proyecto de Captura de CO2 (DOE)
40.3%
21.6%
16.4%
12.5%
9.4%
Activo
Terminado
Finalizado
Potencial
Pausado

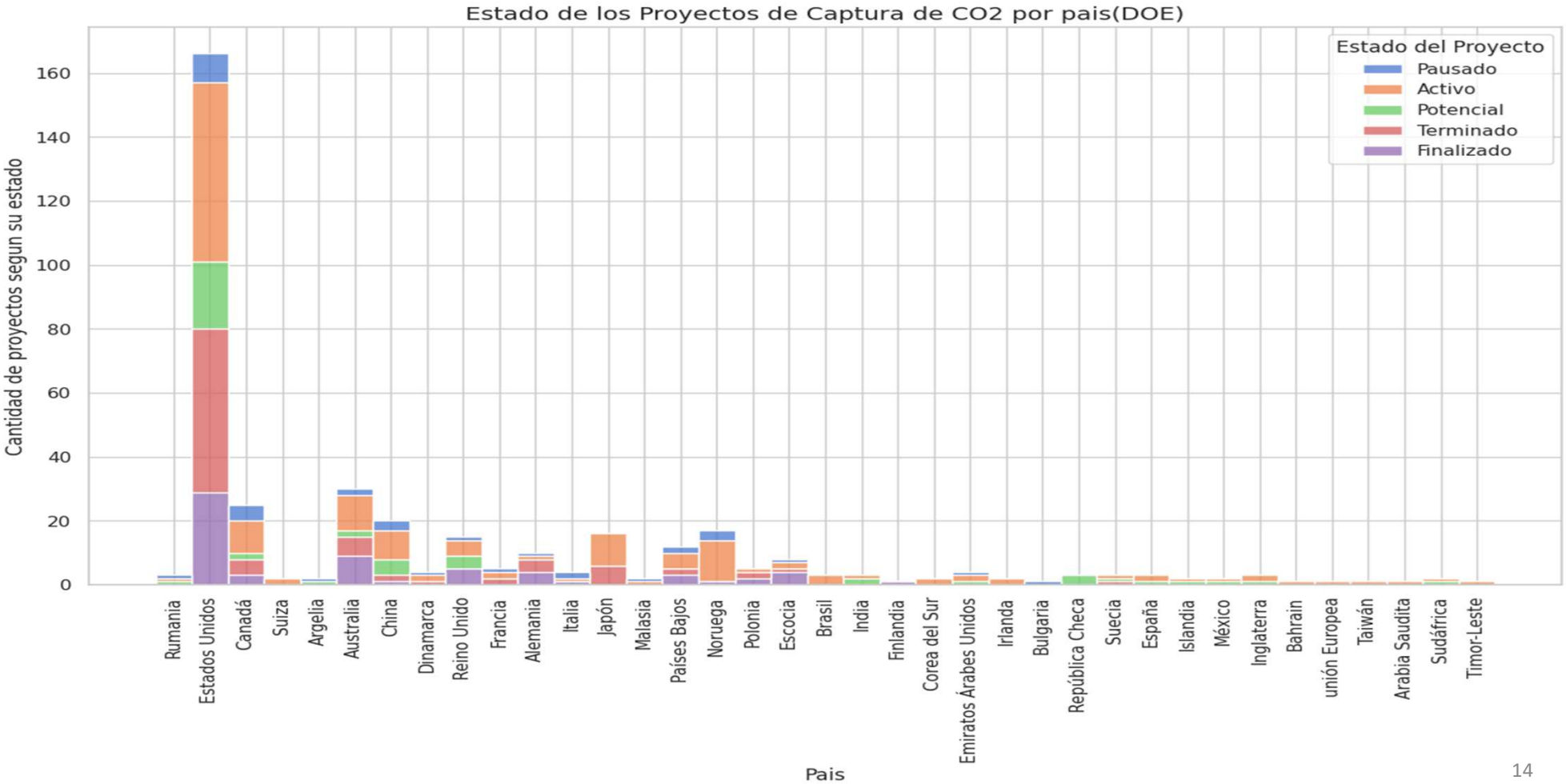
Data set 1


Tipos de Proyectos VS Estado (DOE)
Estado del Proyecto
Pausado
Activo
Potencial
Terminado
Finalizado
Cantidad de proyectos segun su Estado
0
20
40
60
80
100
120
140
160
CCS
Captura
Almacenamiento
Transporte
T&S
CCU
Cadena completa
Tipo de Proyecto

Data set 1


Tipos de Proyectos de Captura de CO2 por pais (DOE)
Tipo de Proyecto
CCS
Captura
Almacenamiento
Transporte
T&S
CCU
Cadena completa
Cantidad de proyectos segun su tipo
0
20
40
60
80
100
120
140
160
Rumania
Estados Unidos
Canadá
Suiza
Argelia
Australia
China
Dinamarca
Reino Unido
Francia
Alemania
Italia
Japón
Malasia
Países Bajos
Noruega
Polonia
Escocia
Brasil
India
Finlandia
Corea del Sur
Emiratos Árabes Unidos
Irlanda
Bulgaria
República Checa
Suecia
España
Islandia
México
Inglaterra
Bahrain
unión Europea
Taiwán
Arabia Saudita
Sudáfrica
Timor-Leste
Pais

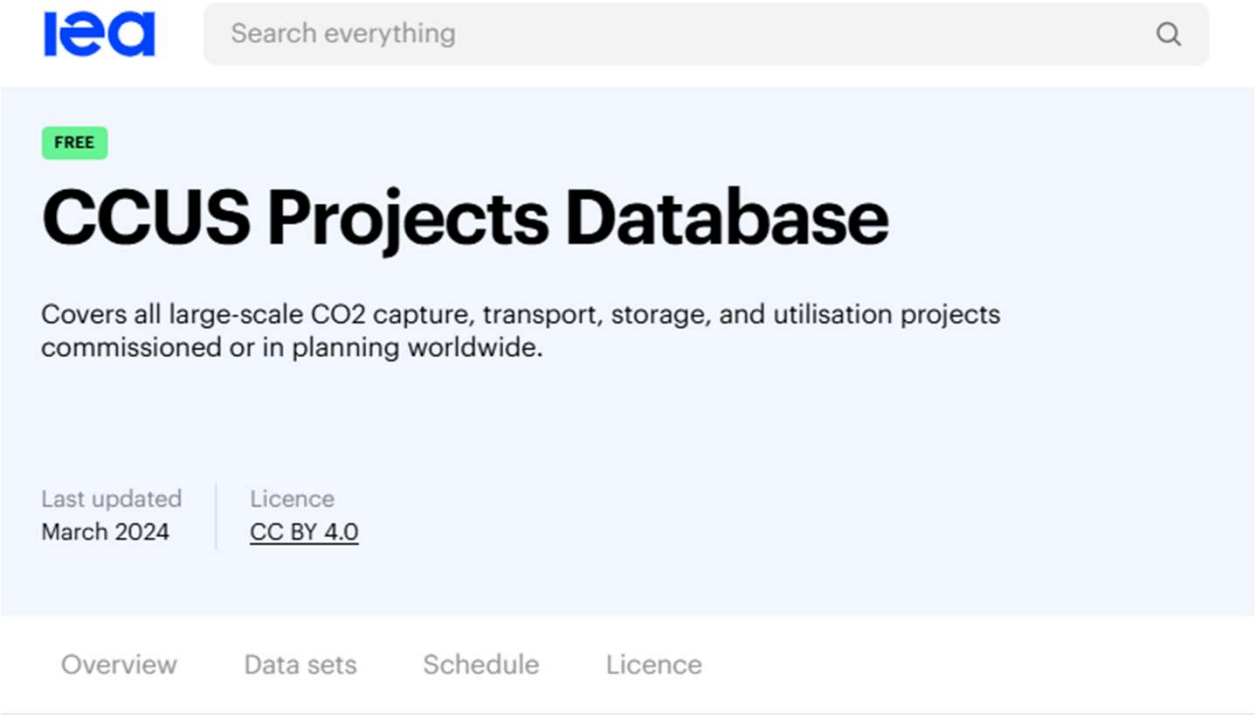
Data set 1


Estado de los Proyectos de Captura de CO2 por pais(DOE)
Estado del Proyecto
Pausado
Activo
Potencial
Terminado
Finalizado
Cantidad de proyectos segun su estado
0
20
40
60
80
100
120
140
160
Rumania
Estados Unidos
Canadá
Suiza
Argelia
Australia
China
Dinamarca
Reino Unido
Francia
Alemania
Italia
Japón
Malasia
Países Bajos
Noruega
Polonia
Escocia
Brasil
India
Finlandia
Corea del Sur
Emiratos Árabes Unidos
Irlanda
Bulgaria
República Checa
Suecia
España
Islandia
México
Inglaterra
Bahrain
unión Europea
Taiwán
Arabia Saudita
Sudáfrica
Timor-Leste
Pais

Data set2

https://www.iea.org/energy-system/carbon-capture-utilisation-and-storage/co2-capture-and-utilisation

Fuente del Data set 2



Data set 2 Visualización de datos

skimpy summary

Data Summary

Dataframe	Values
Number of rows	844
Number of columns	17

Data Types

Column Type	Count
string	11
datetime64	4
float64	2

number

column	NA	NA %	mean	sd	p0	p25	p50	p75	p100	hist
Announced capacity (Mt CO2/yr)	230	27.251184834123222	3.39	11.89	0	0.4	1.2	3	250	
Estimated capacity by IEA (Mt CO2/yr)	199	23.578199052132703	3.45	11.77	0.001	0.4	1.3	3	250	

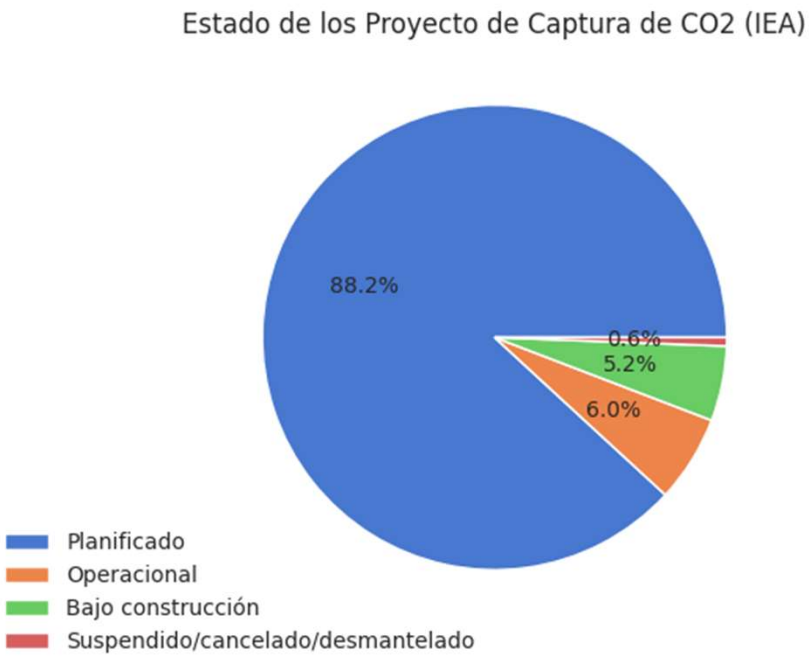
datetime

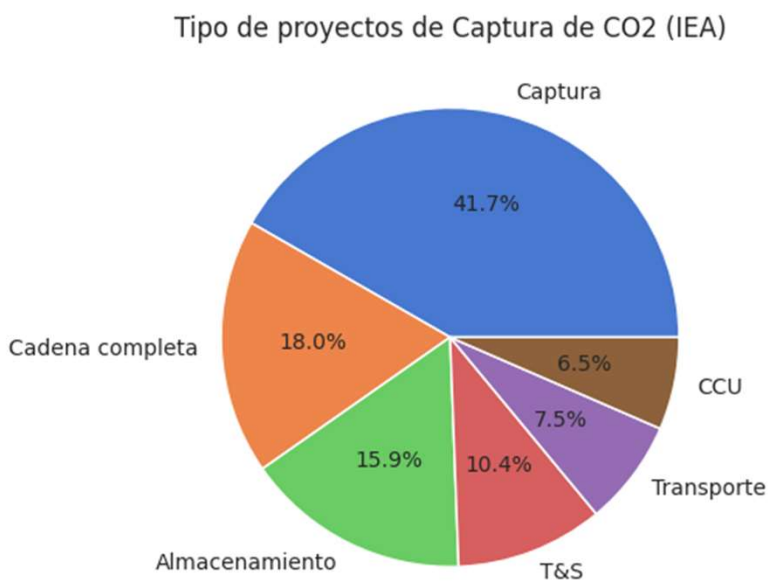
column	NA	NA %	first	last	frequency
Announcement	20	2.3696682464454977	1982-01-01	2024-01-01	None
FID	589	69.7867298578199	1984-01-01	2029-01-01	None
Operation	291	34.47867298578199	1972-01-01	2040-01-01	None
Suspension/decommissioning	838	99.28909952606635	2004-01-01	2026-01-01	None

Data set 2 Estadísticos descriptivos

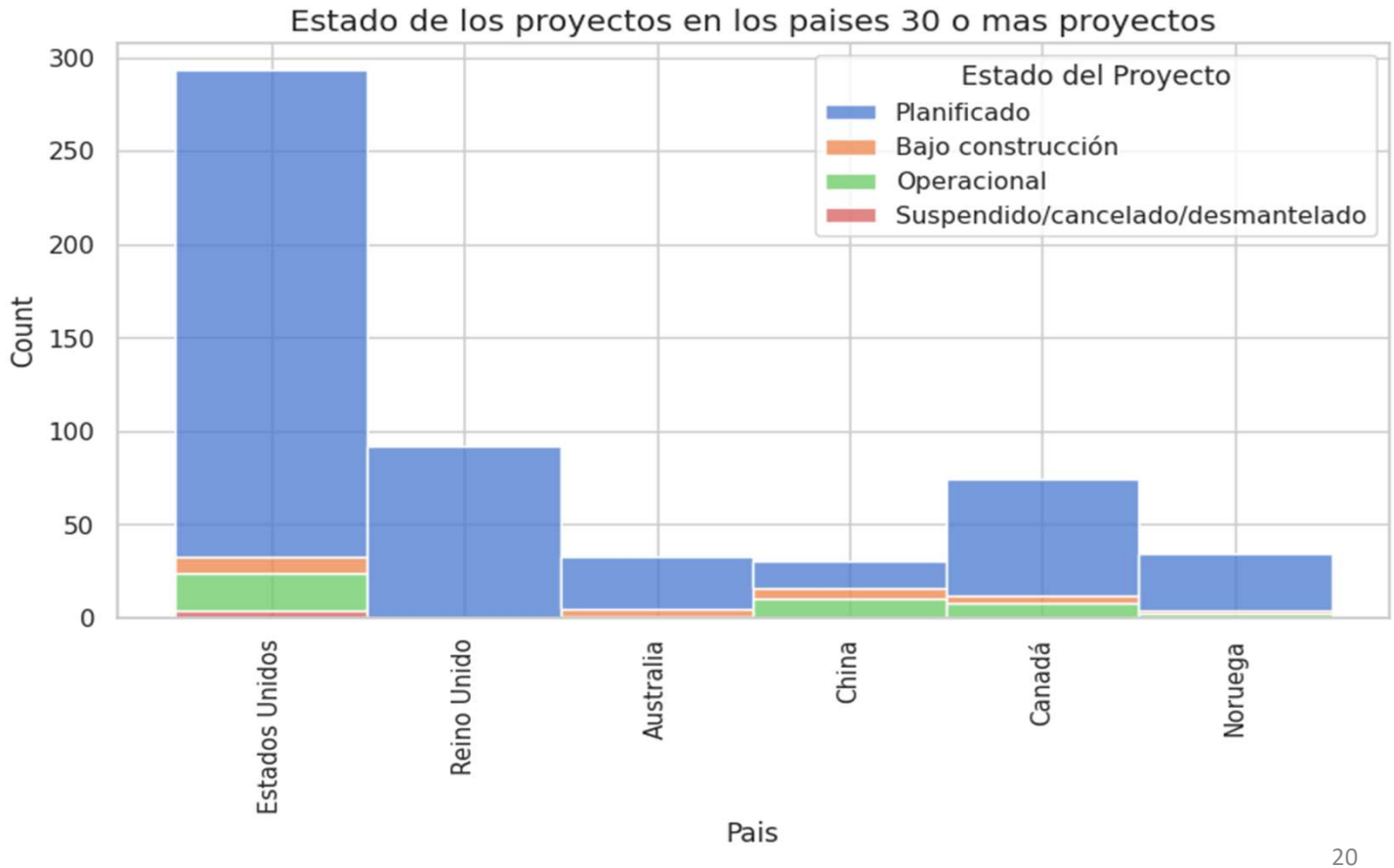
	Announcement	FID	Operation	Suspension/decommissioning	Announced capacity (Mt CO2/yr)	Estimated capacity by IEA (Mt CO2/yr)
count	824	255	553	6	614.000000	645.000000
mean	2020-10-27 23:21:33.203883776	2022-11-06 03:34:35.294117632	2026-03-05 22:47:05.316455680	2017-09-01 04:00:00	3.389782	3.449963
min	1982-01-01 00:00:00	1984-01-01 00:00:00	1972-01-01 00:00:00	2004-01-01 00:00:00	0.000000	0.001000
25%	2021-01-01 00:00:00	2023-01-01 00:00:00	2025-01-01 00:00:00	2012-07-02 00:00:00	0.400000	0.400000
50%	2022-01-01 00:00:00	2024-01-01 00:00:00	2027-01-01 00:00:00	2020-01-01 12:00:00	1.200000	1.300000
75%	2023-01-01 00:00:00	2024-01-01 00:00:00	2029-01-01 00:00:00	2024-07-02 06:00:00	3.000000	3.000000
max	2024-01-01 00:00:00	2029-01-01 00:00:00	2040-01-01 00:00:00	2026-01-01 00:00:00	250.000000	250.000000
std	NaN	NaN	NaN	NaN	11.885486	11.767173

Data set 2 Diagramas de distribución



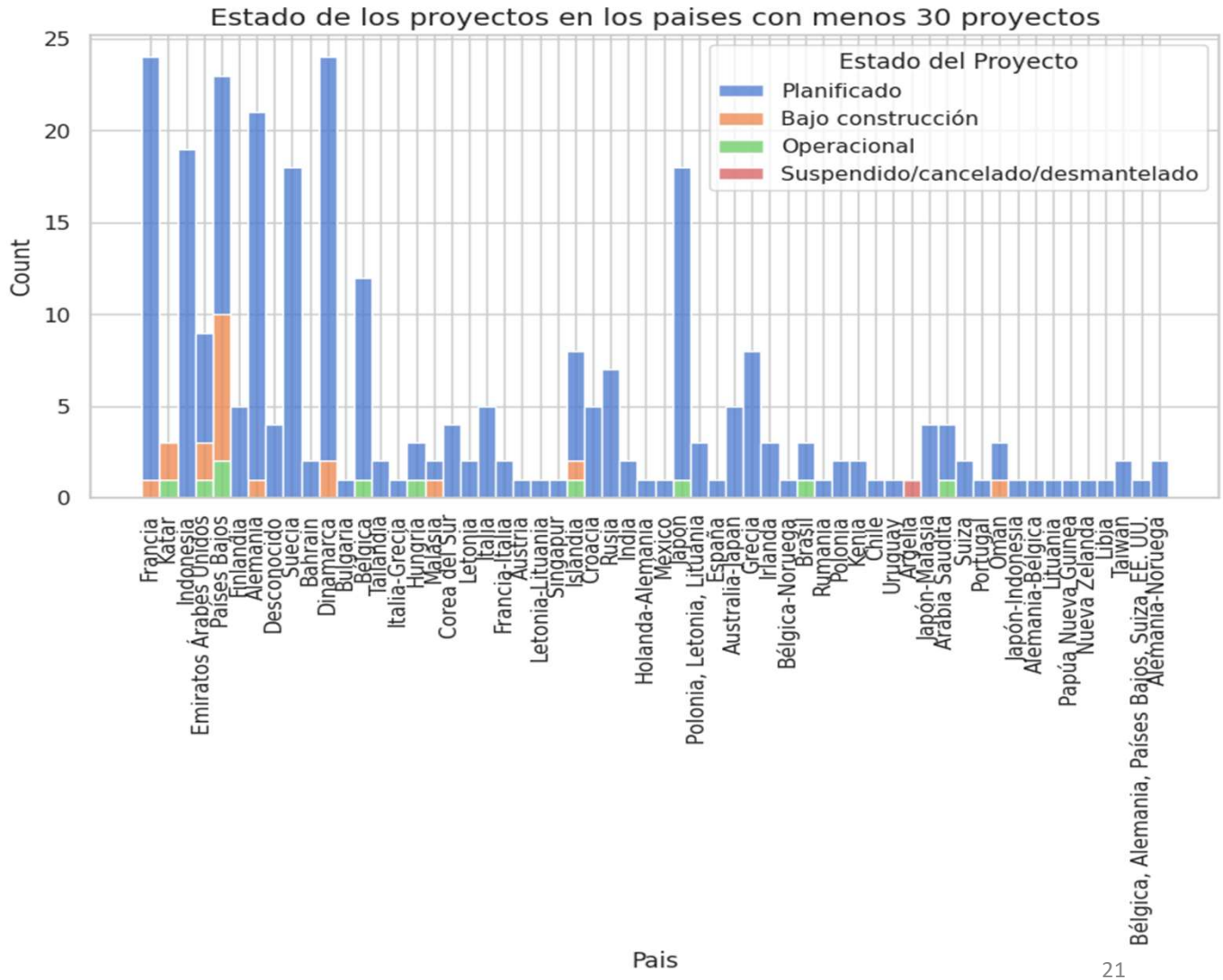


Data set 2


Estado de los proyectos en los paises 30 o mas proyectos
Estado del Proyecto
Planificado
Bajo construcción
Operacional
Suspendido/cancelado/desmantelado
Count
300
250
200
150
100
50
0
Estados Unidos
Reino Unido
Australia
China
Canadá
Noruega
Pais

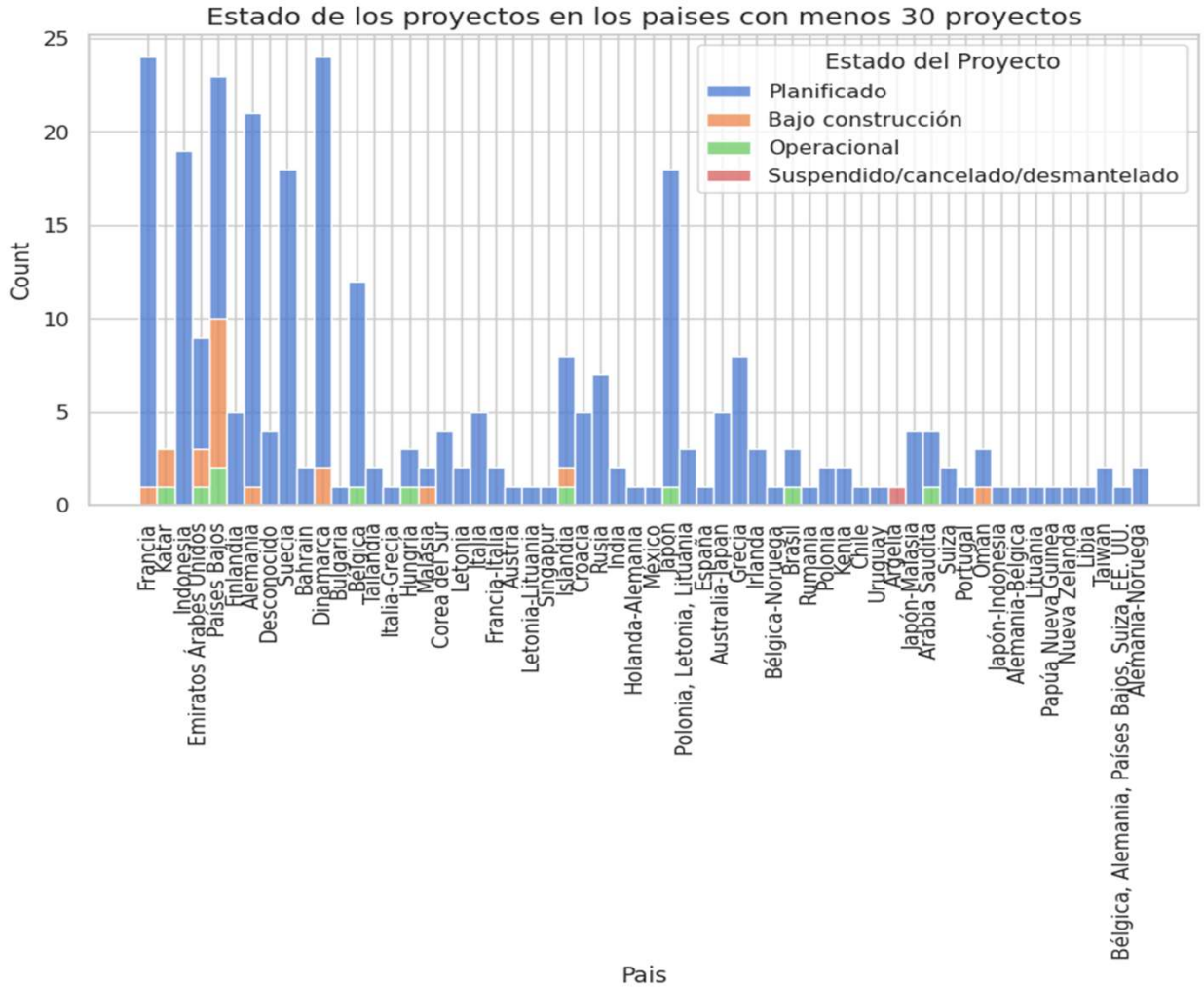
Data set 2

Francia
Katar
Indonesia
Emiratos Árabes Unidos
Países Bajos
Finlandia
Alemania
Desconocido
Suecia
Bahrain
Dinamarca
Bulgaria
Bélgica
Tailandia
Italia-Grecia
Hungría
Malasia
Corea del Sur
Letonia
Italia
Francia-Italia
Austria
Letonia-Lituania
Singapur
Islandia
Croacia
Rusia
India

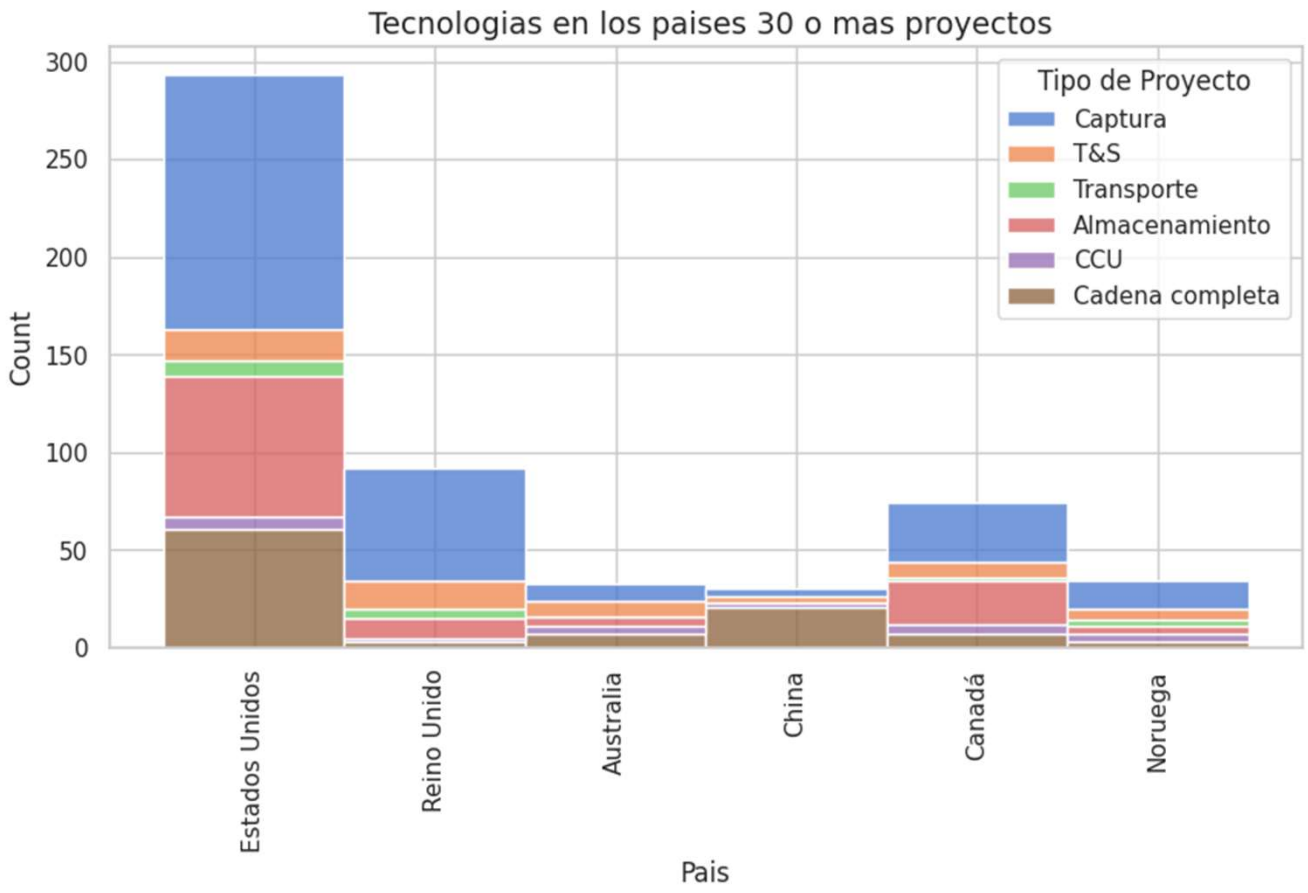

Estado de los proyectos en los paises con menos 30 proyectos
Estado del Proyecto
Planificado
Bajo construcción
Operacional
Suspendido/cancelado/desmantelado
Count
0
5
10
15
20
25
Francia
Katar
Indonesia
Emiratos Árabes Unidos
Países Bajos
Finlandia
Alemania
Desconocido
Suecia
Bahrain
Dinamarca
Bulgaria
Bélgica
Tailandia
Italia-Grecia
Hungría
Malasia
Corea del Sur
Letonia
Italia
Francia-Italia
Austria
Letonia-Lituania
Singapur
Islandia
Croacia
Rusia
India
Holanda-Alemania
Mexico
Japón
Polonia, Letonia, Lituania
España
Australia-Japan
Grecia
Irlanda
Bélgica-Noruega
Brasil
Rumania
Polonia
Kenia
Chile
Uruguay
Argelia
Japón-Malasia
Arabia Saudita
Suiza
Portugal
Oman
Japón-Indonesia
Alemania-Bélgica
Lituania
Papúa Nueva Guinea
Nueva Zelanda
Libia
Taiwan
Bélgica, Alemania, Países Bajos, Suiza, EE. UU.
Alemania-Noruega
Pais

Data set 2

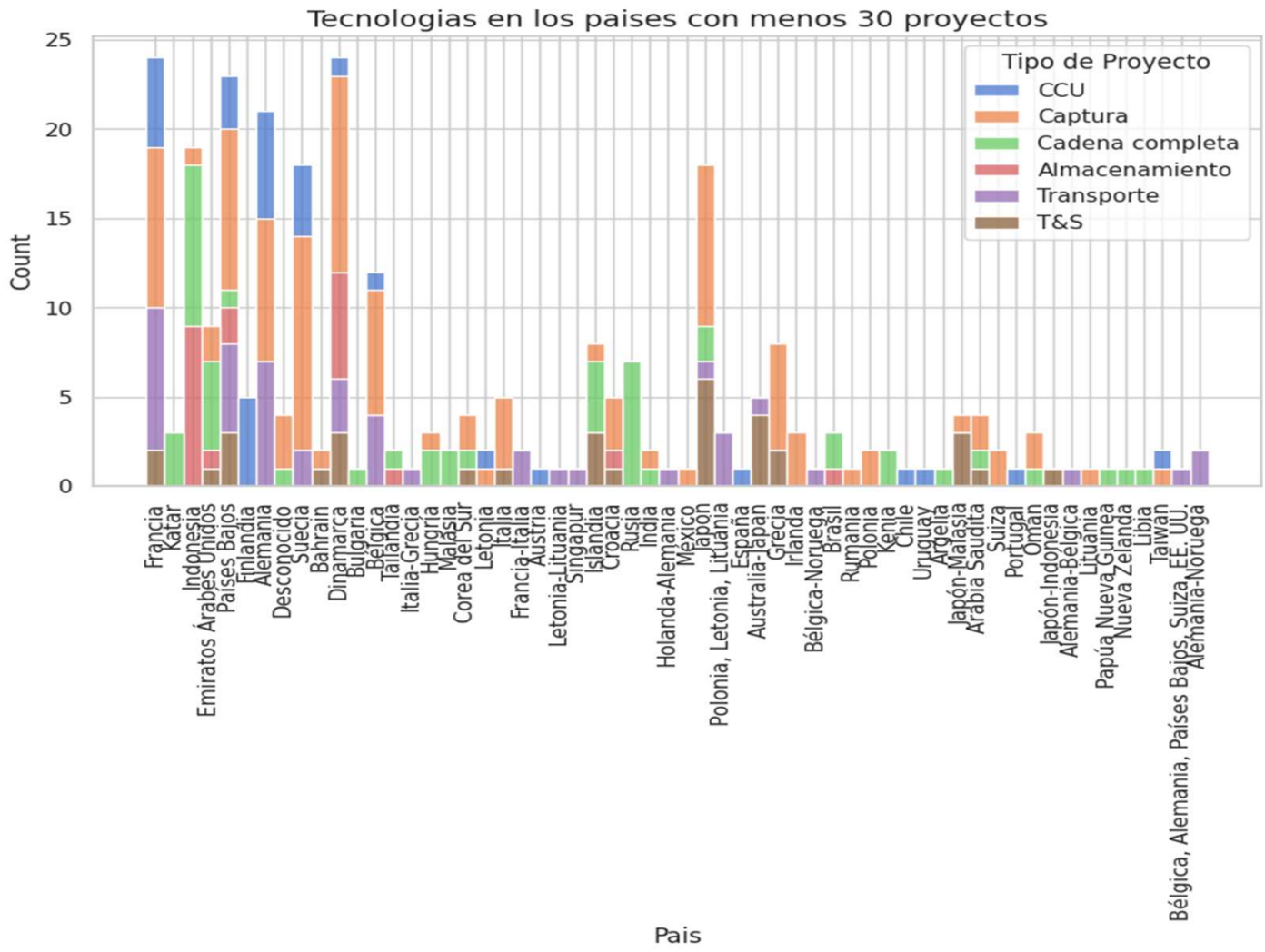
Holanda-Alemania
México
Japón
Polonia, Letonia, Lituania
España
Australia-Japan
Grecia
Irlanda
Bélgica-Noruega
Brasil
Rumania
Polonia
Kenia
Chile
Uruguay
Argelia
Japón-Malasia
Arabia Saudita
Suiza
Portugal
Oman
Japón-Indonesia
Alemania-Bélgica
Lituania
Papúa Nueva Guinea
Nueva Zelanda
Libia
Taiwan
Bélgica, Alemania, Países Bajos, Suiza, EE. UU.
Alemania-Noruega


Estado de los proyectos en los paises con menos 30 proyectos
Estado del Proyecto
Planificado
Bajo construcción
Operacional
Suspendido/cancelado/desmantelado
Count
0
5
10
15
20
25
Francia
Katar
Indonesia
Emiratos Árabes Unidos
Países Bajos
Finlandia
Alemania
Desconocido
Suecia
Bahrain
Dinamarca
Bulgaria
Bélgica
Tailandia
Italia-Grecia
Hungría
Malasia
Corea del Sur
Letonia
Italia
Francia-Italia
Austria
Letonia-Lituania
Singapur
Islandia
Croacia
Rusia
India
Holanda-Alemania
México
Japón
Polonia, Letonia, Lituania
España
Australia-Japan
Grecia
Irlanda
Bélgica-Noruega
Brasil
Rumania
Polonia
Kenia
Chile
Uruguay
Argelia
Japón-Malasia
Arabia Saudita
Suiza
Portugal
Oman
Japón-Indonesia
Alemania-Bélgica
Lituania
Papúa Nueva Guinea
Nueva Zelanda
Libia
Taiwan
Bélgica, Alemania, Países Bajos, Suiza, EE. UU.
Alemania-Noruega
Pais

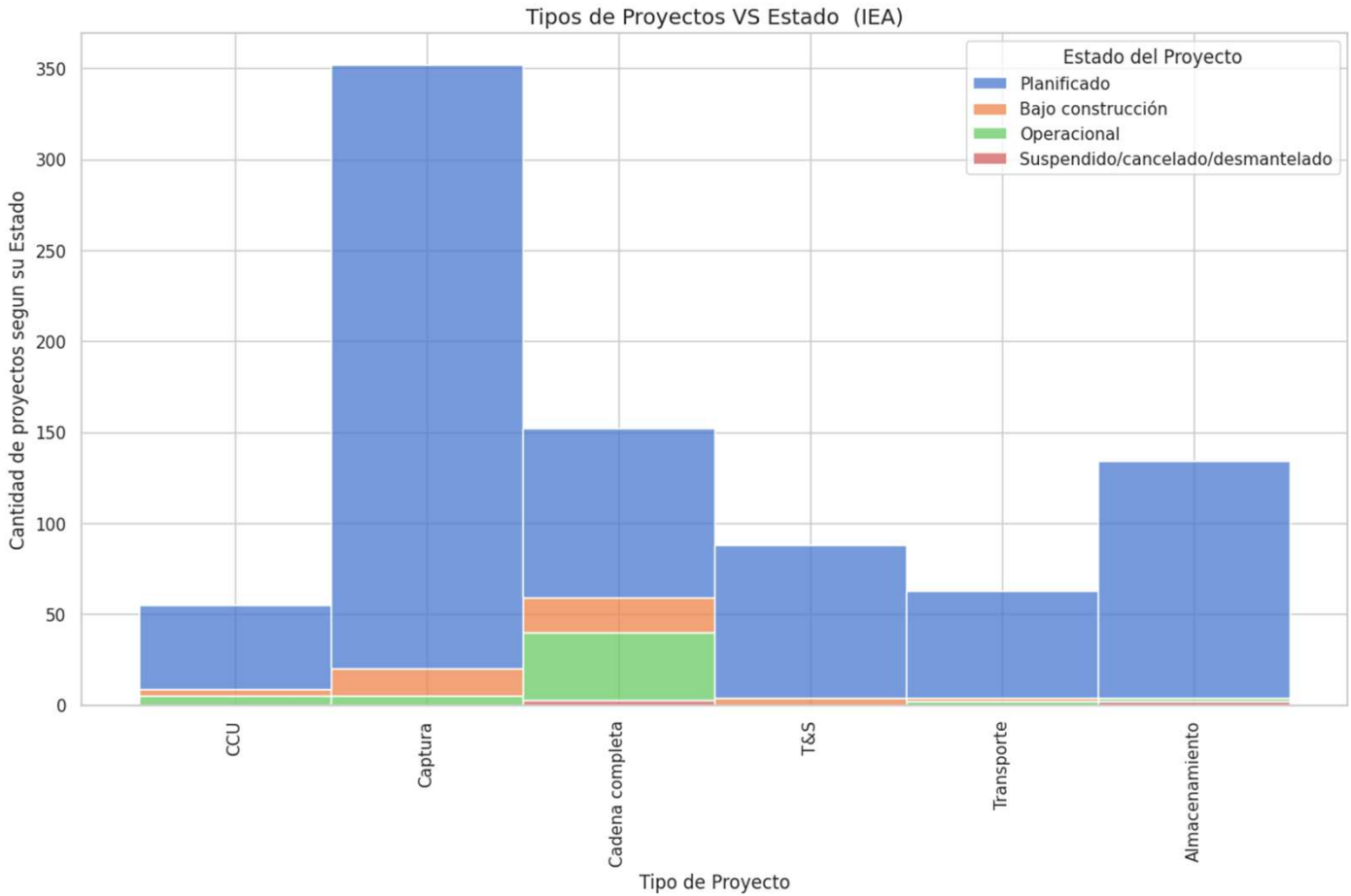
Data set 2


Tecnologias en los paises 30 o mas proyectos
Tipo de Proyecto
Captura
T&S
Transporte
Almacenamiento
CCU
Cadena completa
Count
300
250
200
150
100
50
0
Estados Unidos
Reino Unido
Australia
China
Canadá
Noruega
Pais

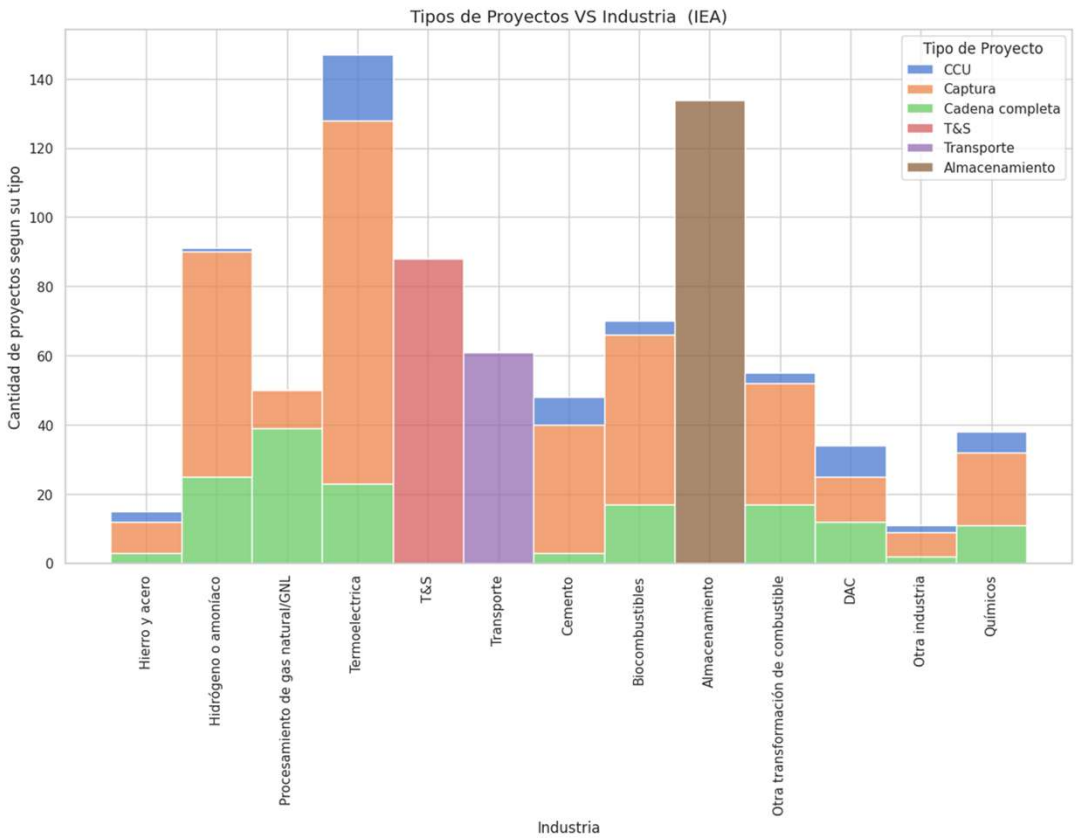
Data set 2


Tecnologias en los paises con menos 30 proyectos
Tipo de Proyecto
CCU
Captura
Cadena completa
Almacenamiento
Transporte
T&S
Count
0
5
10
15
20
25
Francia
Katar
Indonesia
Emiratos Árabes Unidos
Países Bajos
Finlandia
Alemania
Desconocido
Suecia
Bahrain
Dinamarca
Bulgaria
Bélgica
Tailandia
Italia-Grecia
Hungría
Malasia
Corea del Sur
Letonia
Italia
Francia-Italia
Austria
Letonia-Lituania
Singapur
Islandia
Croacia
Rusia
India
Holanda-Alemania
México
Japón
Polonia, Letonia, Lituania
España
Australia-Japan
Grecia
Irlanda
Bélgica-Noruega
Brasil
Rumania
Polonia
Kenia
Chile
Uruguay
Argelia
Japón-Malasia
Arabia Saudita
Suiza
Portugal
Oman
Japón-Indonesia
Alemania-Bélgica
Lituania
Papúa Nueva Guinea
Nueva Zelanda
Libia
Taiwan
Bélgica, Alemania, Países Bajos, Suiza, EE. UU.
Alemania-Noruega
Pais

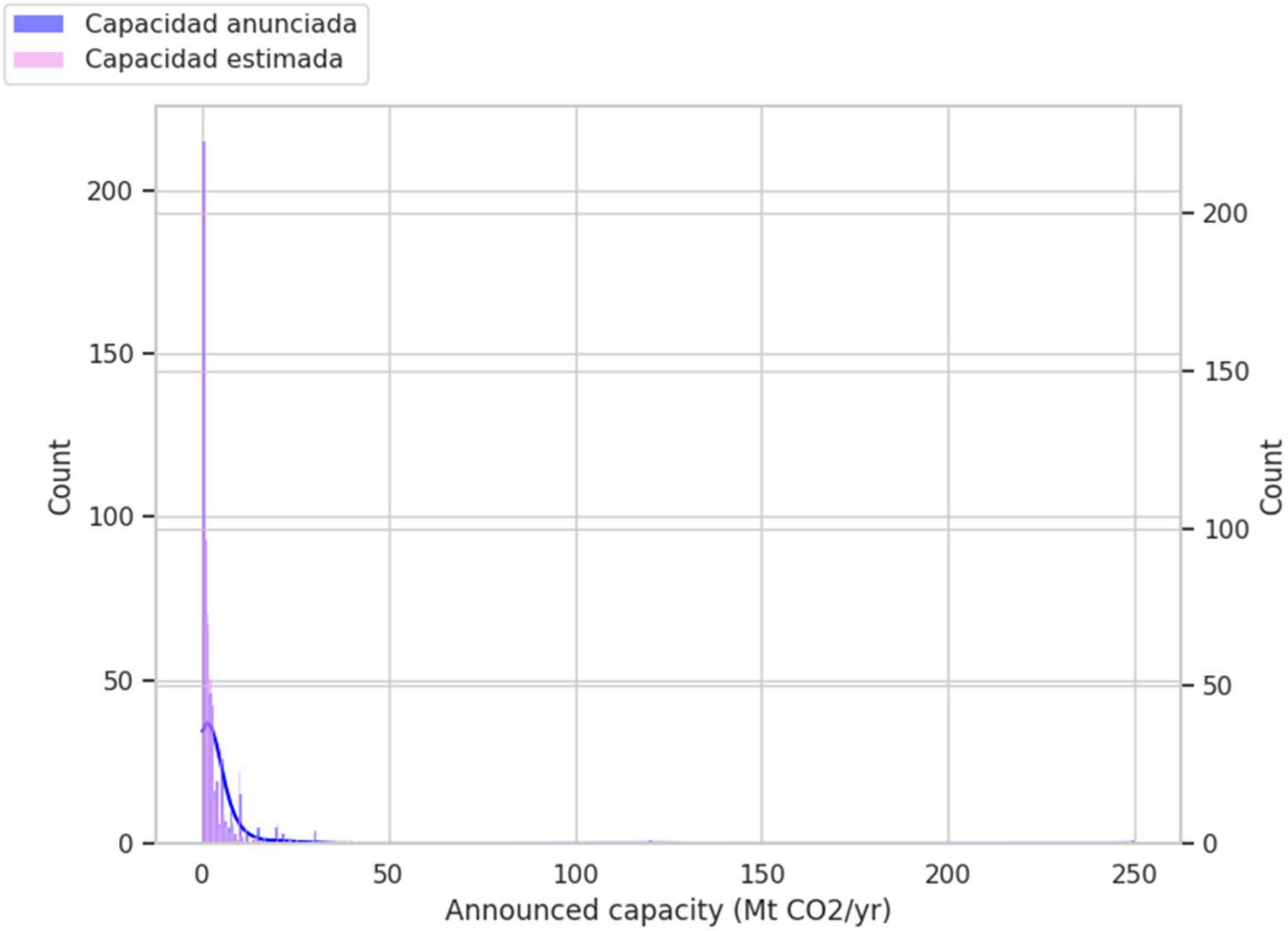
Data set 2


Tipos de Proyectos VS Estado (IEA)
Estado del Proyecto
Planificado
Bajo construcción
Operacional
Suspendido/cancelado/desmantelado
Cantidad de proyectos segun su Estado
0
50
100
150
200
250
300
350
CCU
Captura
Cadena completa
T&S
Transporte
Almacenamiento
Tipo de Proyecto

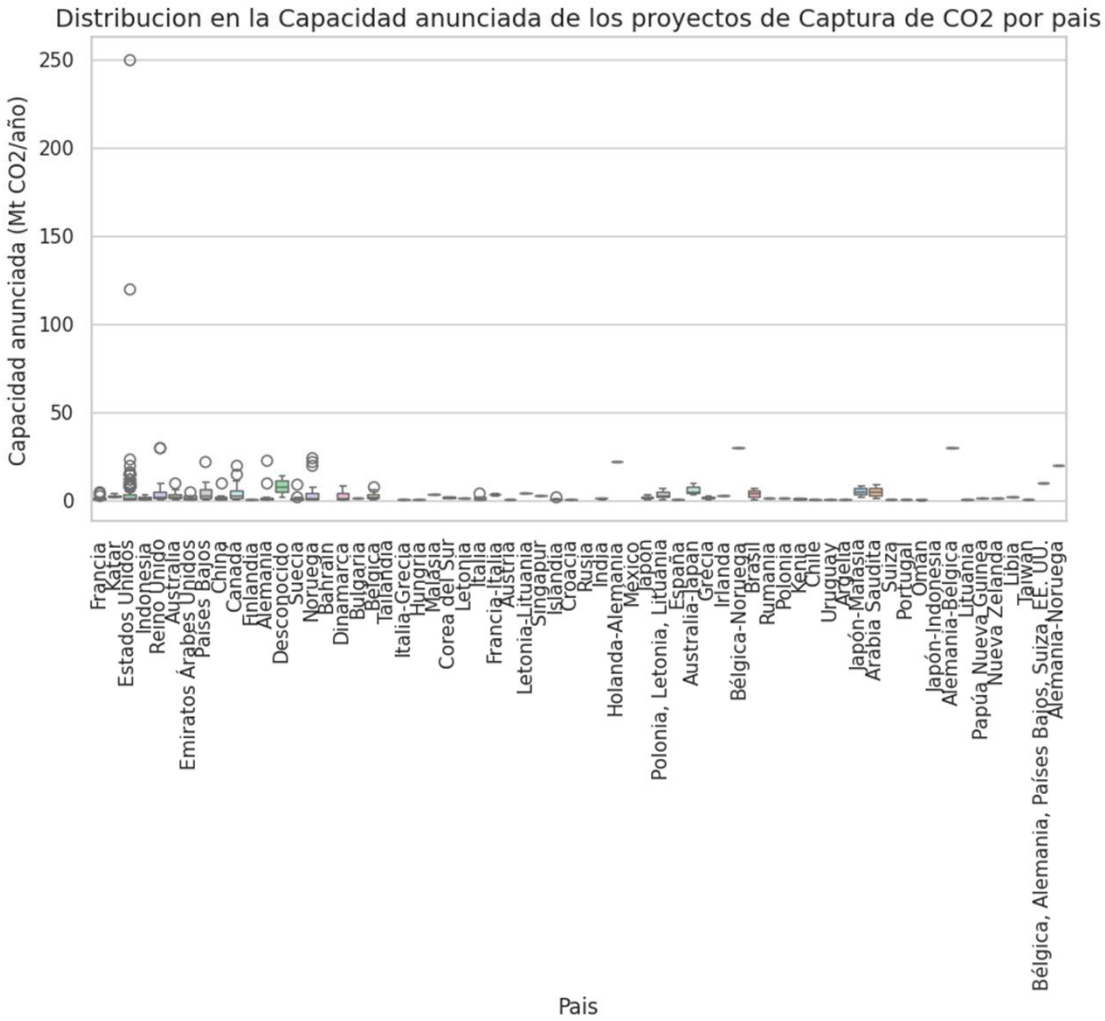
Data set 2


Tipos de Proyectos VS Industria (IEA)
Tipo de Proyecto
CCU
Captura
Cadena completa
T&S
Transporte
Almacenamiento
Cantidad de proyectos segun su tipo
0
20
40
60
80
100
120
140
Hierro y acero
Hidrógeno o amoniaco
Procesamiento de gas natural/GNL
Termoelectrica
T&S
Transporte
Cemento
Biocombustibles
Almacenamiento
Otra transformación de combustible
DAC
Otra industria
Químicos
Industria

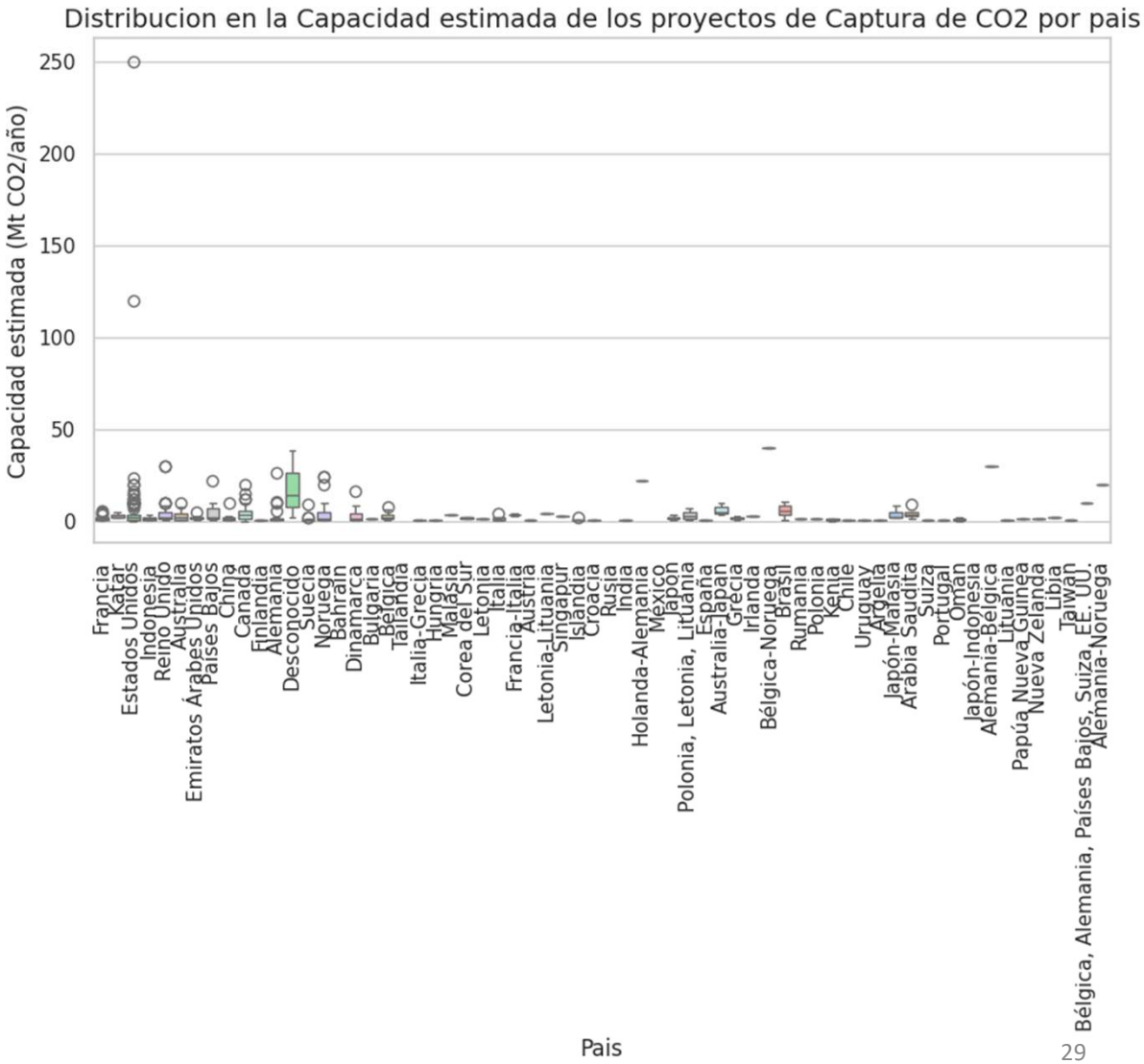
Data set 2
Histograma



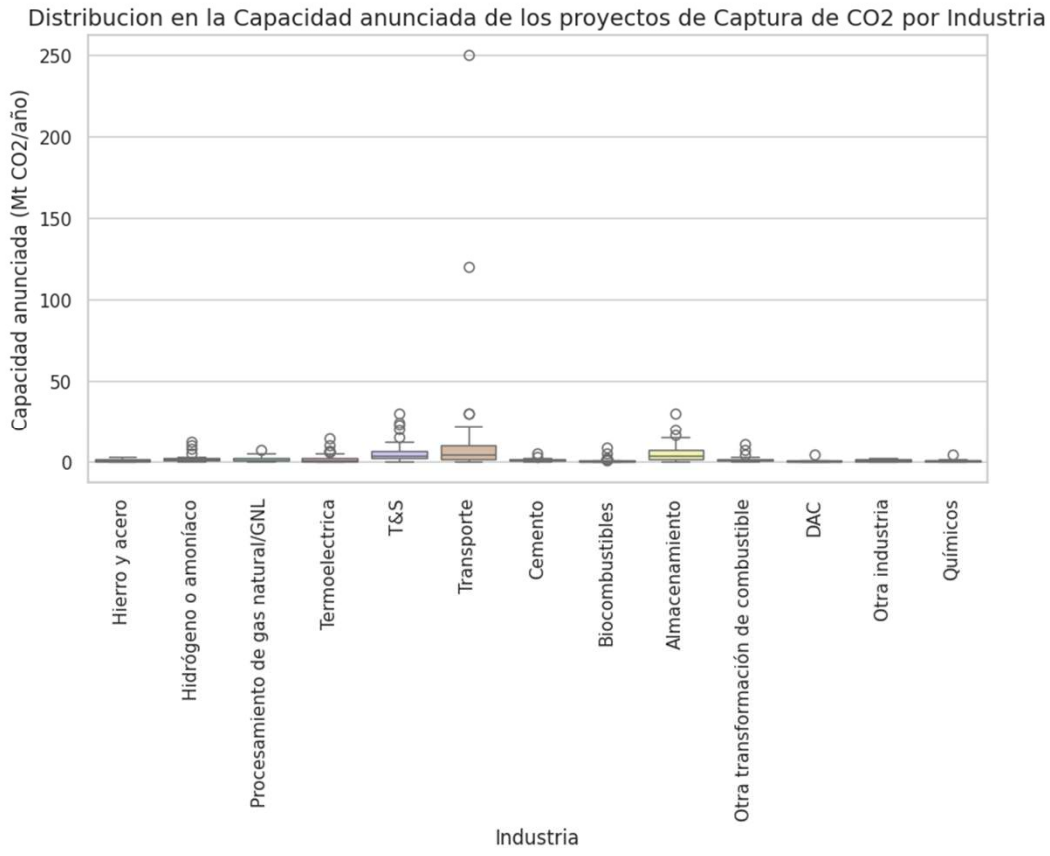
Data set 2
Box plot



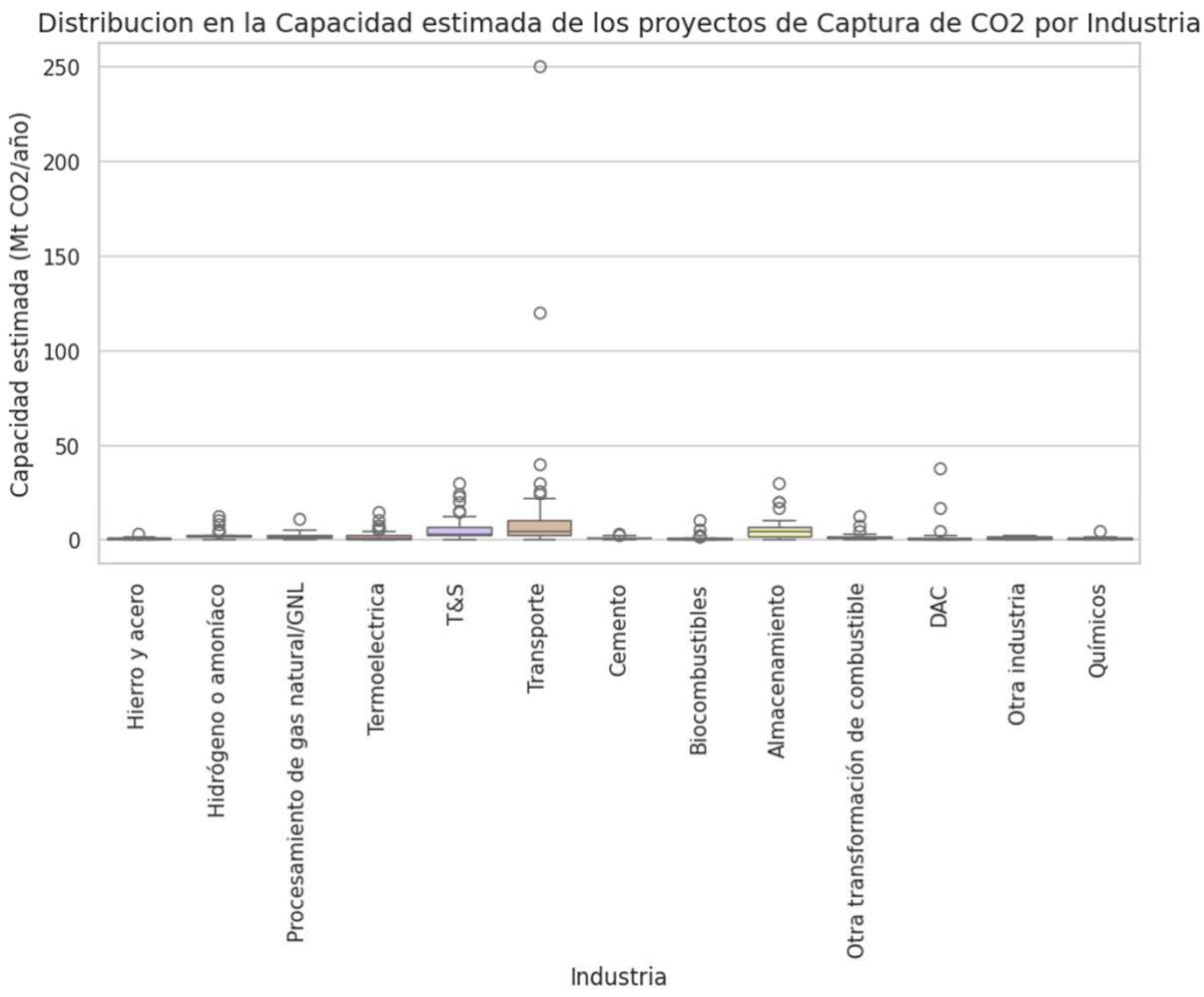
Data set 2
Box plot



Data set 2
Box plot



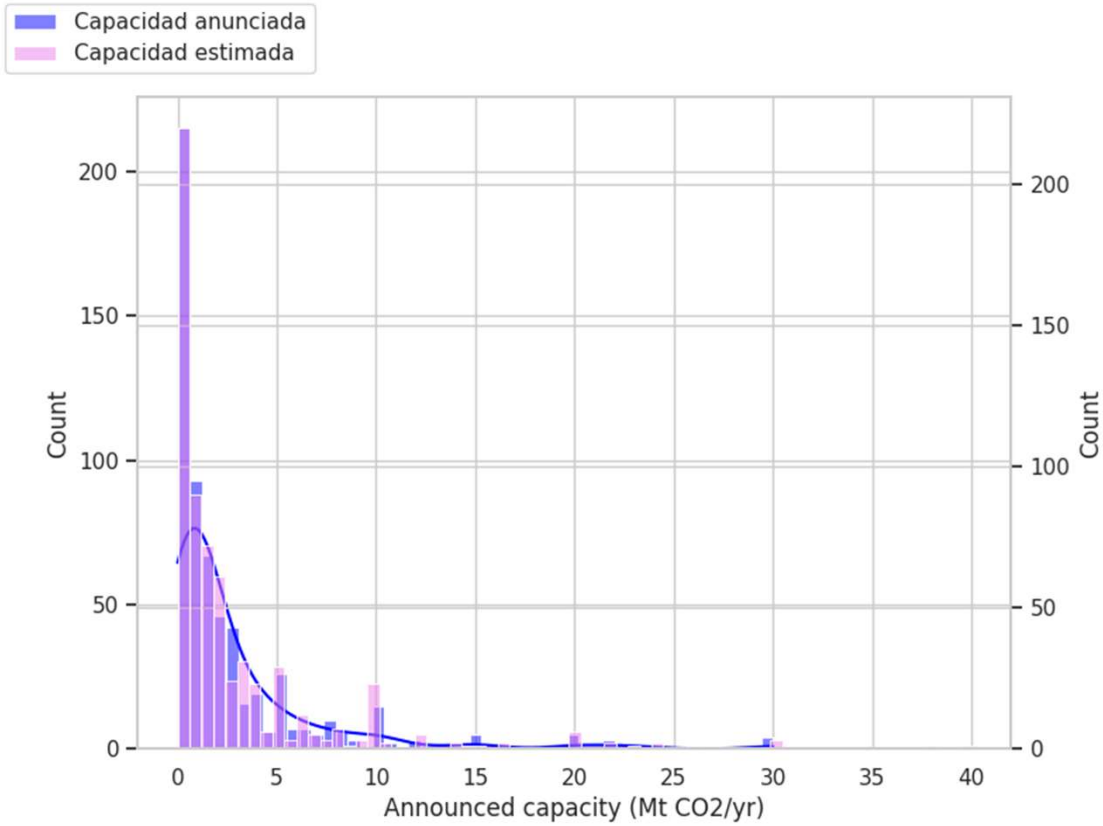
Data set 2 box plot



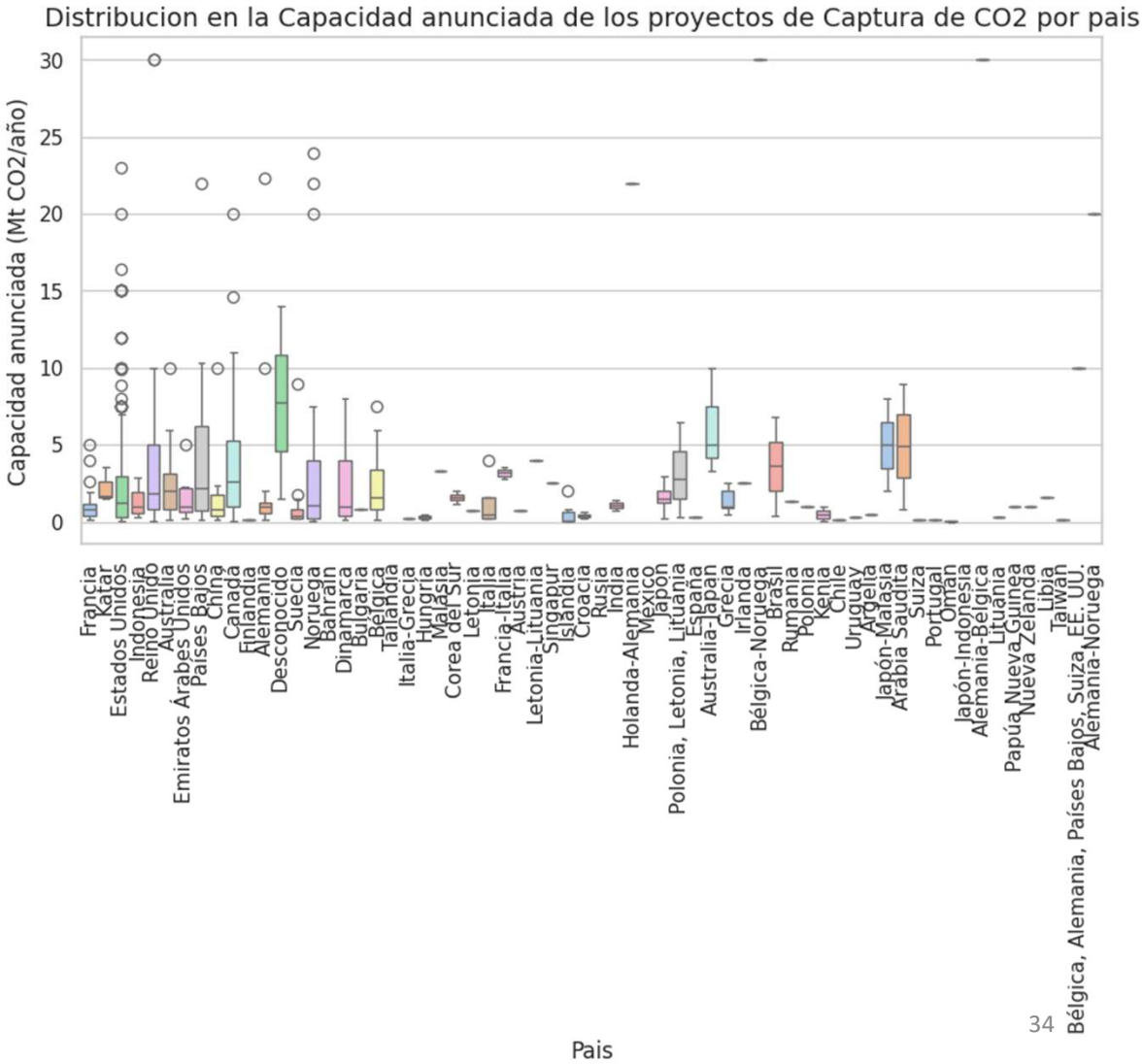
Data set 2 Estadísticos descriptivos sin valores atípicos

	Announcement	FID	Operation	Suspension/decommissioning	Project phase	Announced capacity (Mt CO2/yr)	Estimated capacity by IEA (Mt CO2/yr)
count	822	255	553	6	180.000000	612.000000	643.000000
mean	2020-10-26 00:56:03.503649536	2022-11-06 03:34:35.294117632	2026-03-05 22:47:05.316455680	2017-09-01 04:00:00	1.605556	2.796284	2.885266
min	1982-01-01 00:00:00	1984-01-01 00:00:00	1972-01-01 00:00:00	2004-01-01 00:00:00	1.000000	0.000000	0.001000
25%	2021-01-01 00:00:00	2023-01-01 00:00:00	2025-01-01 00:00:00	2012-07-02 00:00:00	1.000000	0.400000	0.400000
50%	2022-01-01 00:00:00	2024-01-01 00:00:00	2027-01-01 00:00:00	2020-01-01 12:00:00	1.000000	1.200000	1.270000
75%	2023-01-01 00:00:00	2024-01-01 00:00:00	2029-01-01 00:00:00	2024-07-02 06:00:00	2.000000	3.000000	3.000000
max	2024-01-01 00:00:00	2029-01-01 00:00:00	2040-01-01 00:00:00	2026-01-01 00:00:00	4.000000	30.000000	40.000000
std	NaN	NaN	NaN	NaN	0.705107	4.425264	4.768221

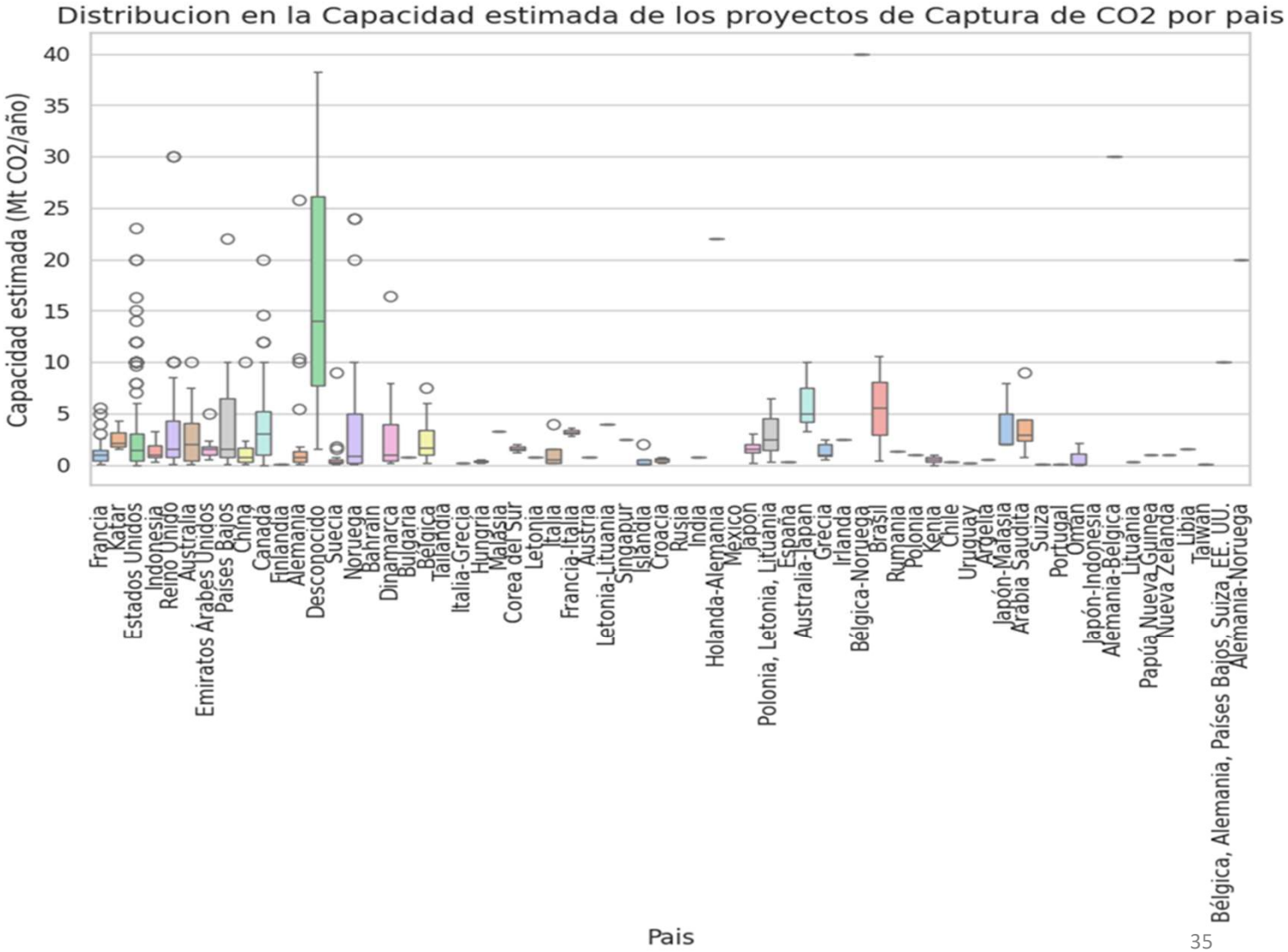
Data set 2
Histograma sin valores atípicos



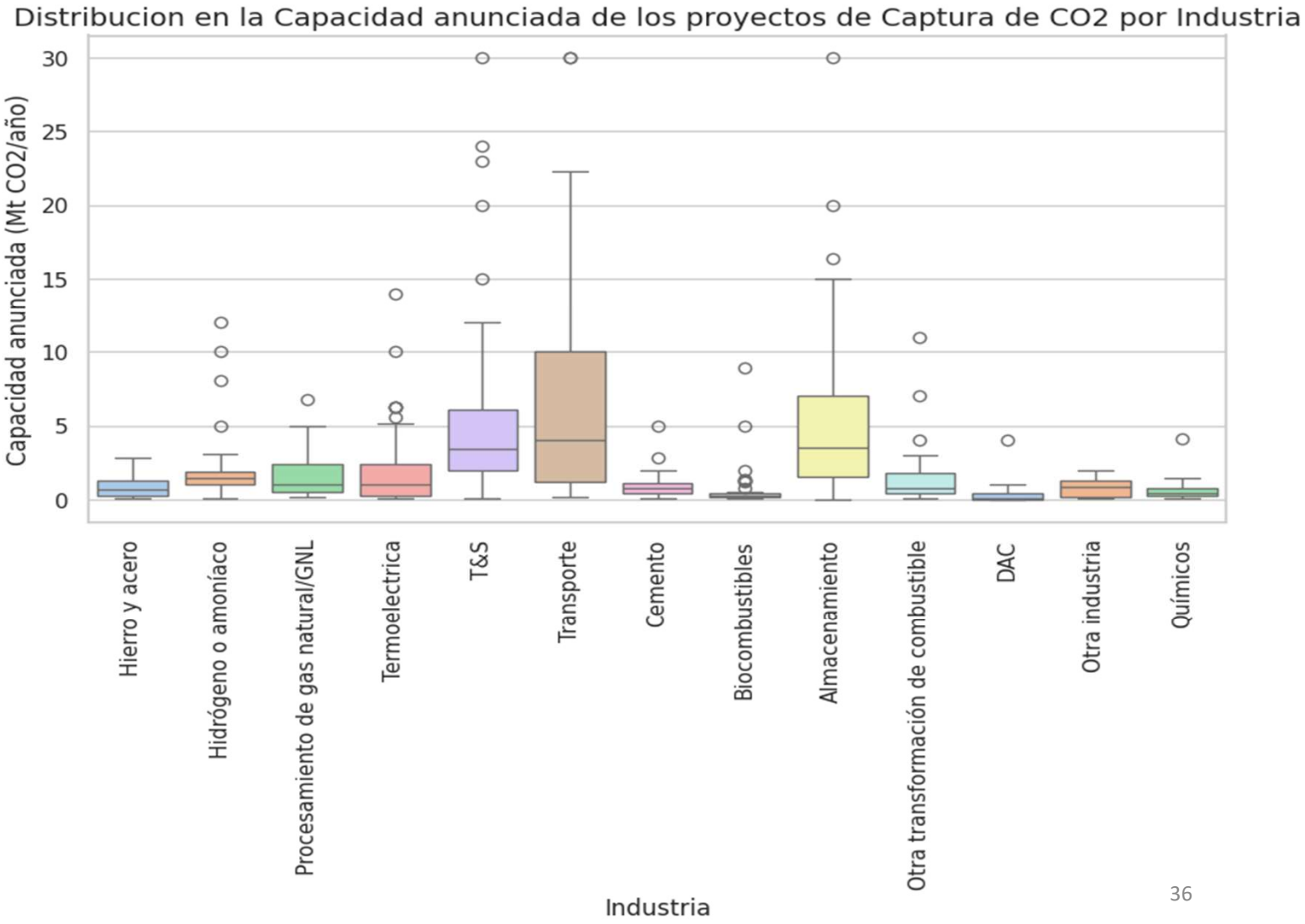
Data set 2
Box Plot sin valores atípicos



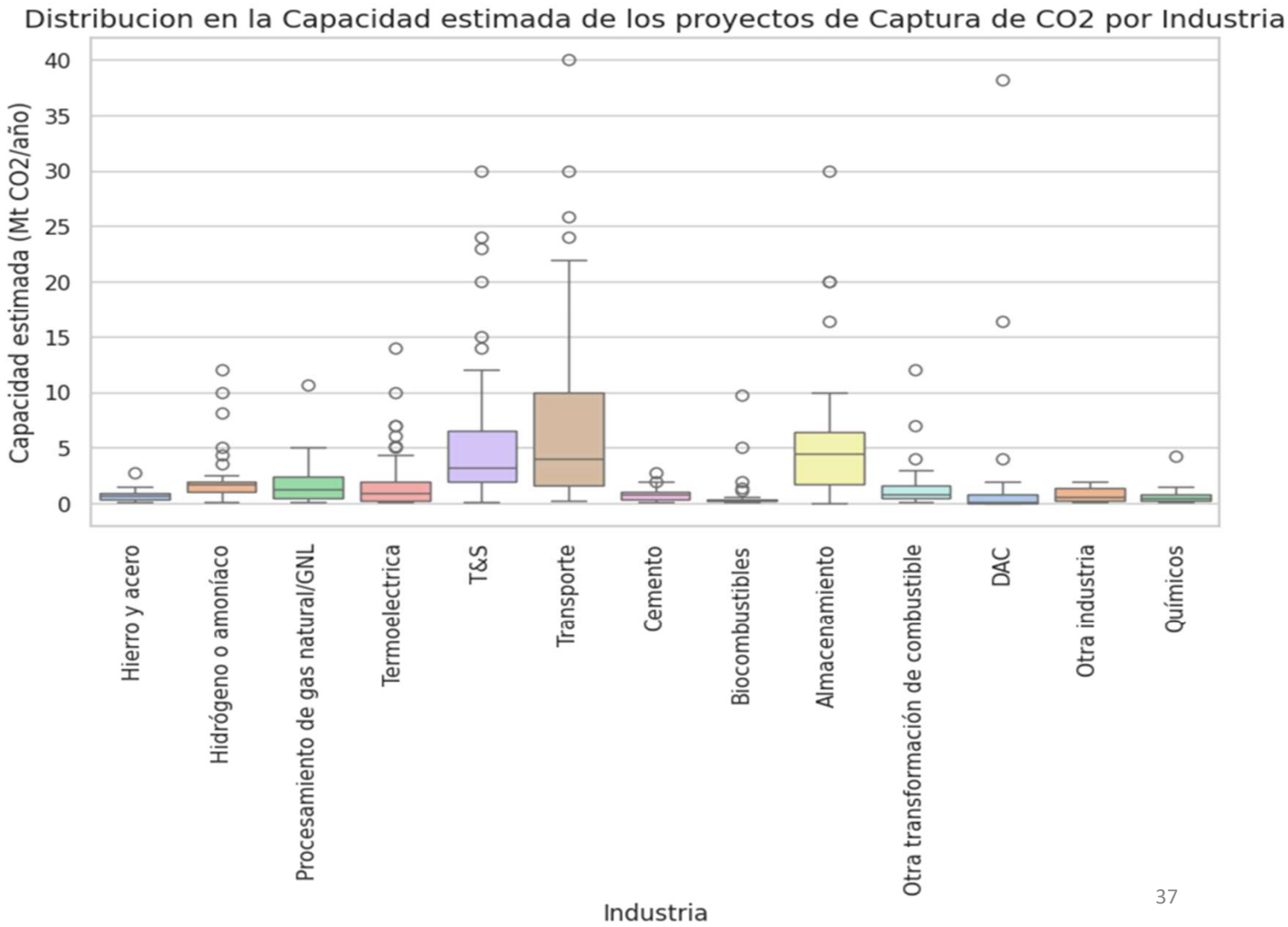
Data set 2 Box Plot sin valores atípicos



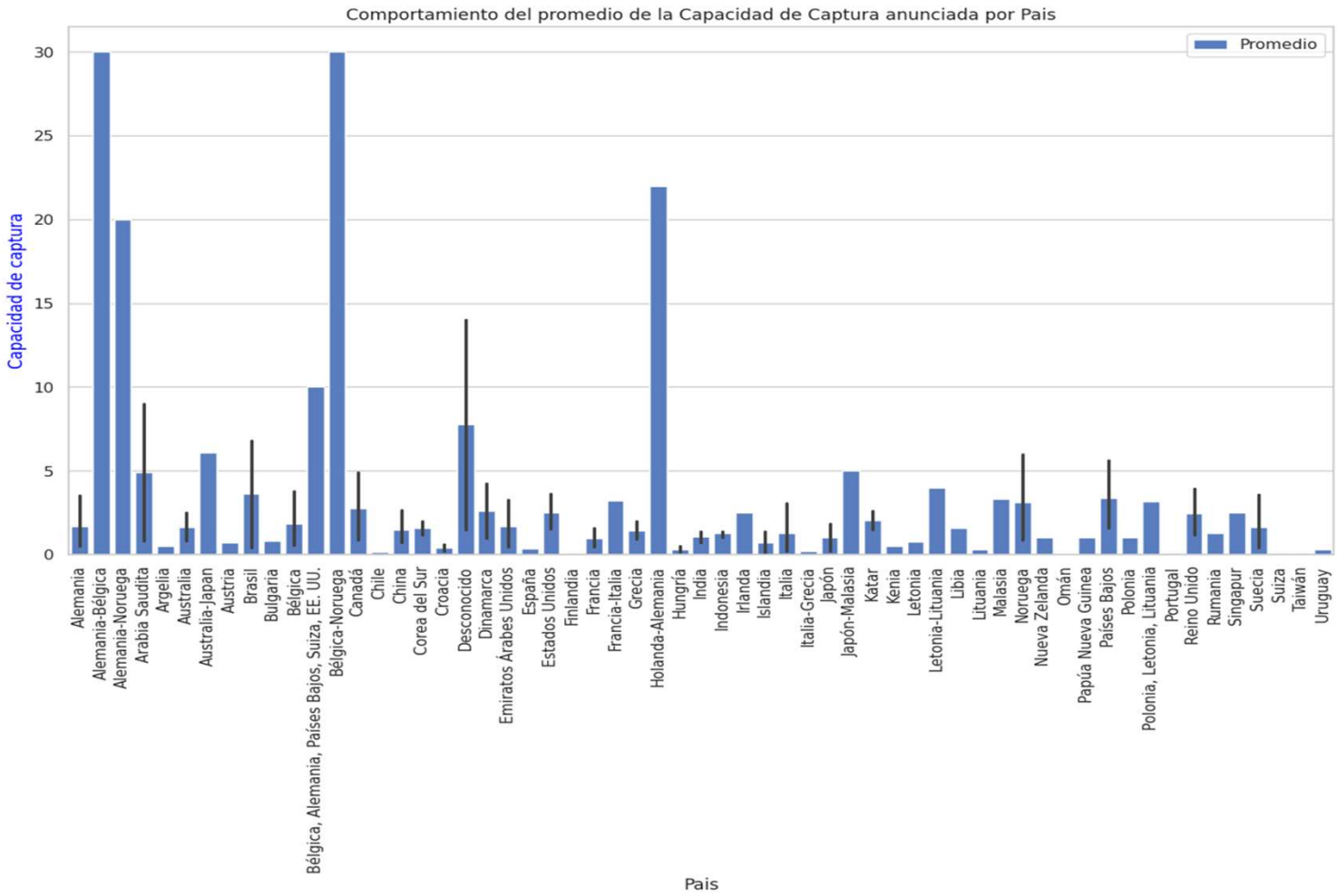
Data set 2 Box Plot sin valores atípicos



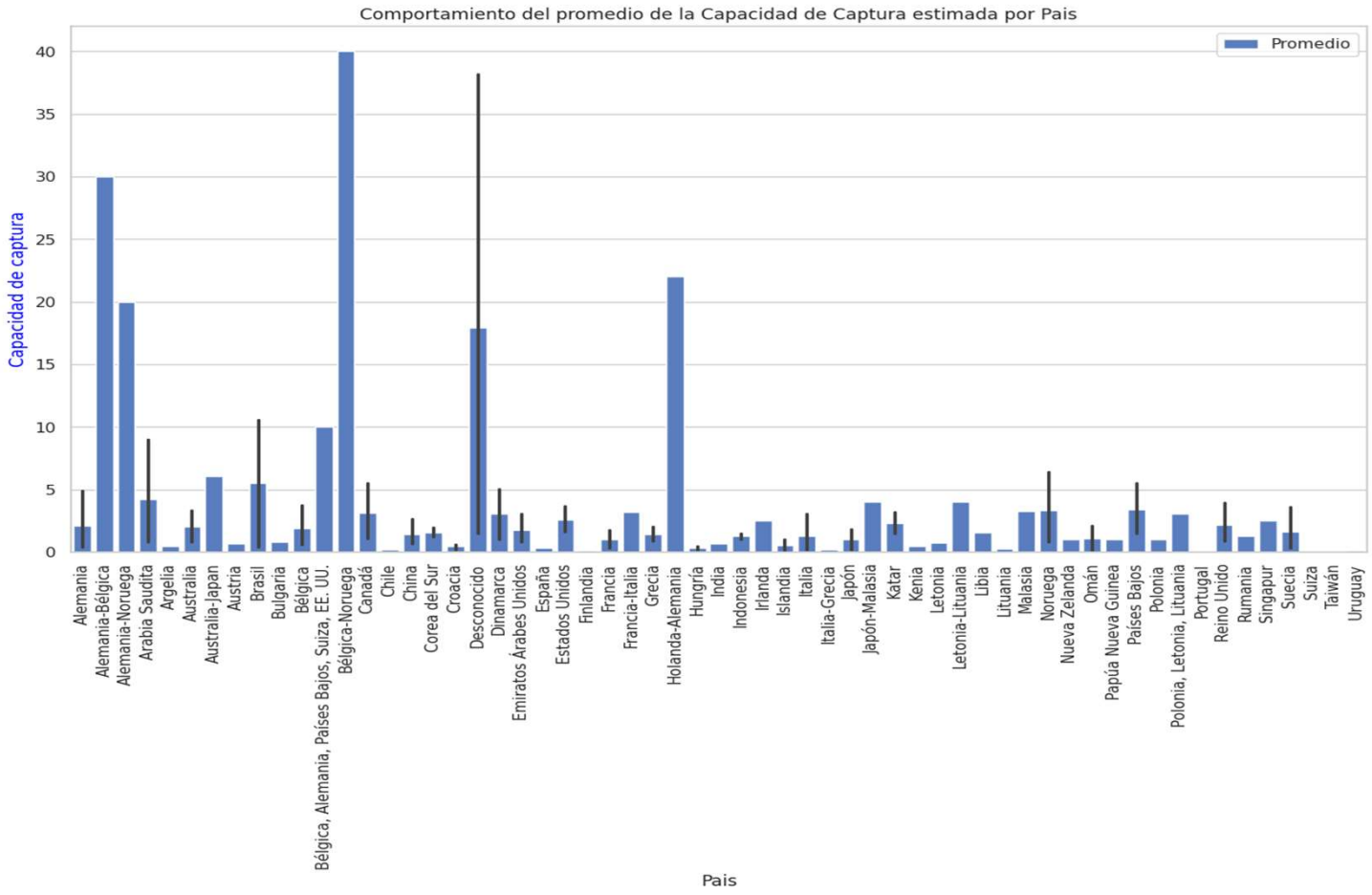
Data set 2 Box Plot sin valores atípicos



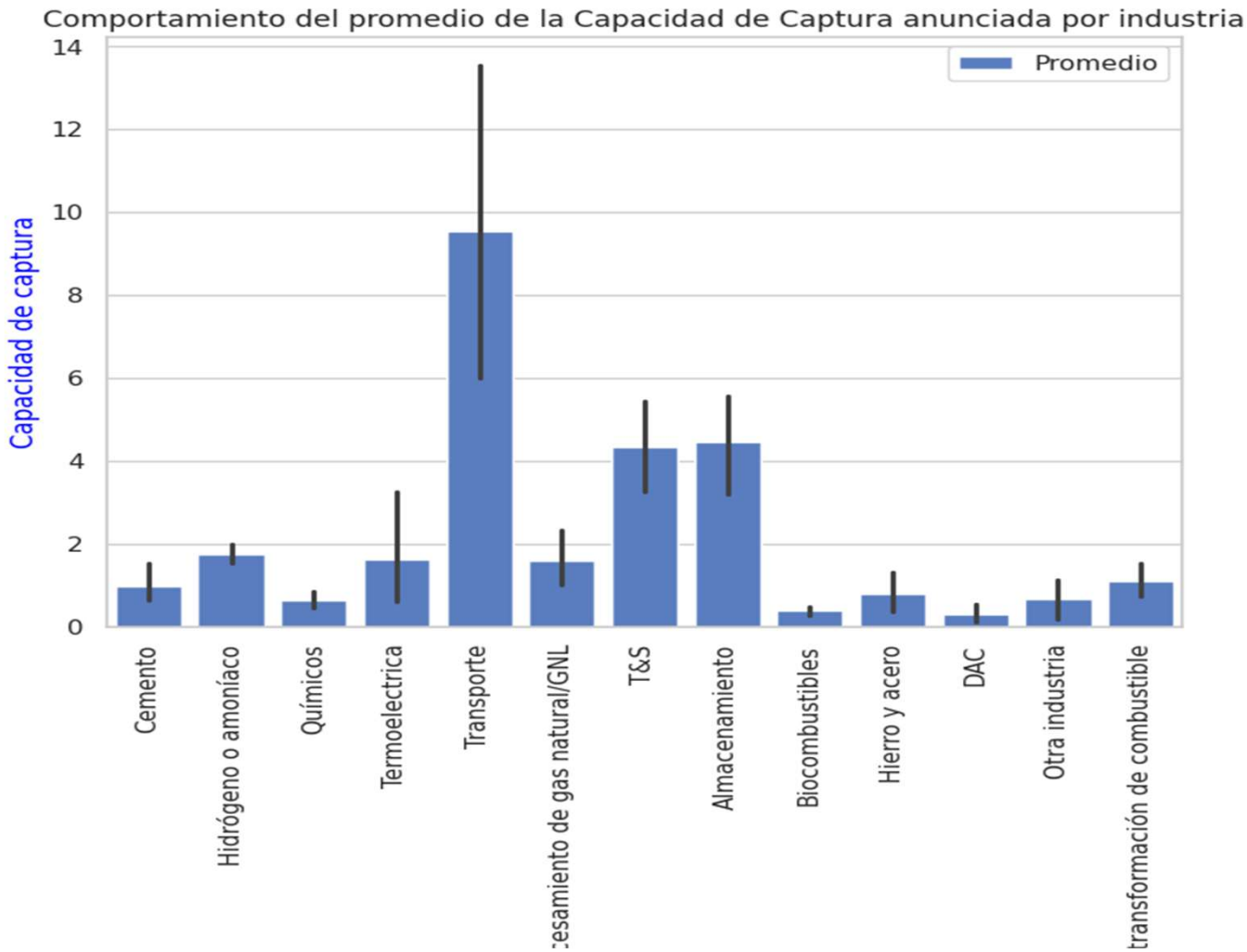
Data set 2


Comportamiento del promedio de la Capacidad de Captura anunciada por Pais
Promedio
Capacidad de captura
0
5
10
15
20
25
30
Alemania
Alemania-Bélgica
Alemania-Noruega
Arabia Saudita
Argelia
Australia
Australia-Japan
Austria
Brasil
Bulgaria
Bélgica
Bélgica, Alemania, Países Bajos, Suiza, EE. UU.
Bélgica-Noruega
Canadá
Chile
China
Corea del Sur
Croacia
Desconocido
Dinamarca
Emiratos Árabes Unidos
España
Estados Unidos
Finlandia
Francia
Francia-Italia
Grecia
Holanda-Alemania
Hungría
India
Indonesia
Irlanda
Islandia
Italia
Italia-Grecia
Japón
Japón-Malasia
Katar
Kenia
Letonia
Letonia-Lituania
Libia
Lituania
Malasia
Noruega
Nueva Zelanda
Omán
Papúa Nueva Guinea
Países Bajos
Polonia
Polonia, Letonia, Lituania
Portugal
Reino Unido
Rumania
Singapur
Suecia
Suiza
Taiwán
Uruguay
Pais

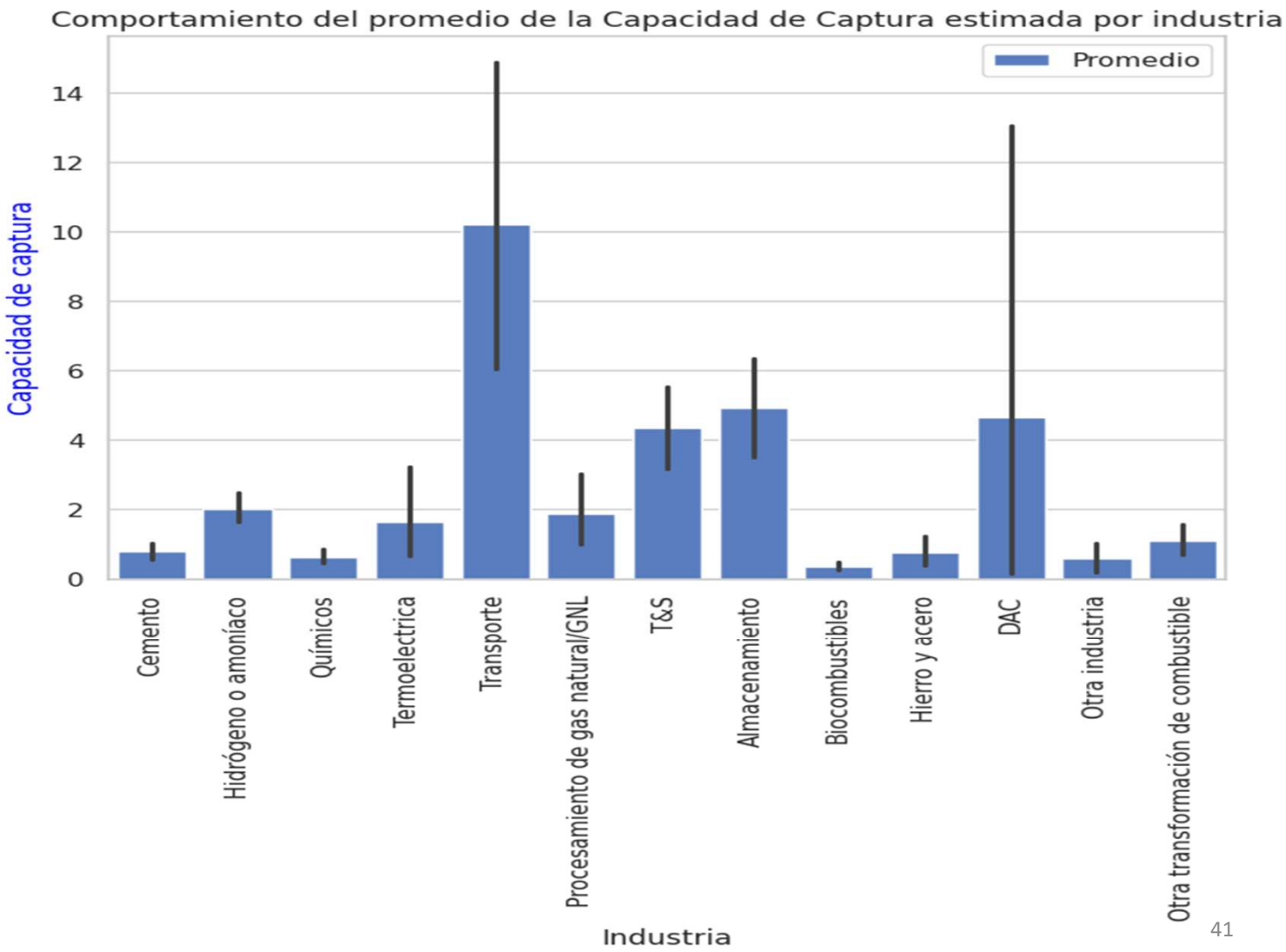
Data set 2


Comportamiento del promedio de la Capacidad de Captura estimada por Pais
Promedio
Capacidad de captura
0
5
10
15
20
25
30
35
40
Alemania
Alemania-Bélgica
Alemania-Noruega
Arabia Saudita
Argelia
Australia
Australia-Japan
Austria
Brasil
Bulgaria
Bélgica
Bélgica, Alemania, Países Bajos, Suiza, EE. UU.
Bélgica-Noruega
Canadá
Chile
China
Corea del Sur
Croacia
Desconocido
Dinamarca
Emiratos Árabes Unidos
España
Estados Unidos
Finlandia
Francia
Francia-Italia
Grecia
Holanda-Alemania
Hungría
India
Indonesia
Irlanda
Islandia
Italia
Italia-Grecia
Japón
Japón-Malasia
Katar
Kenia
Letonia
Letonia-Lituania
Libia
Lituania
Malasia
Noruega
Nueva Zelanda
Omán
Papúa Nueva Guinea
Países Bajos
Polonia
Polonia, Letonia, Lituania
Portugal
Reino Unido
Rumania
Singapur
Suecia
Suiza
Taiwán
Uruguay
Pais

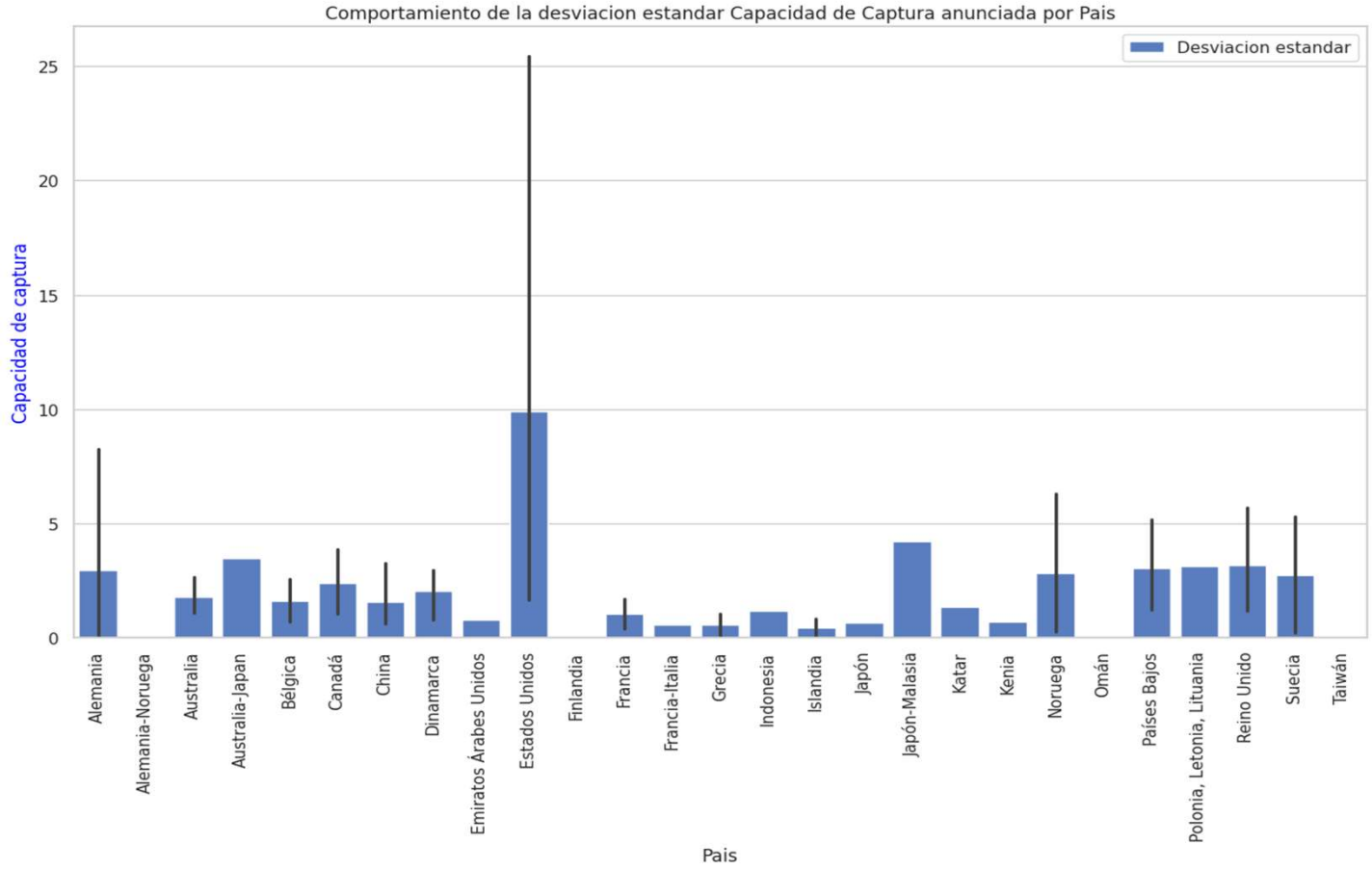
Data Set 2


Comportamiento del promedio de la Capacidad de Captura anunciada por industria
Promedio
Capacidad de captura
0
2
4
6
8
10
12
14
Cemento
Hidrógeno o amoníaco
Químicos
Termoelectrica
Transporte
esamiento de gas natural/GNL
T&S
Almacenamiento
Biocombustibles
Hierro y acero
DAC
Otra industria
transformación de combustible

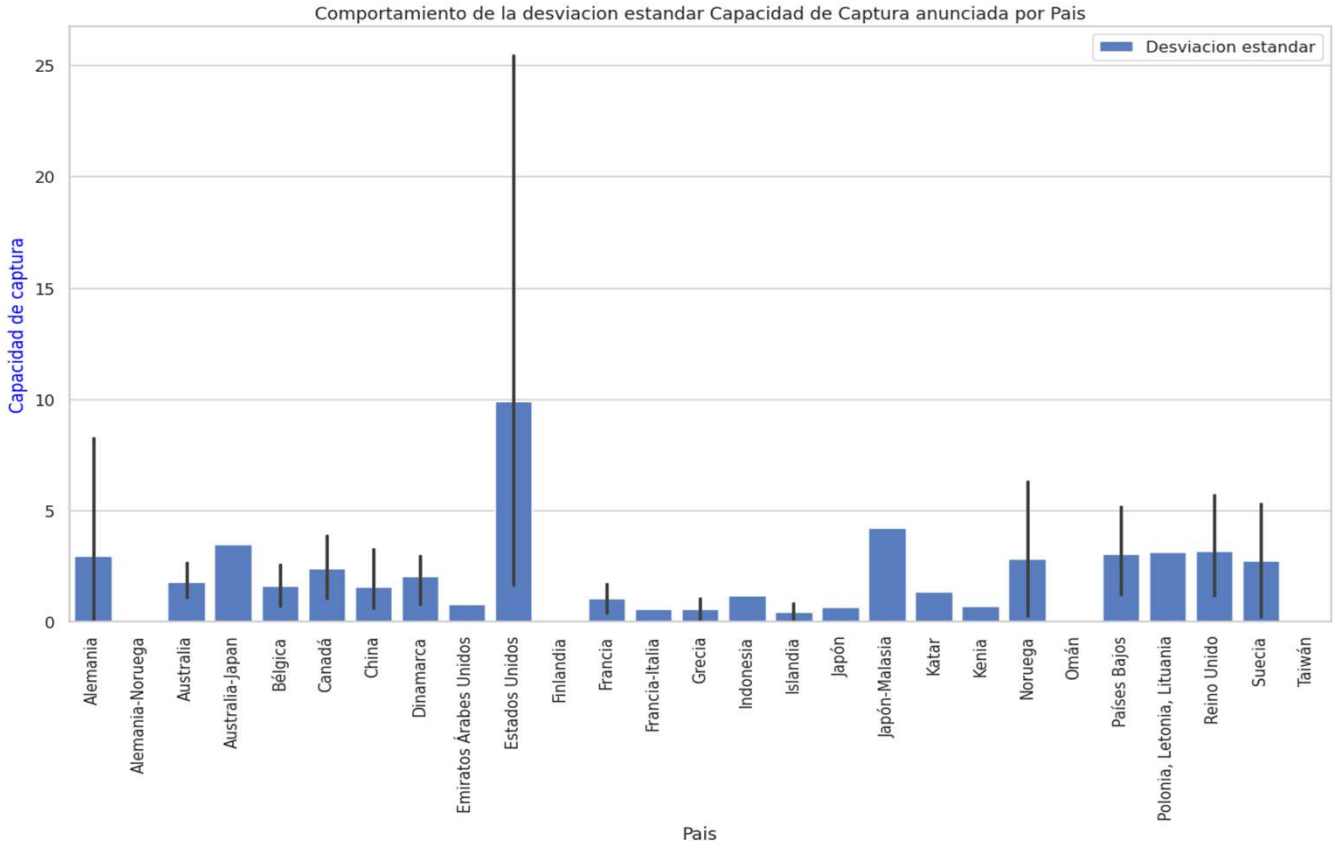
Data set 2


Comportamiento del promedio de la Capacidad de Captura estimada por industria
Promedio
Capacidad de captura
0
2
4
6
8
10
12
14
Cemento
Hidrógeno o amoníaco
Químicos
Termoelectrica
Transporte
Procesamiento de gas natural/GNL
T&S
Almacenamiento
Biocombustibles
Hierro y acero
DAC
Otra industria
Otra transformación de combustible
Industria

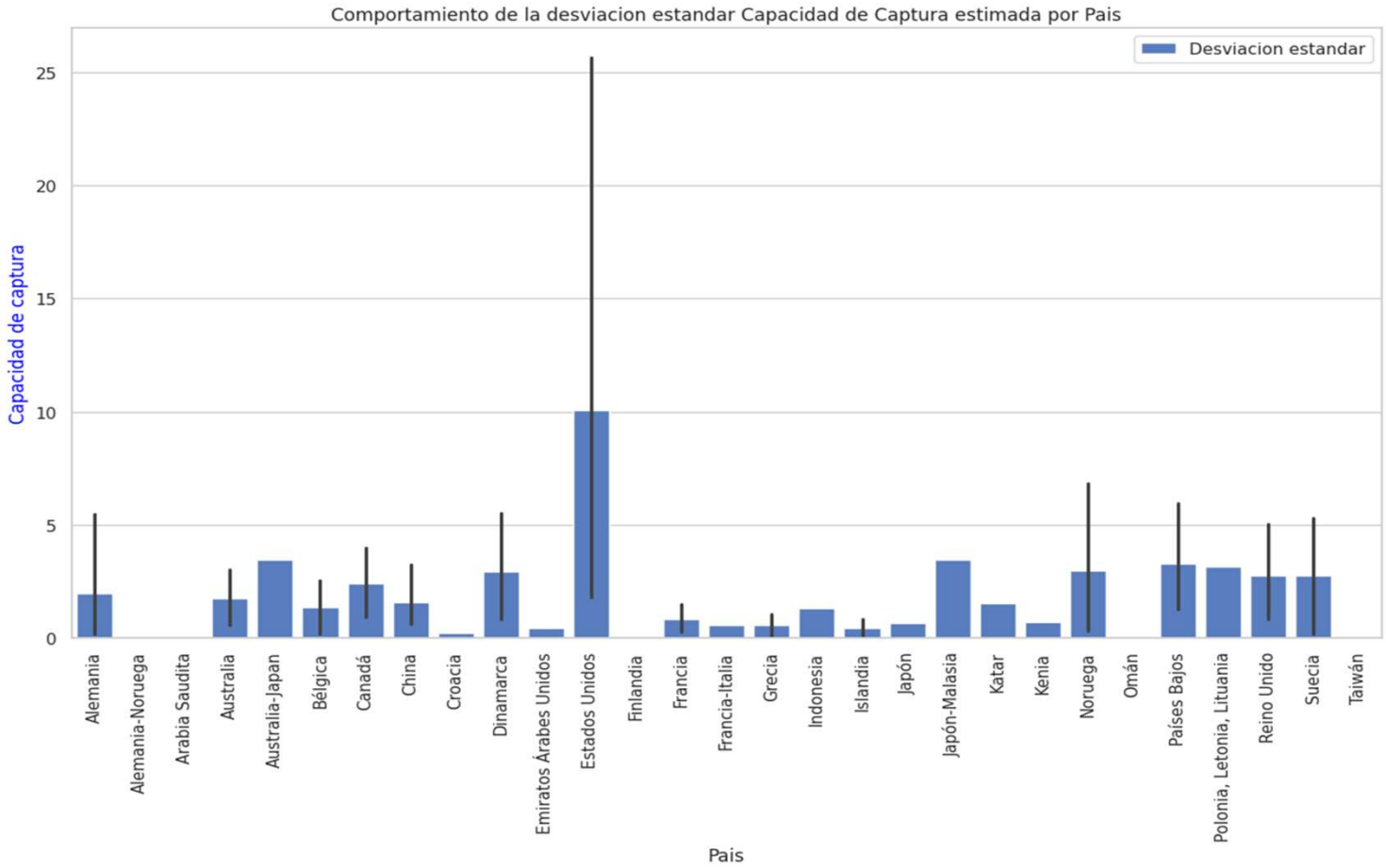
Data set 2


Comportamiento de la desviacion estandar Capacidad de Captura anunciada por Pais
Desviacion estandar
Capacidad de captura
0
5
10
15
20
25
Alemania
Alemania-Noruega
Australia
Australia-Japan
Bélgica
Canadá
China
Dinamarca
Emiratos Árabes Unidos
Estados Unidos
Finlandia
Francia
Francia-Italia
Grecia
Indonesia
Islandia
Japón
Japón-Malasia
Katar
Kenia
Noruega
Omán
Países Bajos
Polonia, Letonia, Lituania
Reino Unido
Suecia
Taiwán
Pais

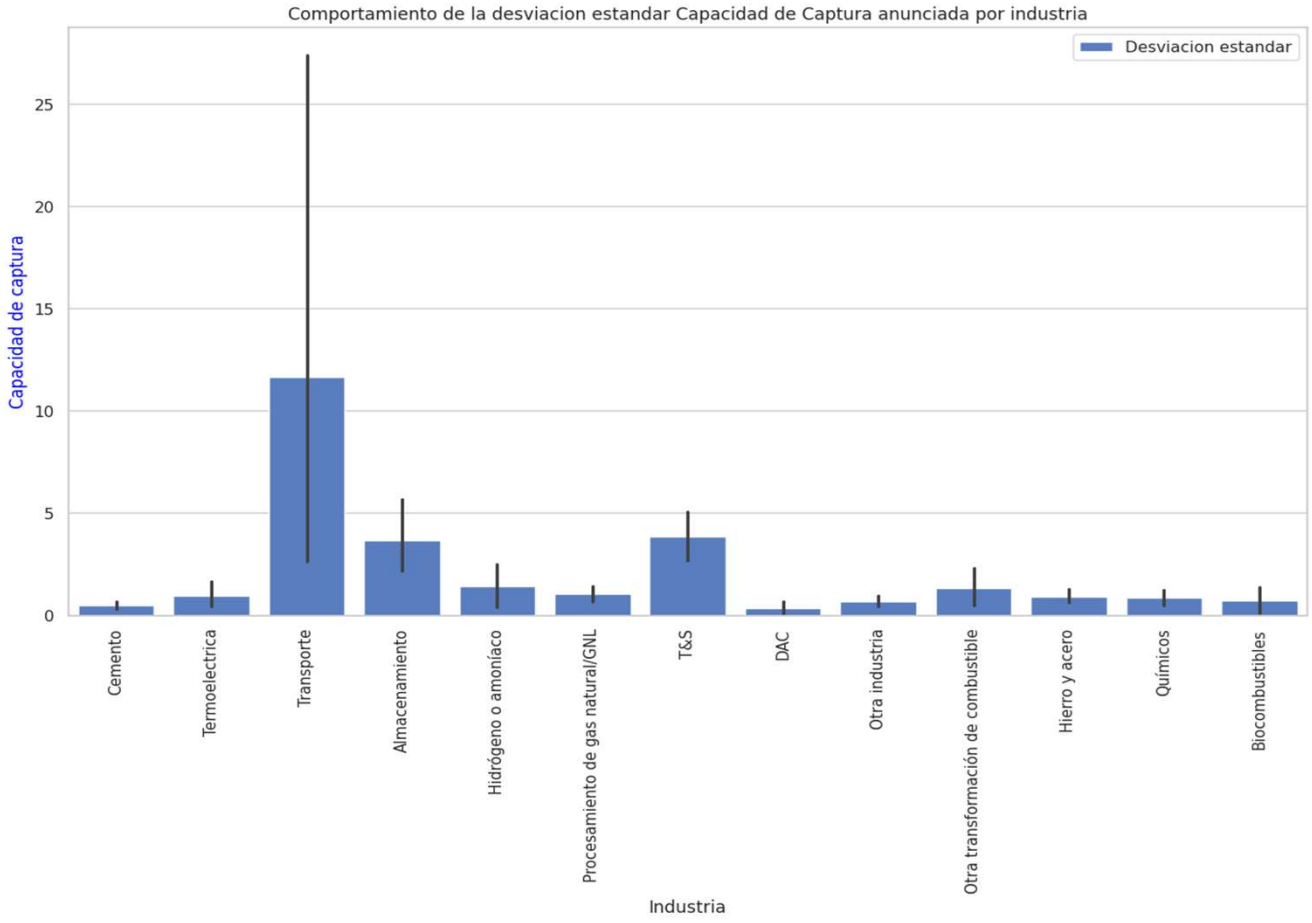
Data set 2


Comportamiento de la desviacion estandar Capacidad de Captura anunciada por Pais
Desviacion estandar
Capacidad de captura
0
5
10
15
20
25
Alemania
Alemania-Noruega
Australia
Australia-Japan
Bélgica
Canadá
China
Dinamarca
Emiratos Árabes Unidos
Estados Unidos
Finlandia
Francia
Francia-Italia
Grecia
Indonesia
Islandia
Japón
Japón-Malasia
Katar
Kenia
Noruega
Omán
Países Bajos
Polonia, Letonia, Lituania
Reino Unido
Suecia
Taiwán
Pais

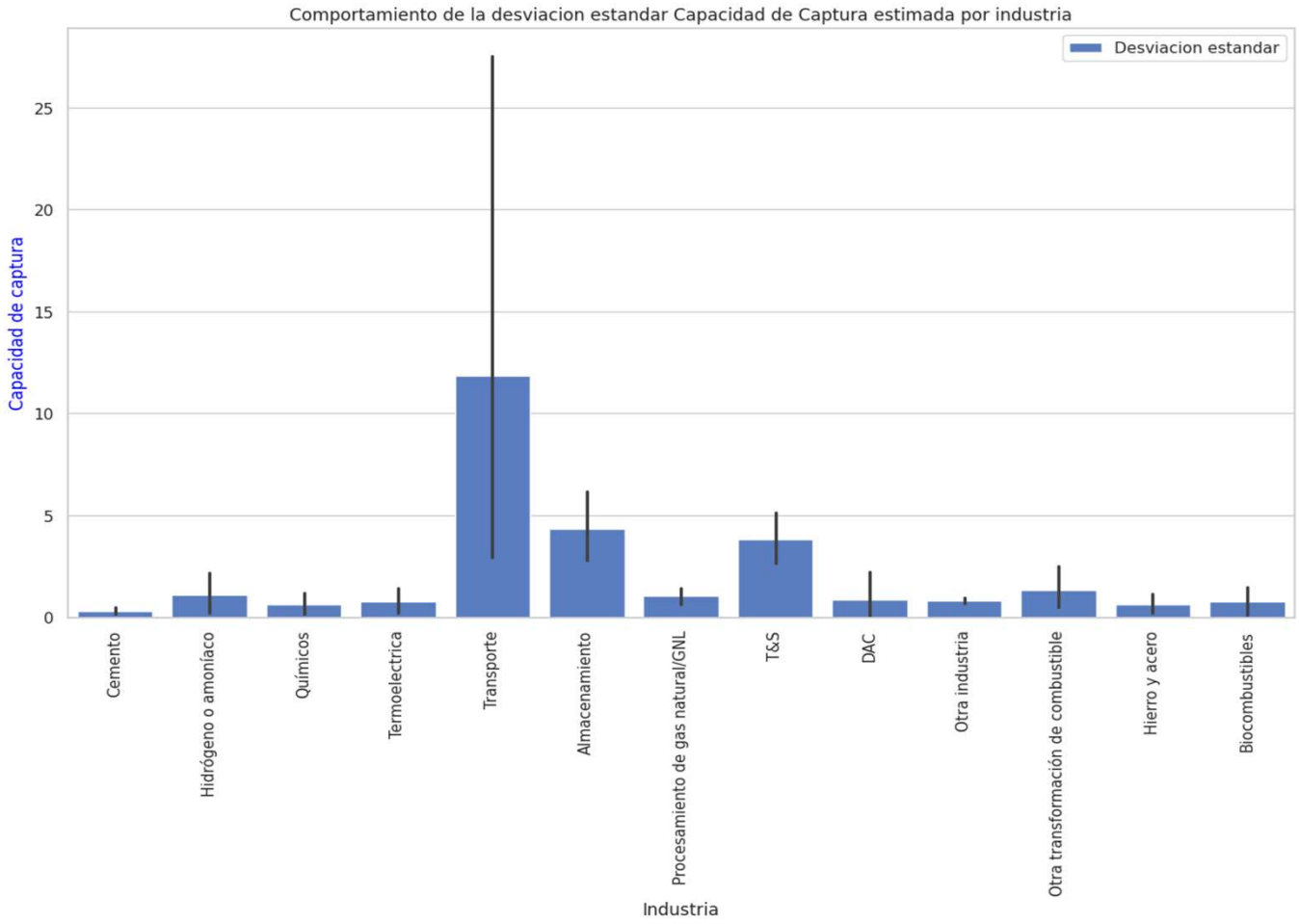
Data set 2


Comportamiento de la desviacion estandar Capacidad de Captura estimada por Pais
Desviacion estandar
Capacidad de captura
0
5
10
15
20
25
Alemania
Alemania-Noruega
Arabia Saudita
Australia
Australia-Japan
Bélgica
Canadá
China
Croacia
Dinamarca
Emiratos Árabes Unidos
Estados Unidos
Finlandia
Francia
Francia-Italia
Grecia
Indonesia
Islandia
Japón
Japón-Malasia
Katar
Kenia
Noruega
Omán
Países Bajos
Polonia, Letonia, Lituania
Reino Unido
Suecia
Taiwán
Pais

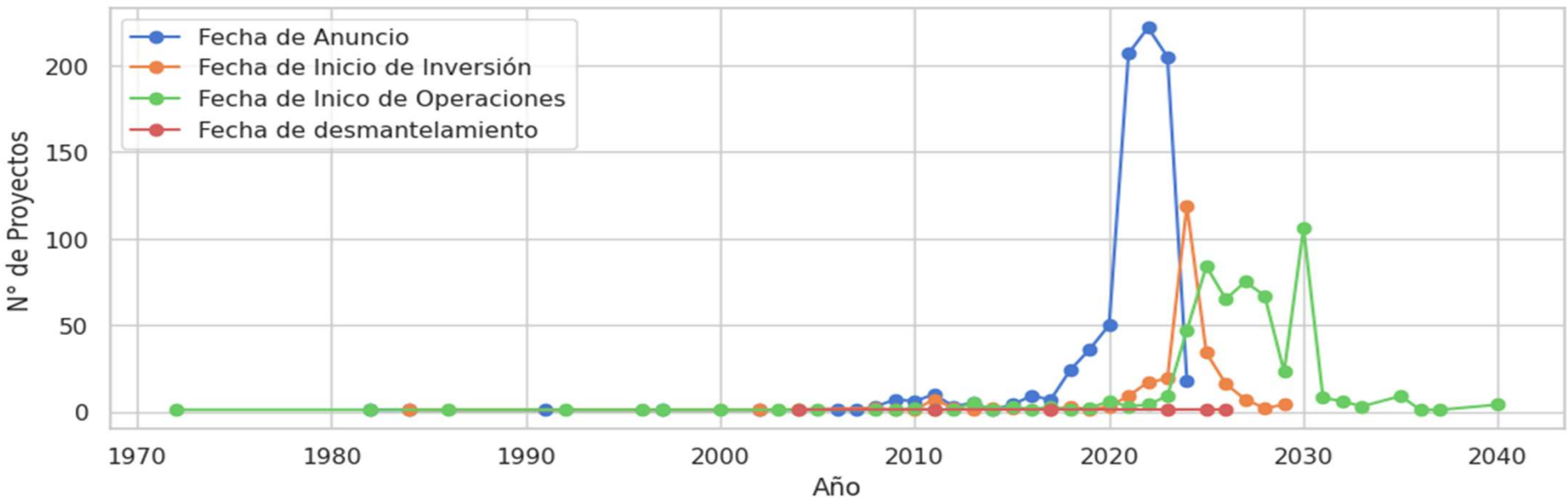
Data set 2 Diagrama de barras con barras de error



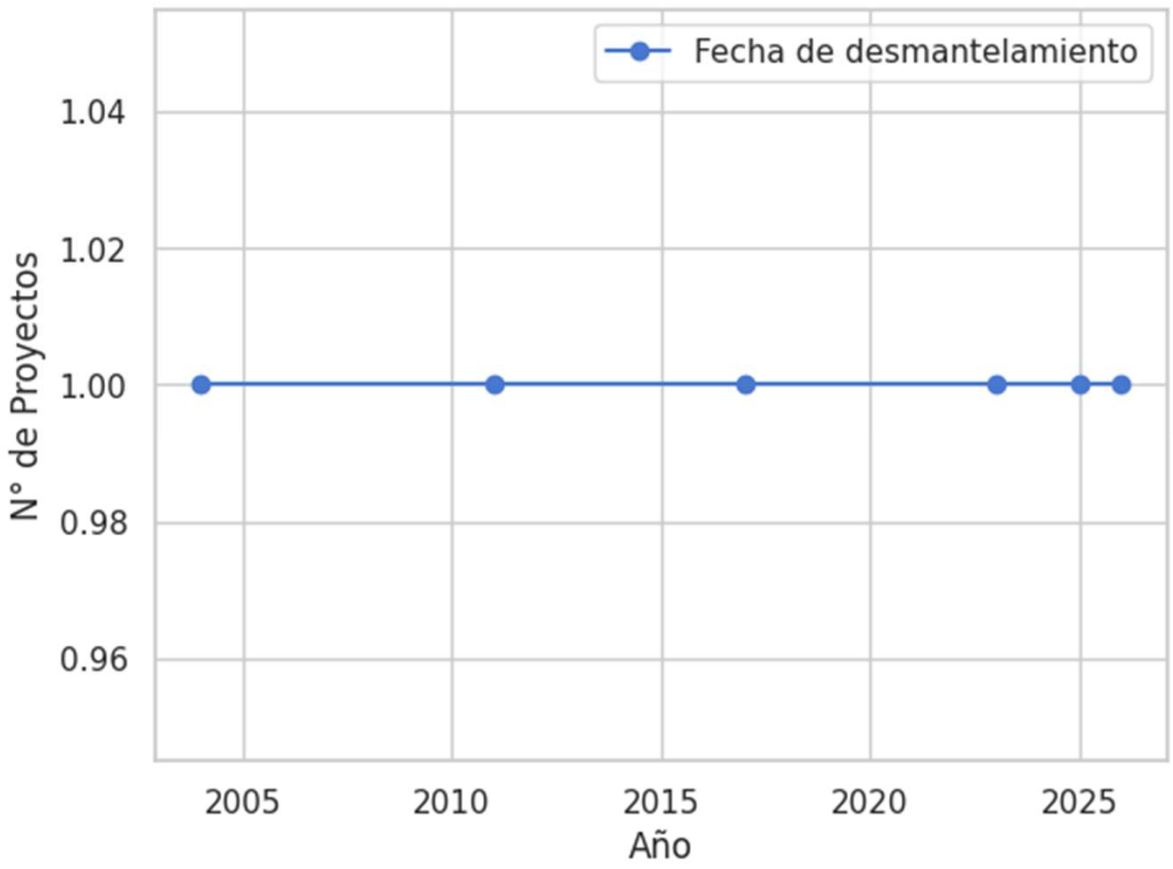
Data set 2 Diagrama de barras con barras de error



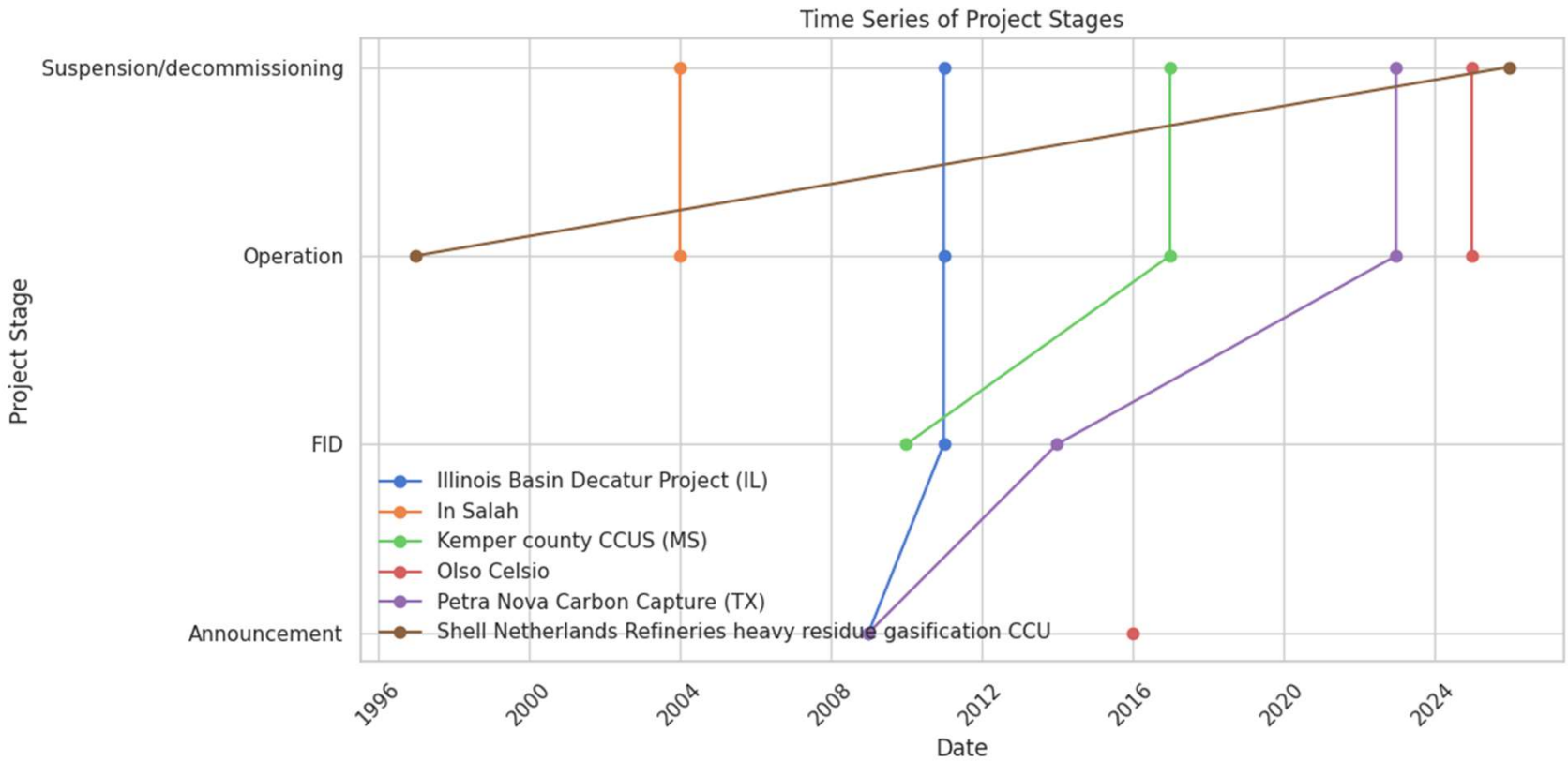
Data set 2 Serie de tiempo



Data set 2


Fecha de desmantelamiento
1.04
1.02
1.00
0.98
0.96
N° de Proyectos
2005
2010
2015
2020
2025
Año

Data set 2


Time Series of Project Stages
Suspension/decommissioning
Operation
FID
Announcement
Project Stage
Illinois Basin Decatur Project (IL)
In Salah
Kemper county CCUS (MS)
Olso Celsio
Petra Nova Carbon Capture (TX)
Shell Netherlands Refineries heavy residue gasification CCU
1996
2000
2004
2008
2012
2016
2020
2024
Date

Data set 2
Exploración para el modelamiento del PCA

Antes de la limpieza de valores atípicos

```
promedio de Capacidad anunciada, Varianza de capacidad anunciada; [np.float64(3.3897817589576547), 141.26476983483792]
promedio de Capacidad estimada, Varianza de Capacidad estimada:  [np.float64(3.4499627906976746), 138.46636432470035]
```

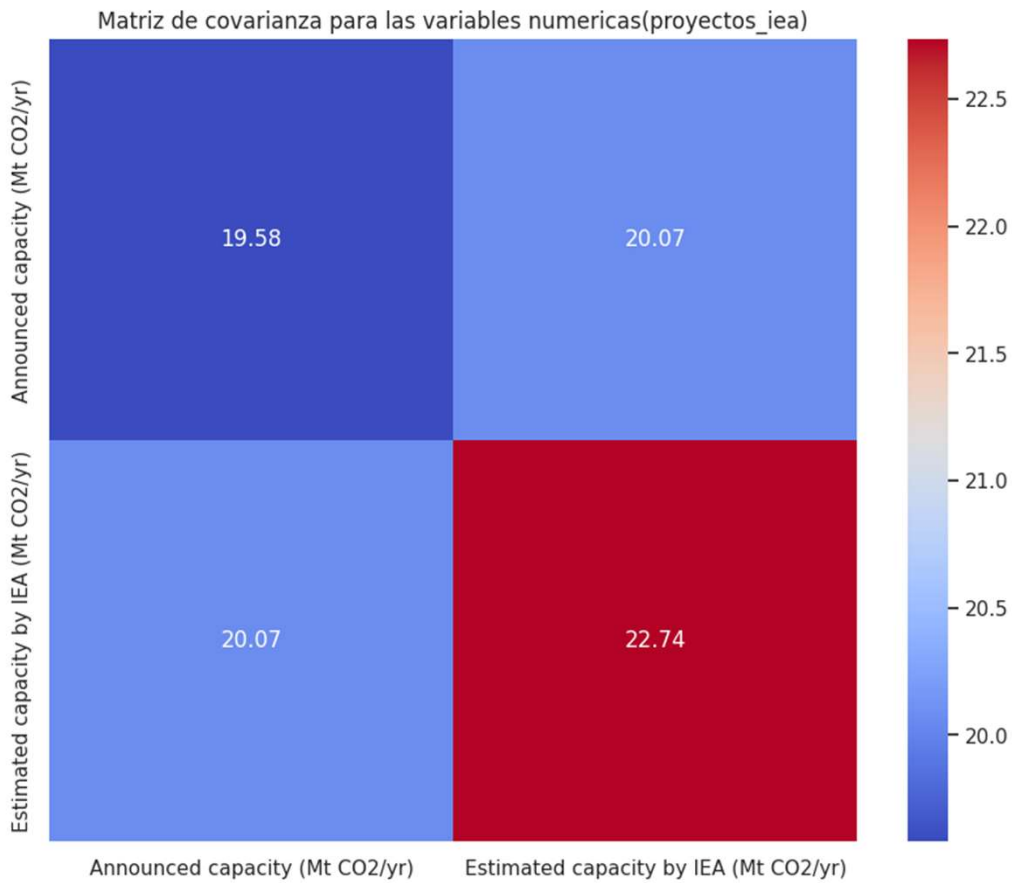
Luego de la limpieza de valores atípicos

```
promedio de Capacidad anunciada, Varianza de capacidad anunciada; [np.float64(2.79628431372549), 19.58296361461439]
promedio de Capacidad estimada, Varianza de Capacidad estimada:  [np.float64(2.885265940902022), 22.735927002373984]
```

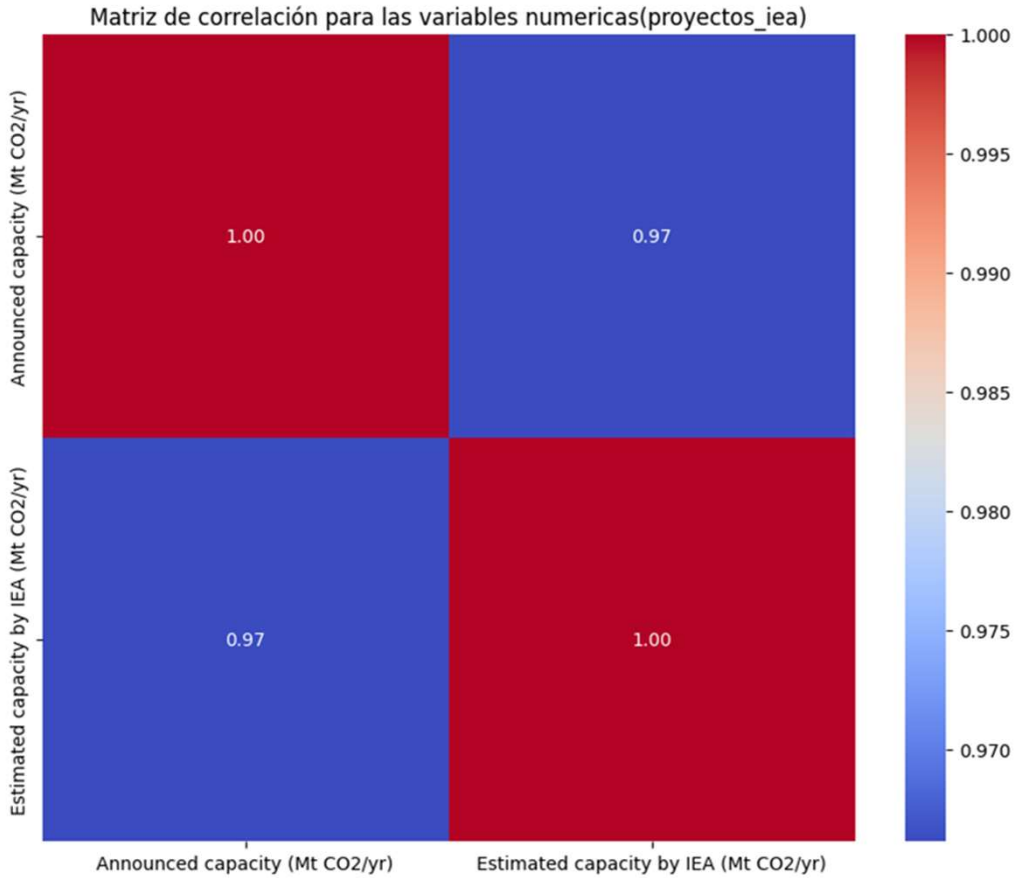
Data set 2
Matriz de Covarianza

	Announced capacity (Mt CO2/yr)	Estimated capacity by IEA (Mt CO2/yr)
Announced capacity (Mt CO2/yr)	19.582964	20.067968
Estimated capacity by IEA (Mt CO2/yr)	20.067968	22.735927

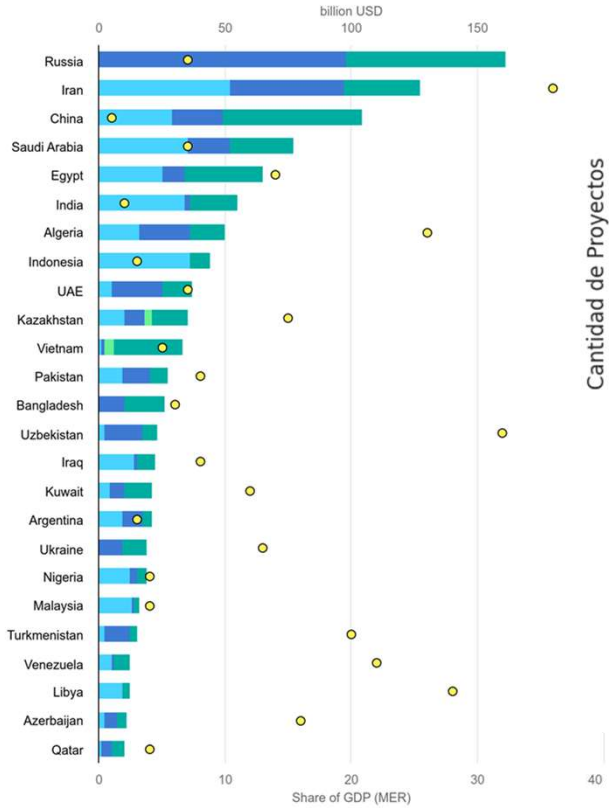
Data set 2
Matriz de Covarianza-Mapa de Calor

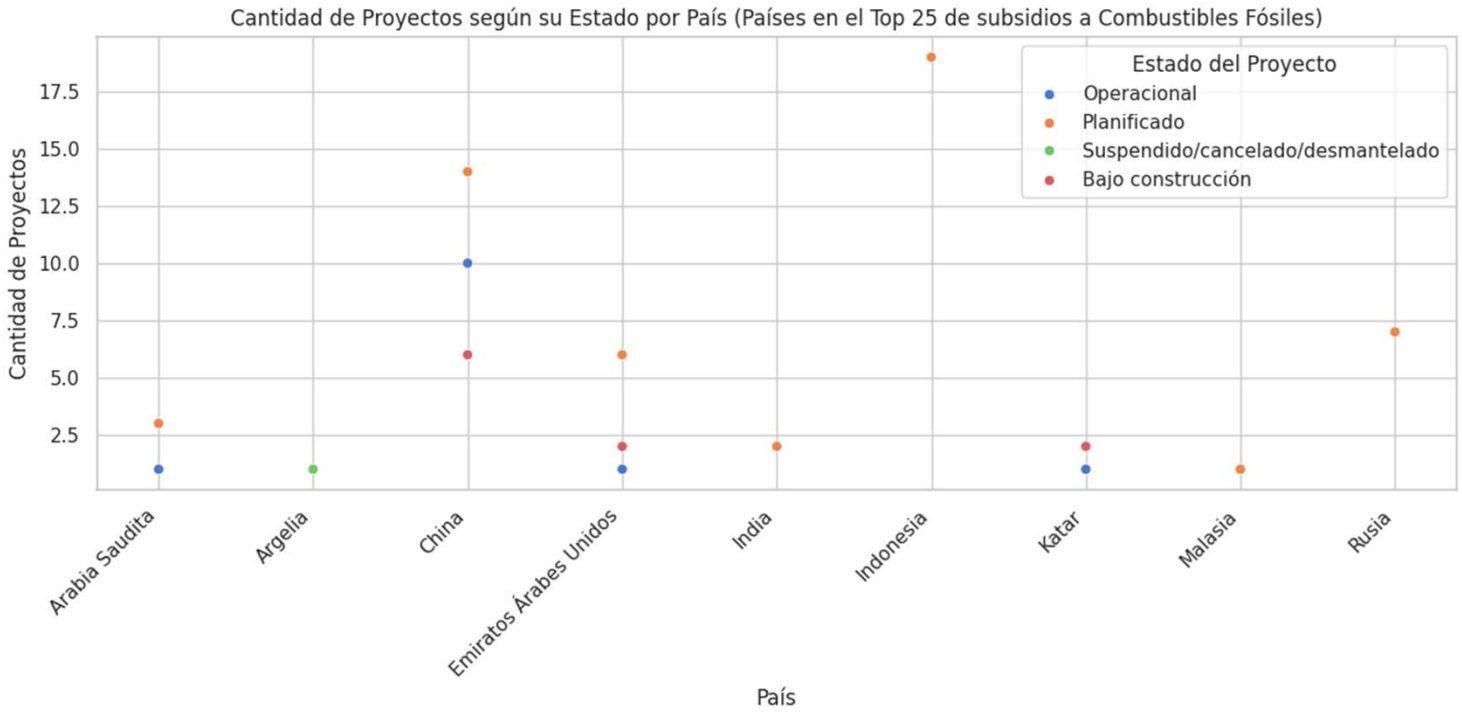


Data set 2
Matriz de Correlación-Mapa de Calor



Top 25 de Países con subsidios de Combustibles Fósiles





Value of fossil-fuel subsidies by fuel in the top 25 countries, 2022

Data set 3 : China

https://www.youtube.com/watch?v=LYSNER4Dnjs

Fuente del Data set 3

View PDF Download full issue

ELSEVIER

Journal of Cleaner Production
Volume 454, 15 May 2024, 142280

A preliminary assessment of CO_2 capture, transport, and storage network for China's steel sector

Yuancheng Lin [a b], Tubagus Aryandi Gunawan [c], Cecelia Isaac [c], Hongxi Luo [c], Fangwei Cheng [c], Eric D. Larson [c], Chris Greig [c], Linwei Ma [a b], Zheng Li [a b]

Show more

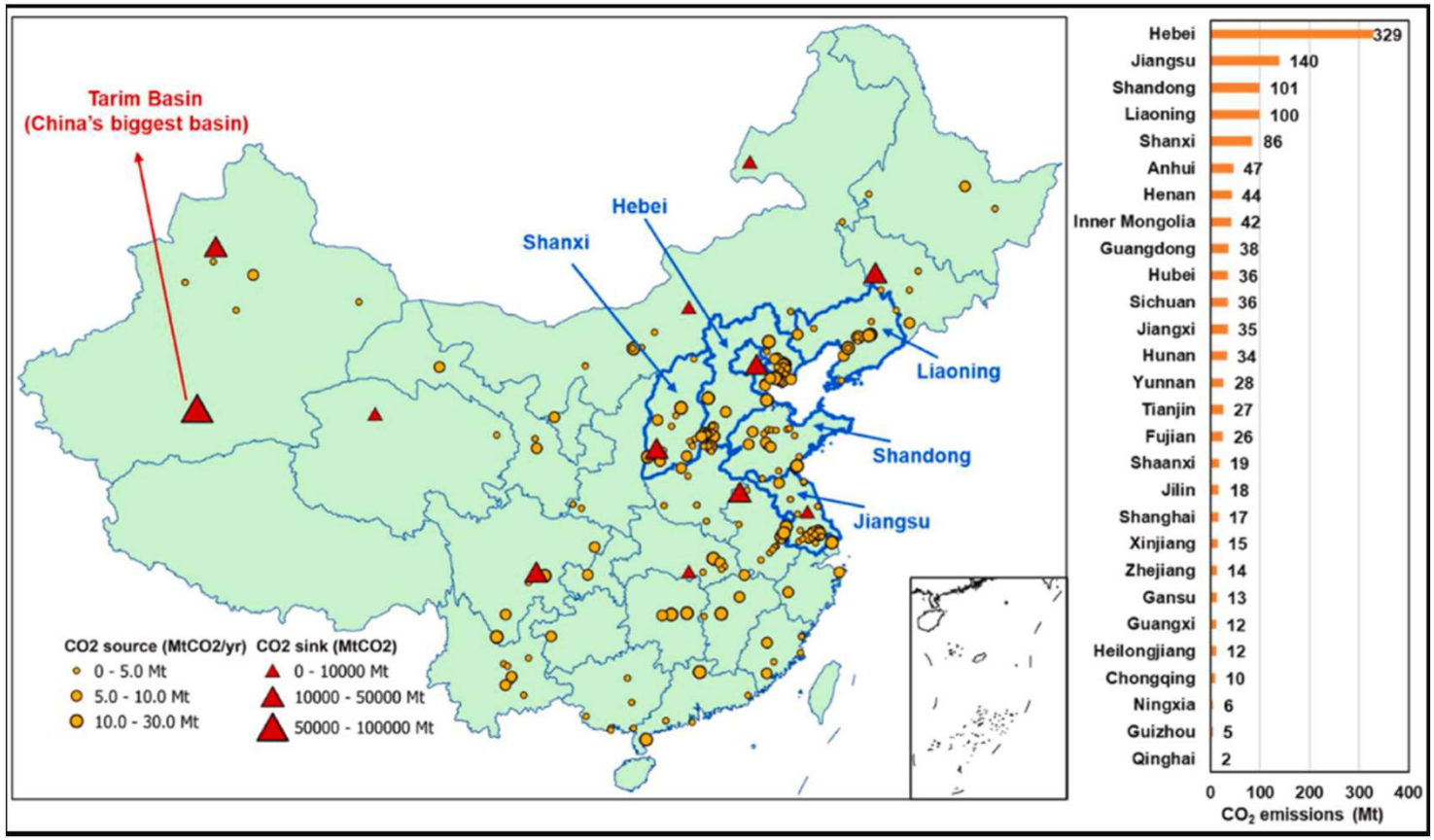
+ Add to Mendeley Share Cite

https://doi.org/10.1016/j.jclepro.2024.142280

Get rights and content

Full text access

China Data set 3



Data set 3

```
[4]  1 ## Descripcion de las variables numéricas
     2 df_China.describe()
```

	Captured CO2/Mt	Captured cost/($/t)
count	227.000000	227.000000
mean	5.690749	54.897357
std	4.946819	10.671026
min	0.200000	34.700000
25%	2.450000	46.800000
50%	4.000000	54.200000
75%	7.700000	60.750000
max	28.300000	104.200000

Data set 3

```
# Libreria Skimpy permite realizar estadistiscos descriptivos de variables categóricas
skim(df_China)
```

skimpy summary

Data Summary

Dataframe	Values
Number of rows	227
Number of columns	3

Data Types

Column Type	Count
float64	2
string	1

number

column	NA	NA %	mean	sd	p0	p25	p50	p75	p100	hist
Captured CO2/Mt	0	0	5.691	4.947	0.2	2.45	4	7.7	28.3	
Captured cost/($/t)	0	0	54.9	10.67	34.7	46.8	54.2	60.75	104.2	

string

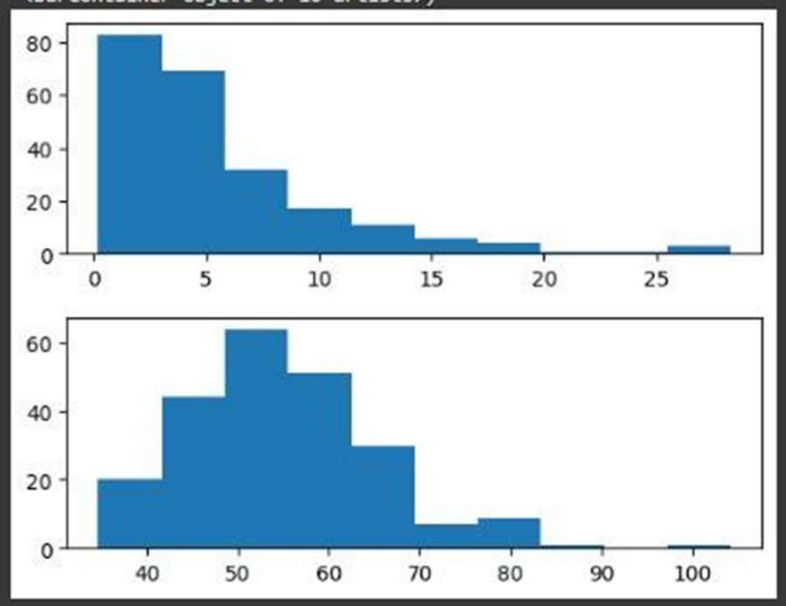
column	NA	NA %	shortest	longest	min	max	chars per row	words per row	total words
Province	0	0	Anhui	Inner Mongolia	Anhui	Zhejiang	6.67	1	236

End

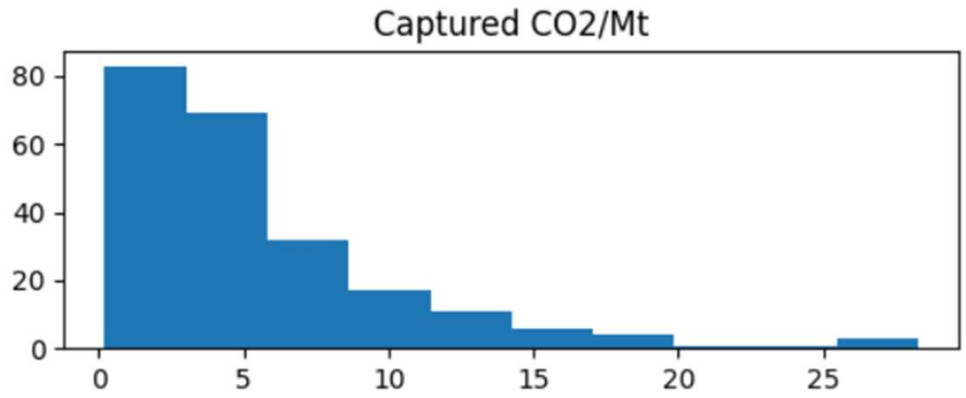
Data set 3

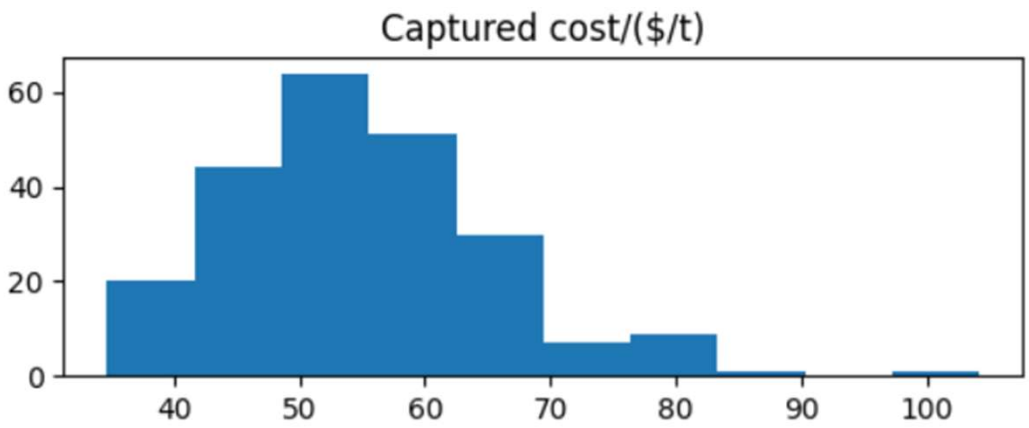
```
[34]
# Analisis estadistico - Medidas de Tendencia Central
plt.figure(figsize=(6,2))            ## Configuración del tamaño del grafico
plt.hist(df_China["Captured CO2/Mt"])  ## Histograma de Captured CO2/Mt
plt.figure(figsize=(6,2))            ## Configuración del tamaño del grafico
plt.hist(df_China["Captured cost/($/t)"])  ## Histograma de Captured CO2/Mt
```

```
(array([20., 44., 64., 51., 30.,  7.,  9.,  1.,  0.,  1.]),
 array([ 34.7 ,  41.65,  48.6 ,  55.55,  62.5 ,  69.45,  76.4 ,  83.35,
        90.3 ,  97.25, 104.2 ]),
 <BarContainer object of 10 artists>)
```

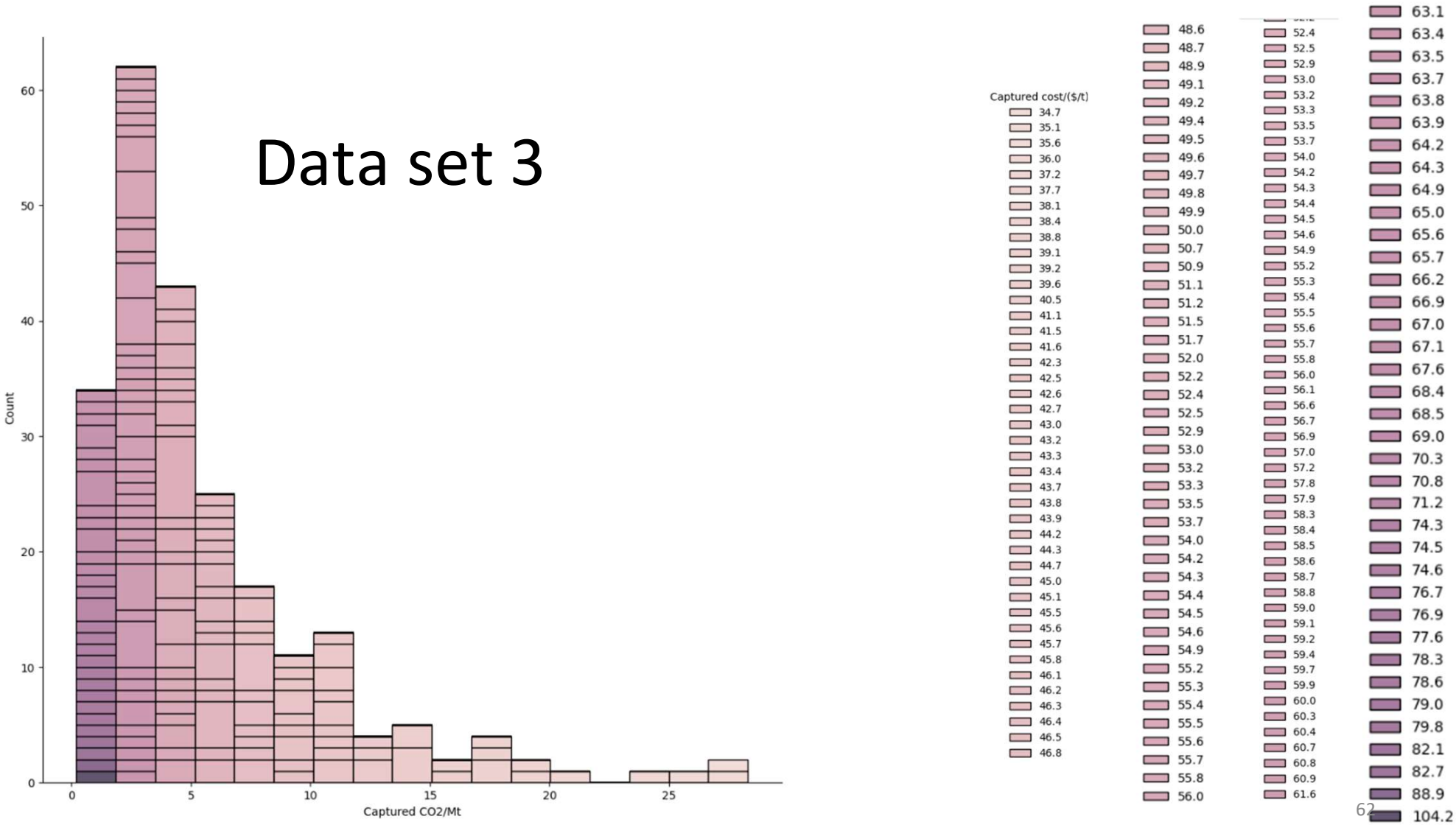


Data set 3


Captured CO2/Mt
80
60
40
20
0
0
5
10
15
20
25

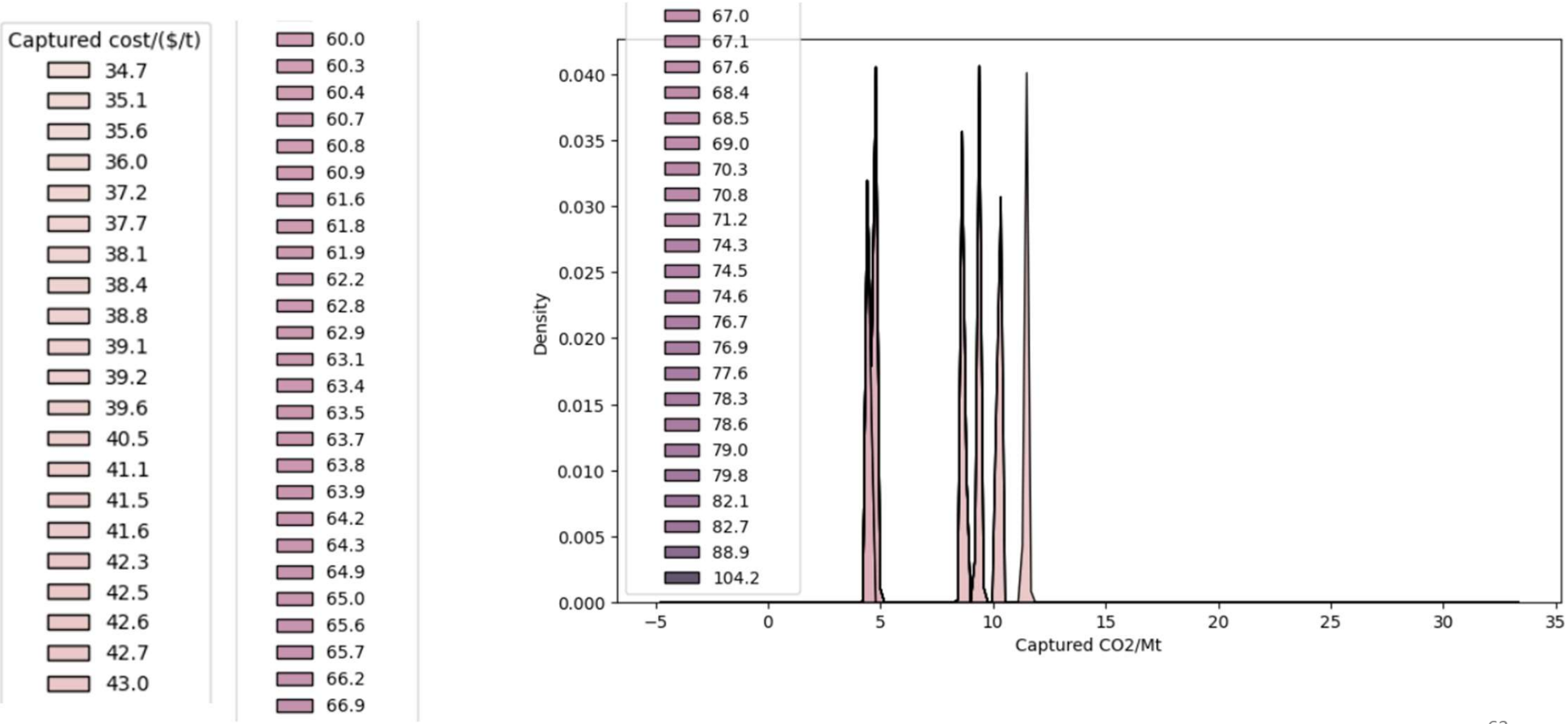

Captured cost/($/t)
60
40
20
0
40
50
60
70
80
90
100

Data set 3

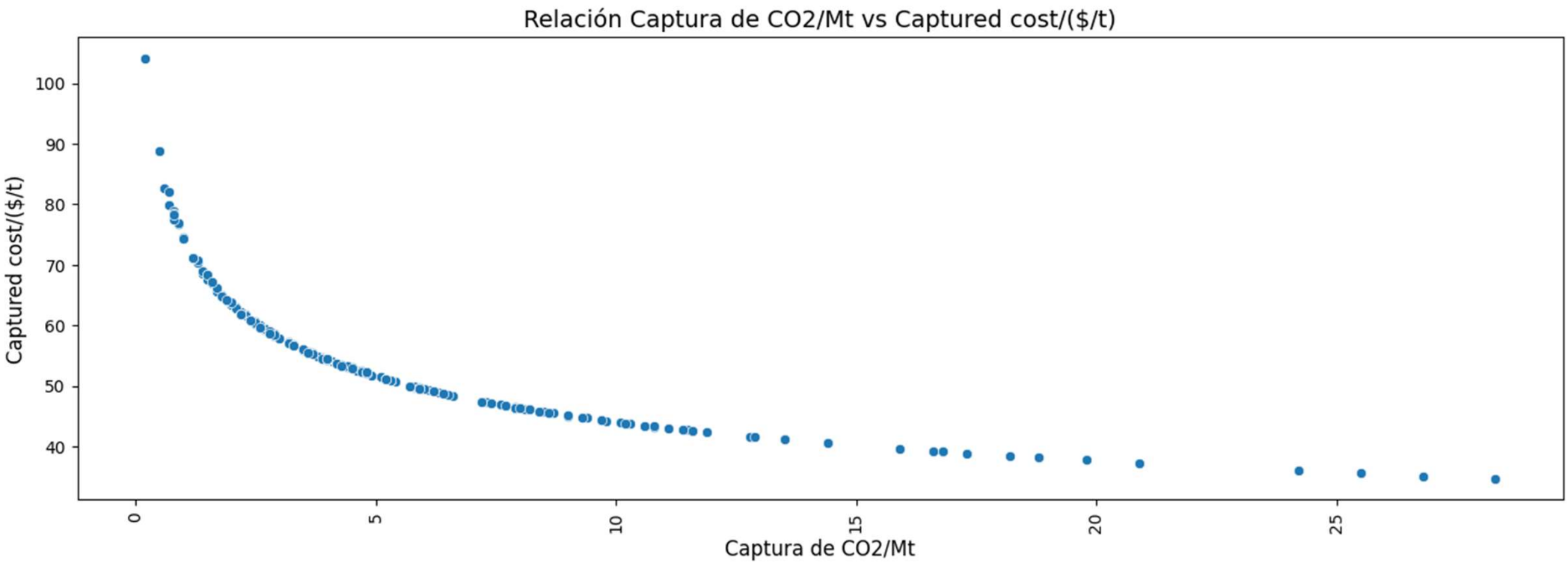

Count
Captured CO2/Mt
0
10
20
30
40
50
60
0
5
10
15
20
25
Captured cost/($/t)
34.7
35.1
35.6
36.0
37.2
37.7
38.1
38.4
38.8
39.1
39.2
39.6
40.5
41.1
41.5
41.6
42.3
42.5
42.6
42.7
43.0
43.2
43.3
43.4
43.7
43.8
43.9
44.2
44.3
44.7
45.0
45.1
45.5
45.6
45.7
45.8
46.1
46.2
46.3
46.4
46.5
46.8
48.6
48.7
48.9
49.1
49.2
49.4
49.5
49.6
49.7
49.8
49.9
50.0
50.7
50.9
51.1
51.2
51.5
51.7
52.0
52.2
52.4
52.5
52.9
53.0
53.2
53.3
53.5
53.7
54.0
54.2
54.3
54.4
54.5
54.6
54.9
55.2
55.3
55.4
55.5
55.6
55.7
55.8
56.0
52.4
52.5
52.9
53.0
53.2
53.3
53.5
53.7
54.0
54.2
54.3
54.4
54.5
54.6
54.9
55.2
55.3
55.4
55.5
55.6
55.7
55.8
56.0
56.1
56.6
56.7
56.9
57.0
57.2
57.8
57.9
58.3
58.4
58.5
58.6
58.7
58.8
59.0
59.1
59.2
59.4
59.7
59.9
60.0
60.3
60.4
60.7
60.8
60.9
61.6
63.1
63.4
63.5
63.7
63.8
63.9
64.2
64.3
64.9
65.0
65.6
65.7
66.2
66.9
67.0
67.1
67.6
68.4
68.5
69.0
70.3
70.8
71.2
74.3
74.5
74.6
76.7
76.9
77.6
78.3
78.6
79.0
79.8
82.1
82.7
88.9
104.2

Data set 3 KDE

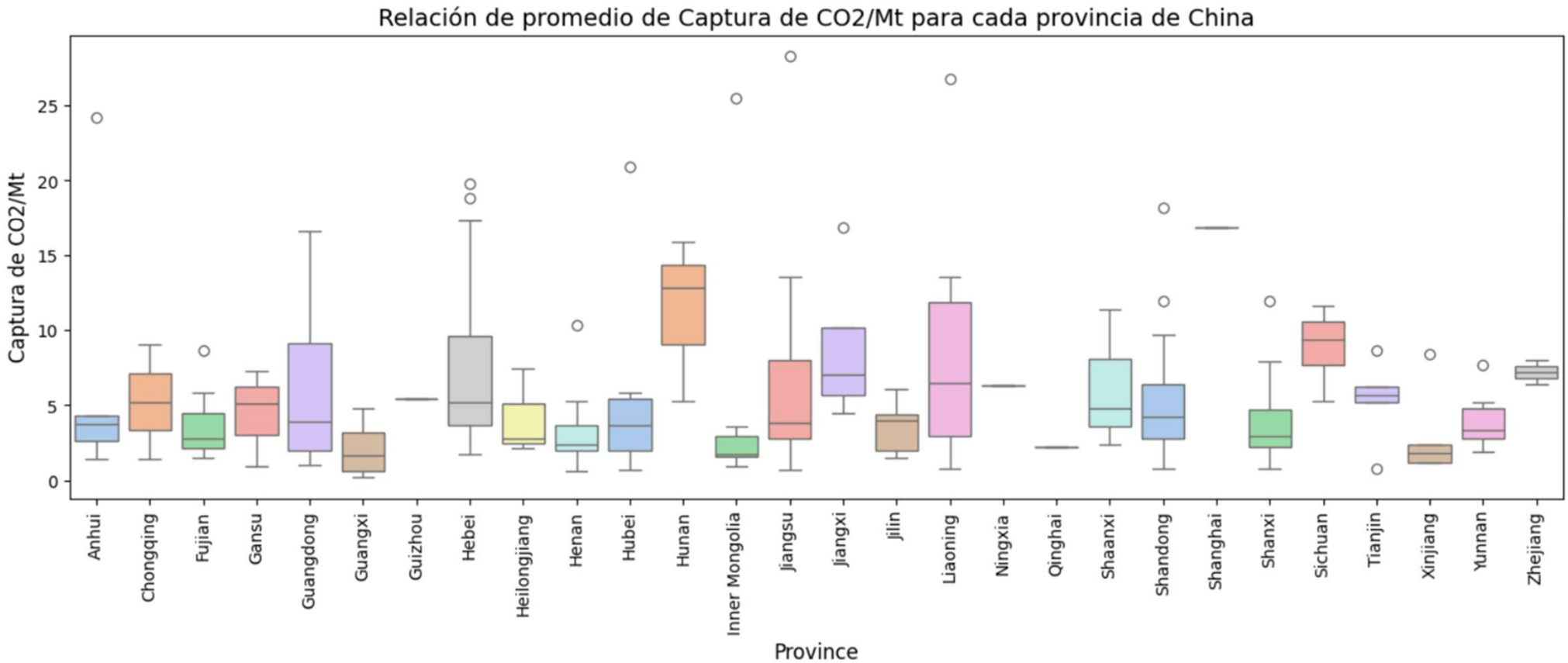
kernel density estimate (KDE)

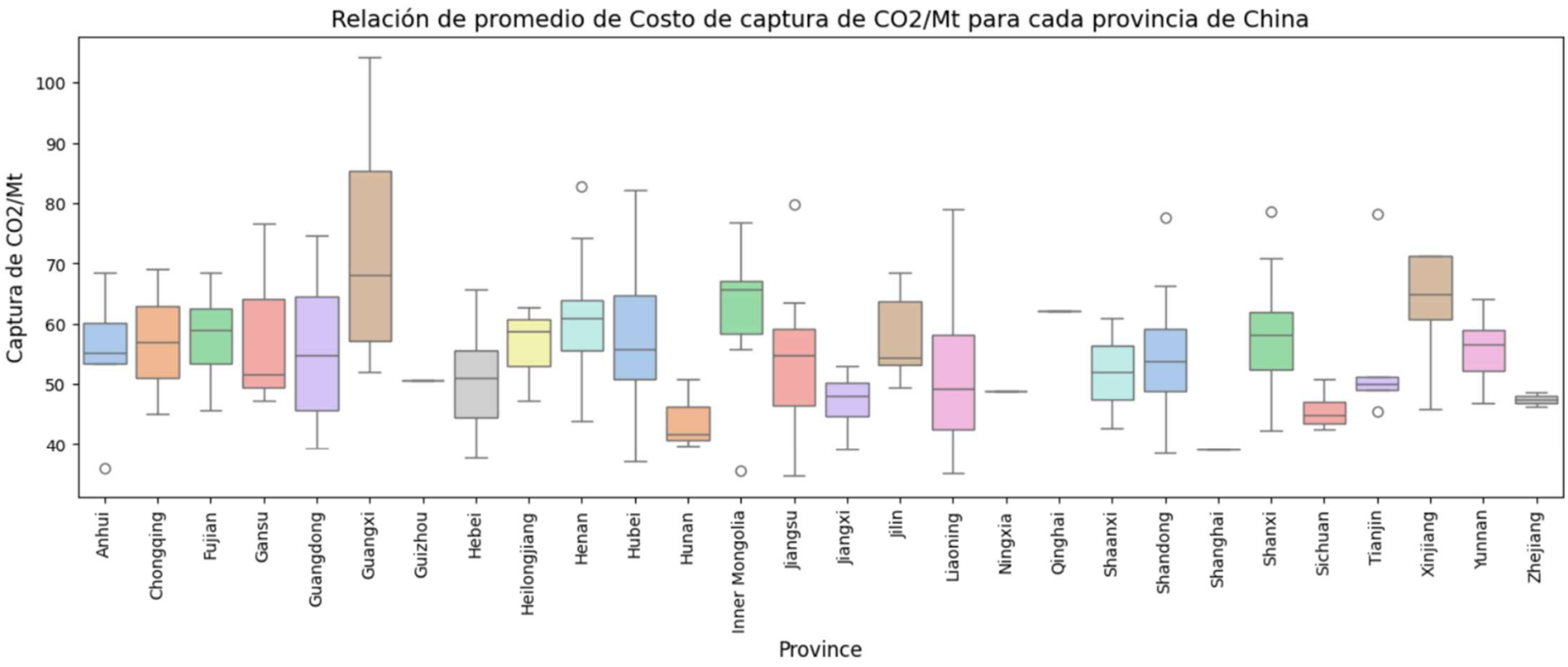


Data set 3 *Scatterplot*

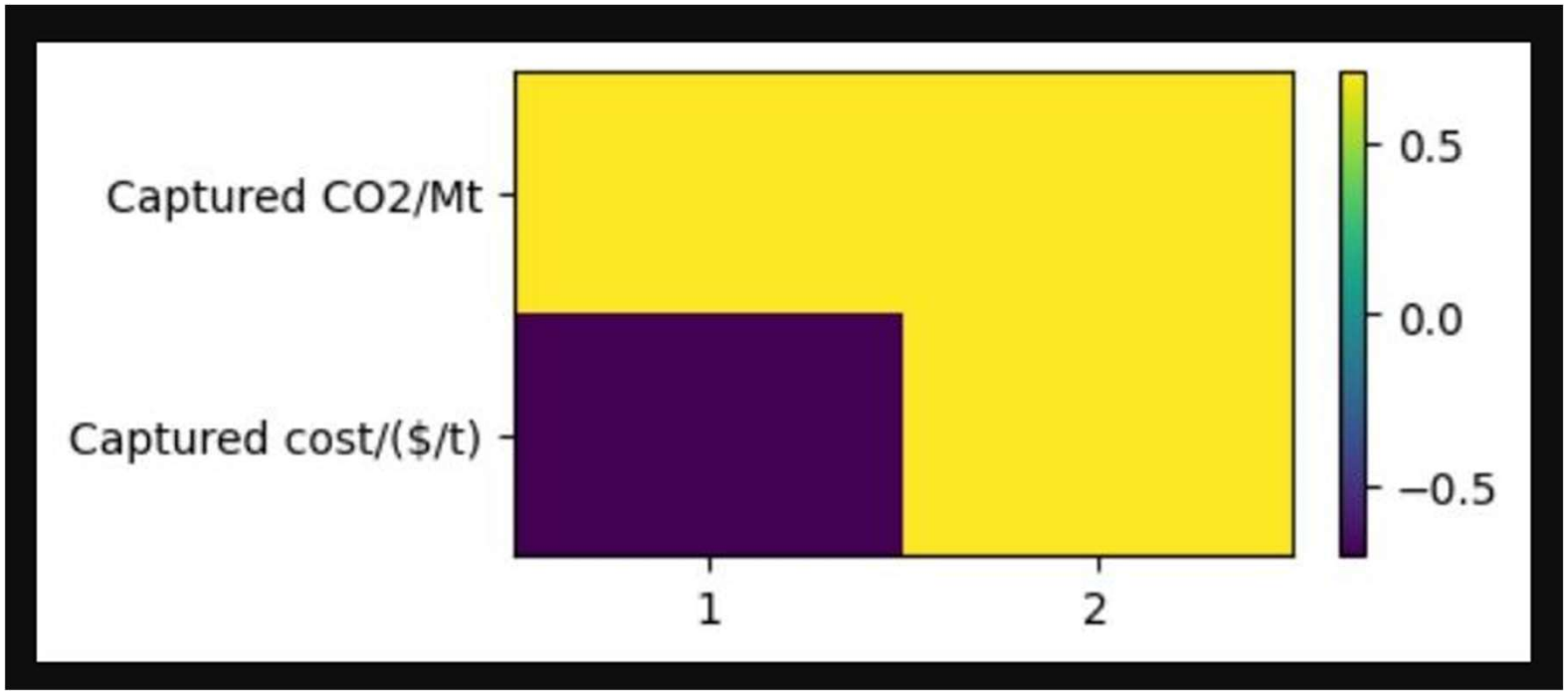


Data set 3

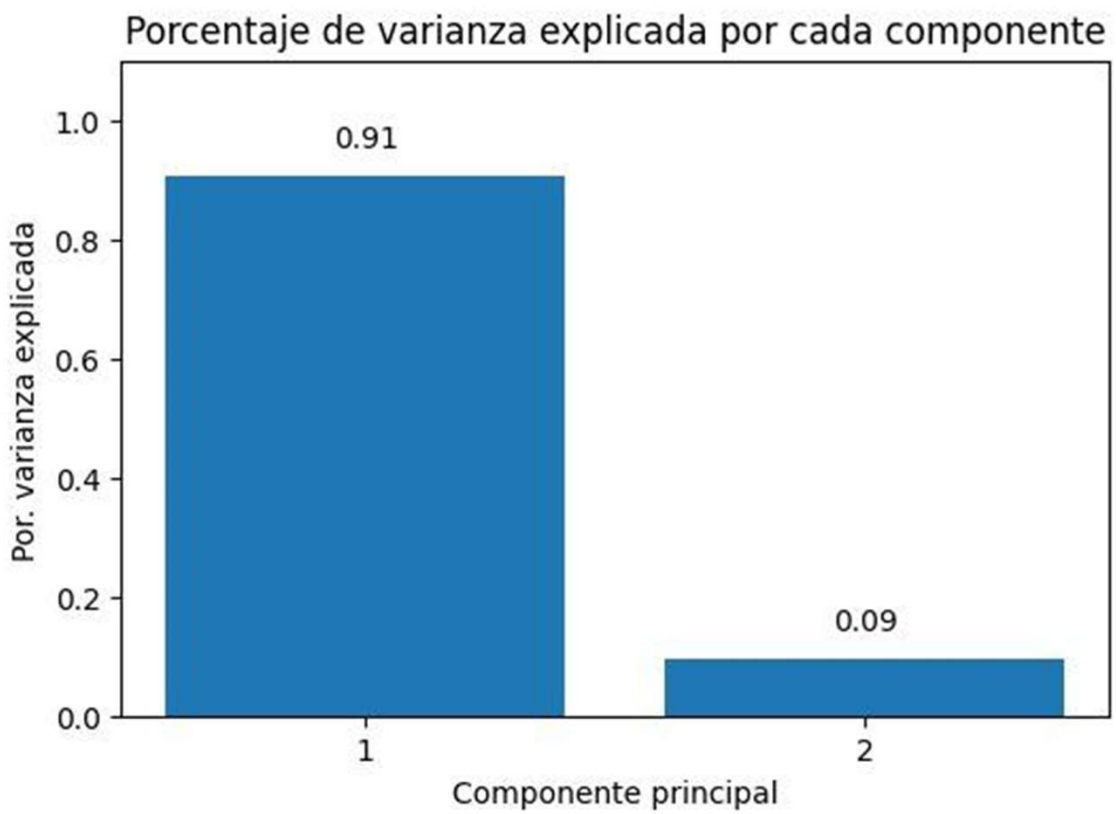

Relación de promedio de Captura de CO2/Mt para cada provincia de China
Captura de CO2/Mt
0
5
10
15
20
25
Anhui
Chongqing
Fujian
Gansu
Guangdong
Guangxi
Guizhou
Hebei
Heilongjiang
Henan
Hubei
Hunan
Inner Mongolia
Jiangsu
Jiangxi
Jilin
Liaoning
Ningxia
Qinghai
Shaanxi
Shandong
Shanghai
Shanxi
Sichuan
Tianjin
Xinjiang
Yunnan
Zhejiang
Province


Relación de promedio de Costo de captura de CO2/Mt para cada provincia de China
Captura de CO2/Mt
100
90
80
70
60
50
40
Anhui
Chongqing
Fujian
Gansu
Guangdong
Guangxi
Guizhou
Hebei
Heilongjiang
Henan
Hubei
Hunan
Inner Mongolia
Jiangsu
Jiangxi
Jilin
Liaoning
Ningxia
Qinghai
Shaanxi
Shandong
Shanghai
Shanxi
Sichuan
Tianjin
Xinjiang
Yunnan
Zhejiang
Province

Data set 3 Modelado PCA - Mapa de Calor



Data set 3 Modelado PCA



Data set 3

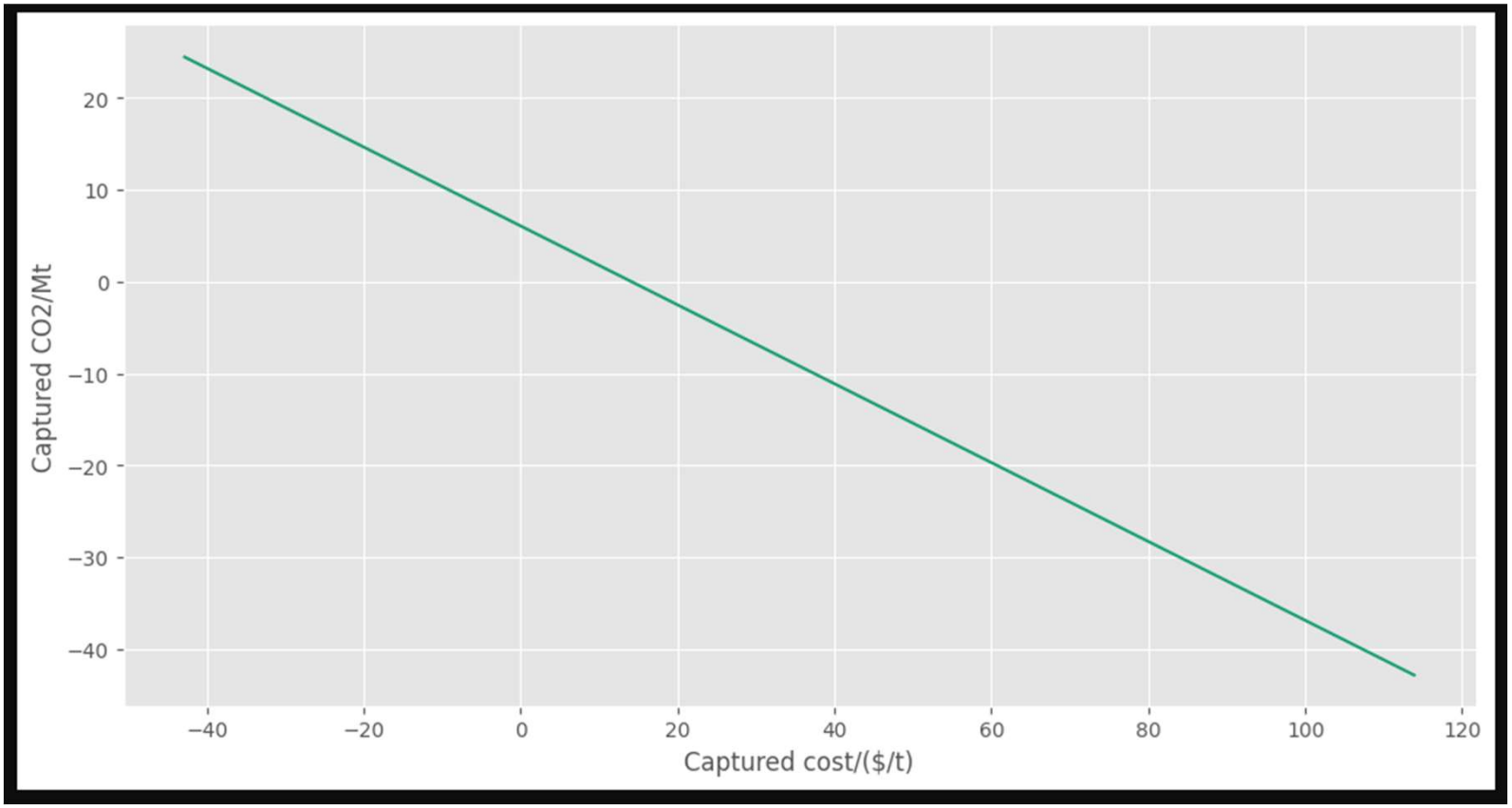
	variable_1	variable_2	r	abs_r
1	Captured CO2/Mt	Captured cost/($/t)	-0.811049	0.811049
2	Captured cost/($/t)	Captured CO2/Mt	-0.811049	0.811049

Matriz de Covarianza

```python
#Matriz de Covarianza
# Selecting only numerical columns for covariance calculation
numerical_df = df_China.select_dtypes(include=['number'])
numerical_df.cov(min_periods=None, ddof=1)
```

	Captured CO2/Mt	Captured cost/($/t)
Captured CO2/Mt	24.471020	-42.813343
Captured cost/($/t)	-42.813343	113.870789

Matriz de Covarianza



```python
# Correlación entre columnas numéricas
def tidy_corr_matrix(corr_mat):
    '''
    Función para convertir una matriz de correlación de pandas en formato tidy
    '''
    corr_mat = corr_mat.stack().reset_index()
    corr_mat.columns = ['variable_1','variable_2','r']
    corr_mat = corr_mat.loc[corr_mat['variable_1'] != corr_mat['variable_2'], :]
    corr_mat['abs_r'] = np.abs(corr_mat['r'])
    corr_mat = corr_mat.sort_values('abs_r', ascending=False)

    return(corr_mat)

corr_matrix = df_China.select_dtypes(include=['float64', 'int']) \
                .corr(method='pearson')
display(tidy_corr_matrix(corr_matrix).head(5))
```

	variable_1	variable_2	r	abs_r
1	Captured CO2/Mt	Captured cost/($/t)	-0.811049	0.811049
2	Captured cost/($/t)	Captured CO2/Mt	-0.811049	0.811049

Journal of CO2 Utilization 89 (2024) 102941

Contents lists available at ScienceDirect

Journal of CO2 Utilization

journal homepage: www.elsevier.com/locate/jcou

Review article

Machine learning of metal-organic framework design for carbon dioxide capture and utilization

Yang Jeong Park [a,b], Sungroh Yoon [b,c,*], Sung Eun Jerng [d,**]

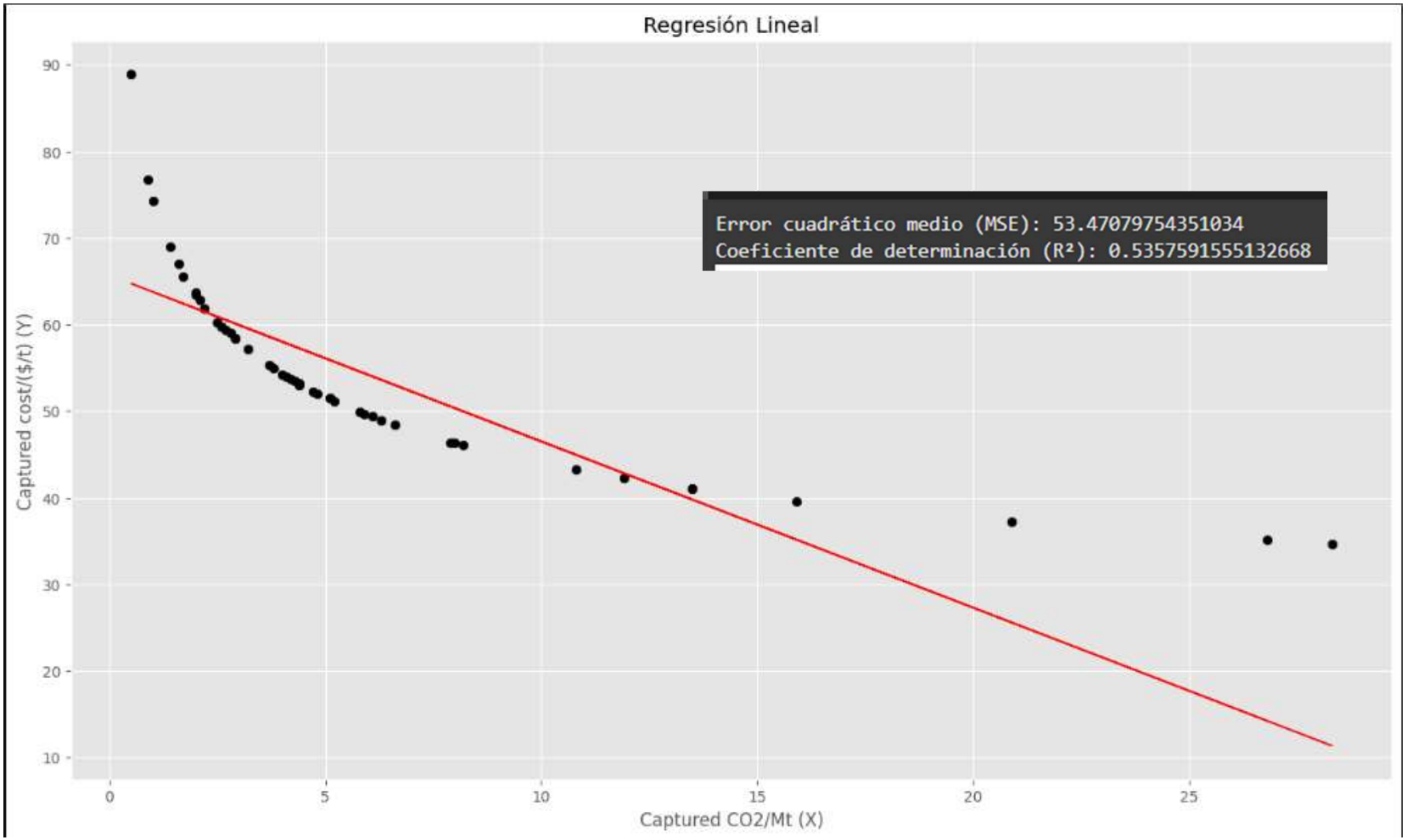
[a] Department of Materials Science and Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, 02139, USA
[b] Department of Electrical and Computer Engineering, Seoul National University, Seoul, 08826, Republic of Korea
[c] Interdisciplinary Program in Artificial Intelligence, Seoul National University, Seoul, 08826, Republic of Korea
[d] Department of Environmental Energy Engineering, The University of Suwon, 17, Wauan-gil, Bongdam-eup, Hwaseong-si, Gyeonggi-do, 18323, Republic of Korea

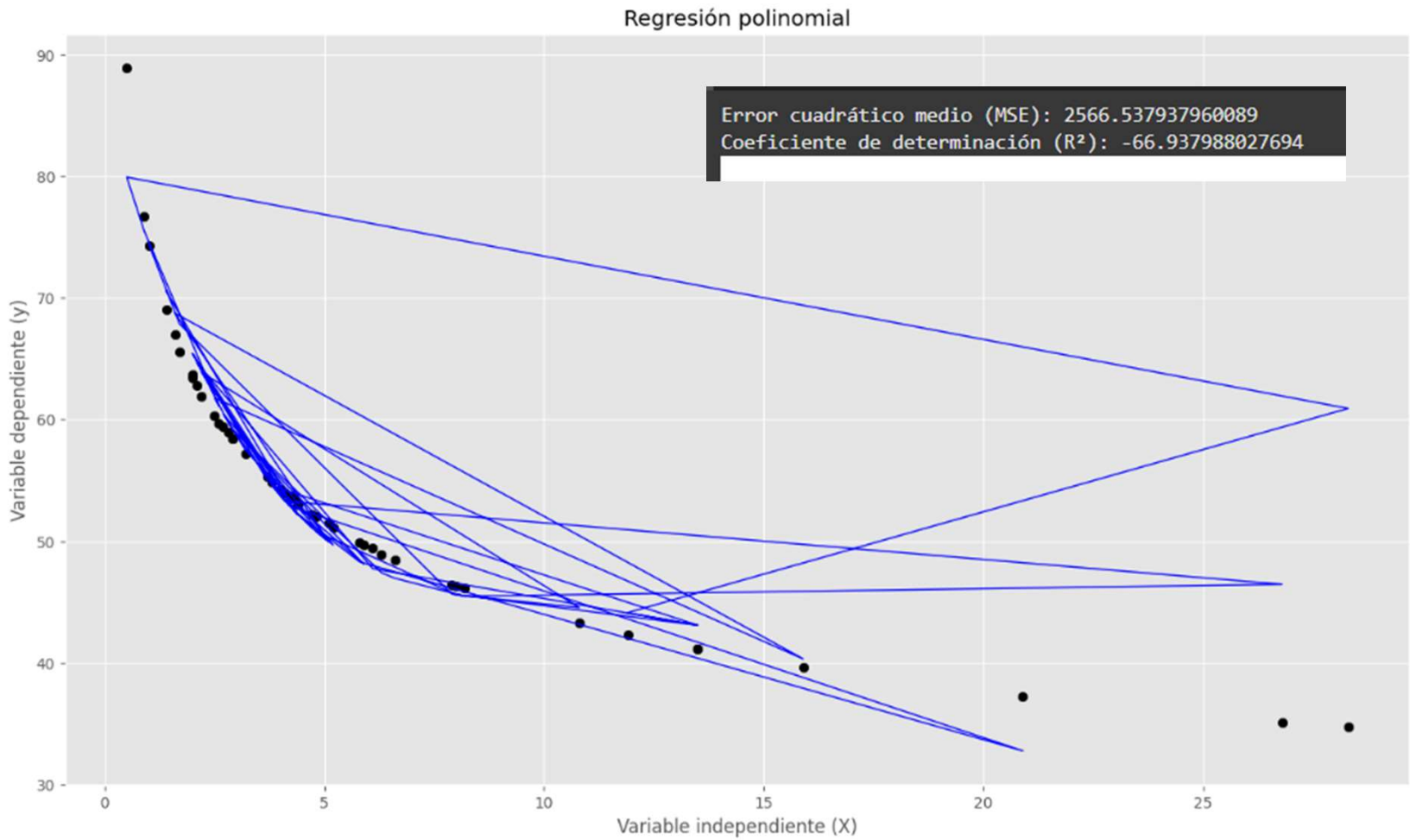
3.2.1. Root Mean Square Error (RMSE)

RMSE measures the square root of the average squared differences between predicted values ($\hat{y}_i$) and actual values (y_i). It is sensitive to large errors, making it useful when larger errors are particularly undesirable. The formula for RMSE is:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}$$

where n is the number of observations.


Regresión Lineal
Error cuadrático medio (MSE): 53.47079754351034
Coeficiente de determinación (R²): 0.5357591555132668
Captured cost/($/t) (Y)
Captured CO2/Mt (X)
90
80
70
60
50
40
30
20
10
0
5
10
15
20
25


Regresión polinomial
Error cuadrático medio (MSE): 2566.537937960089
Coeficiente de determinación (R²): -66.937988027694
Variable dependiente (y)
Variable independiente (X)
90
80
70
60
50
40
30
0
5
10
15
20
25

Código y Dashboard

Github

https://github.com/ManuAF-291196/Proyecto_TalentoTech/blob/data_analysis/Proyecto_CCS.ipynb

Referencias

- Data set 1 : https://netl.doe.gov/carbon-management/carbon-storage/worldwide-ccs-database
- Data set 2: https://www.iea.org/data-and-statistics/data-product/ccus-projects-database
- Data set 3: https://www.sciencedirect.com/science/article/pii/S0959652624017281?via%3Dihub#appsec1

Carbon Capture (CO_2)

2 de abril de 2025

ANÁLISIS DE DATOS DE LA TEMÁTICA DE CAPTURA DE CARBONO (CC) COMO HERRAMIENTA EMERGENTE PARA MITIGAR LAS EMISIONES DE DIÓXIDO DE CARBONO

- JHON FREDY RIOS HOYOS
- MANUEL ARANGO FIGUEROA
- PAULA ANDREA PEREZ ESPITIA

Fuentes de los Datasets

- Data set 1 : https://netl.doe.gov/carbon-management/carbon-storage/worldwide-ccs-database
- Data set 2: https://www.iea.org/data-and-statistics/data-product/ccus-projects-database
- Data set 3: https://www.sciencedirect.com/science/article/pii/S0959652624017281?via%3Dihub#appsec1

Objetivos específicos

1. **Relacionar el estado de proyectos** de las diferentes tecnologías de captura de carbono en función de los países donde se están desarrollando este tipo de proyectos.
2. **Estimar las diferentes correlaciones de las variables** de proyectos de investigación, tales como países, las tecnologías relacionadas con la captura de CO2, la capacidad de captura anunciada y la capacidad de captura estimada, en donde se están desarrollando estos proyectos de investigación.
3. **Establecer una relación** entre los datos de captura de CO2 y el costo de captura de CO2.

Análisis estadístico de datos de la temática de captura de CO_2 a partir de los tres data sets

	Data set 1	Data set 2	Data set 3
Objetivo 1 Visualización preliminar de datos / Análisis estadístico descriptivo	Diagrama de distribución, promedio	Diagrama de distribución, histograma, cajas y bigotes, serie de tiempos	Diagrama de distribución, histograma, cajas y bigotes,
Objetivo 2 (Misión 2) Conceptos estadísticos avanzados	No aplica	Evaluación de promedios y varianzas (No aplica PCA)	Análisis de componentes Principales (PCA) Modelado Regresión lineal Matriz de Covariabilidad
Objetivo 3 Dashboard (Misión 3)		Proyectos de todas las tecnologías a través del tiempo en Streamlit - Mapa	

Planteamiento del problema

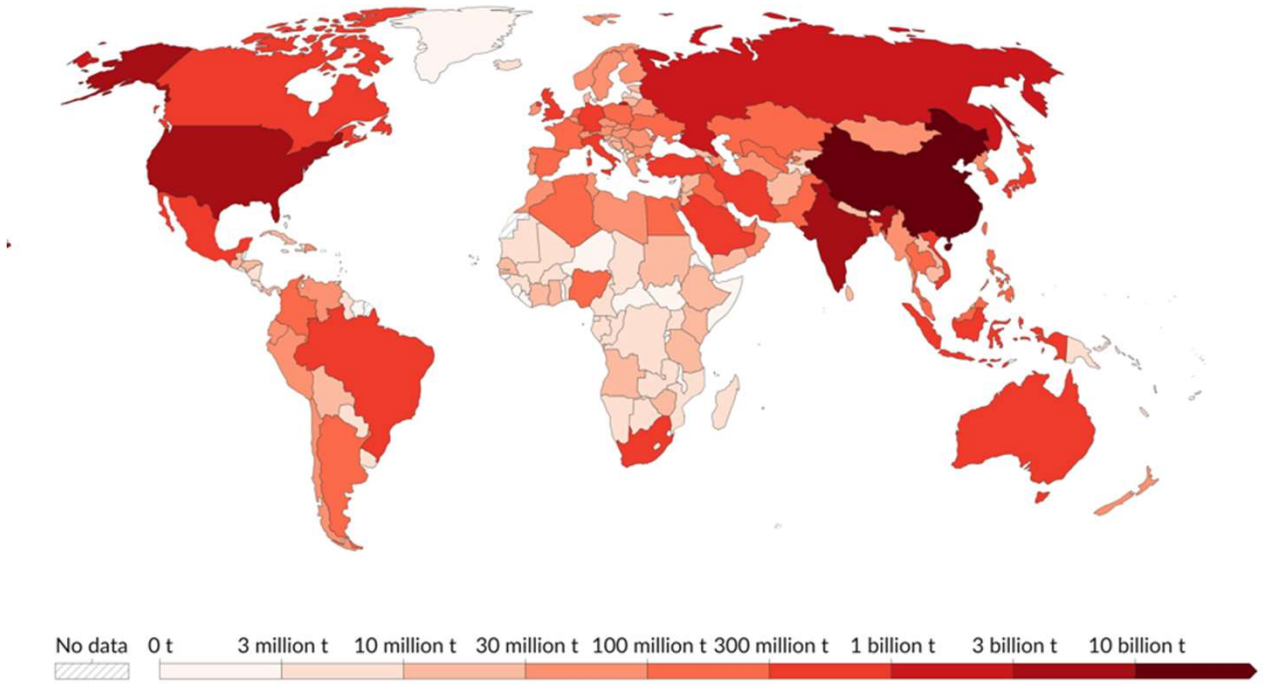
- A pesar del creciente consenso científico sobre la urgencia de mitigar el cambio climático, las **emisiones de dióxido de carbono (CO2)** continúan siendo una preocupación central.

La **descarbonización de la economía global** requiere una estrategia multifacética que incluye la reducción de la dependencia de combustibles fósiles, el aumento de la eficiencia energética y el desarrollo e implementación de tecnologías para capturar el CO2 emitido.

Annual CO_2 emissions, 2023

Our World in Data

Carbon dioxide (CO_2) emissions from fossil fuels and industry[1]. Land-use change is not included.



Data source: Global Carbon Budget (2024)

OurWorldinData.org/co2-and-greenhouse-gas-emissions | CC BY

1. **Fossil emissions**: Fossil emissions measure the quantity of carbon dioxide (CO_2) emitted from the burning of fossil fuels, and directly from industrial processes such as cement and steel production. Fossil CO_2 includes emissions from coal, oil, gas, flaring, cement, steel, and other industrial processes. Fossil emissions do not include land use change, deforestation, soils, or vegetation.

https://ourworldindata.org/grapher/annual-co2-emissions-per-country?tab=map&country=USA~GBR~IND~CHN~FRA~DEU~BRA~COL

Fuente del Data set 1

NATIONAL ENERGY TECHNOLOGY LABORATORY

Research and Programs Business Library

Carbon Capture and Storage Database

Welcome to the National Energy Technology Laboratory's (NETL) Carbon Capture and Storage (CCS) Database, which includes information on active, proposed, and terminated CCS projects worldwide.

Publicly available information has been aggregated to provide a one-stop interactive tool that contains valuable data, including, but not limited to:

- Technologies being developed for capture.
- Evaluation of sites for carbon dioxide (CO_2) storage.
- An estimation of project costs.
- Project description and current status.
- Amount of CO_2 captured/stored.

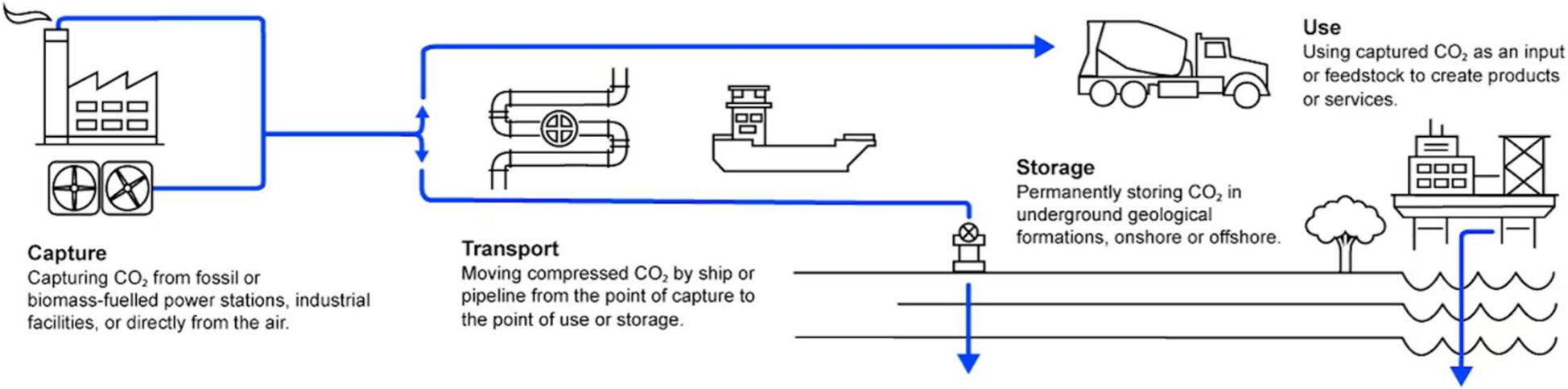
NETL's CCS Database provides the public with information regarding efforts by various industries, public groups, and governments that are being made towards development and eventual deployment of CCS technology. As of April 2018, the database contained 305 total CCS projects worldwide, with 299 sites identified. The 299 site-located projects include 76 capture, 76 storage, and 147 for capture and storage in more than 30 countries across 6 continents. While several of the projects are still in the planning and development stage, and many have been completed, 37 are actively capturing and/or injecting CO_2.

The CCS Database is presented using a Tableau Dashboard which is entirely interactive. Hovering, clicking, and/or dragging any of the icons will customize your view. Adjusting the toolbar on the left allows you to narrow your selection, pan, or zoom in/out on the map.

Diccionario

Captura	Captura de CO_2
CCS	De sus siglas en inglés (Capture Carbon Storage) Captura y almacenamiento de CO_2
Almacenamiento	Almacenamiento de CO_2 solamente.
T & S	De sus siglas en inglés (Transport and Storage) Transporte y almacenamiento de CO_2
Transporte	Transporte posterior del almacenamiento de CO_2
CCU	De sus siglas en inglés Carbon, Capture, Utilisation
Cadena completa	Captura, Almacenamiento, Transporte y utilización

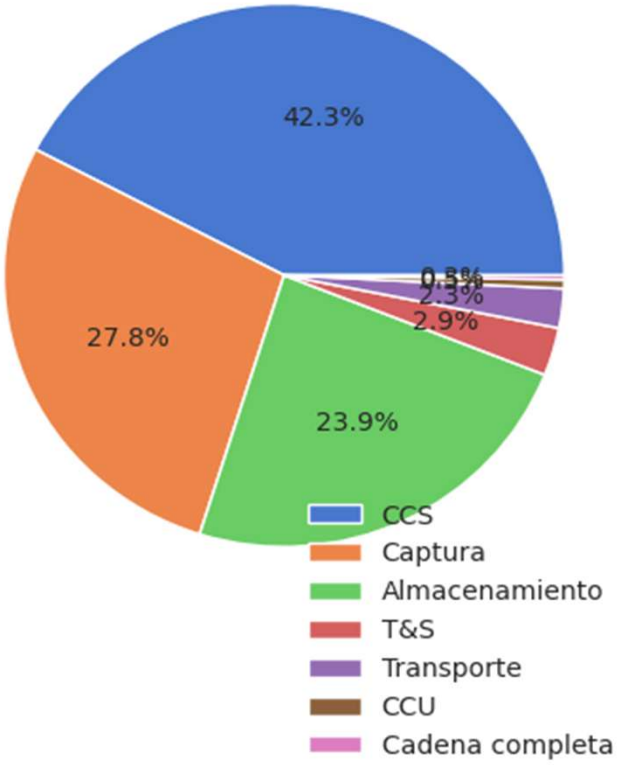
The CCUS Chain



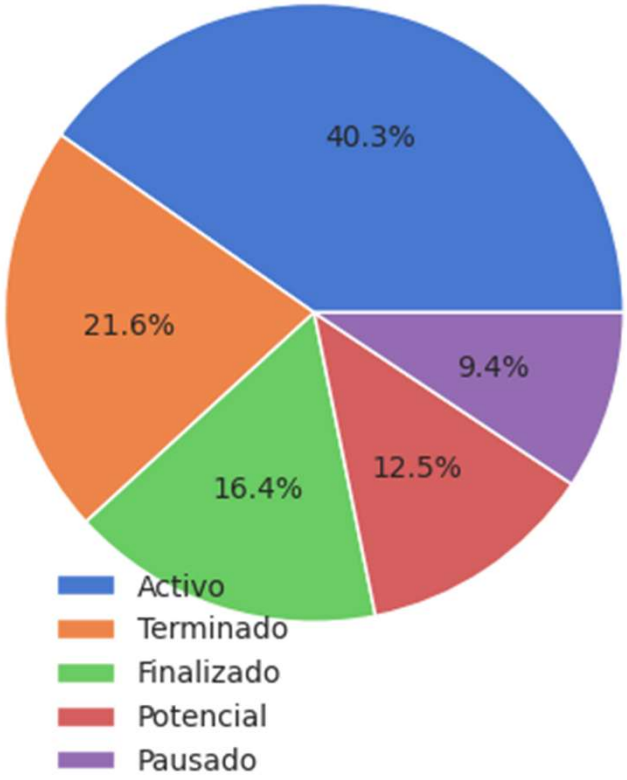
https://www.iea.org/energy-system/carbon-capture-utilisation-and-storage#how-does-ccus-work

Data set 1

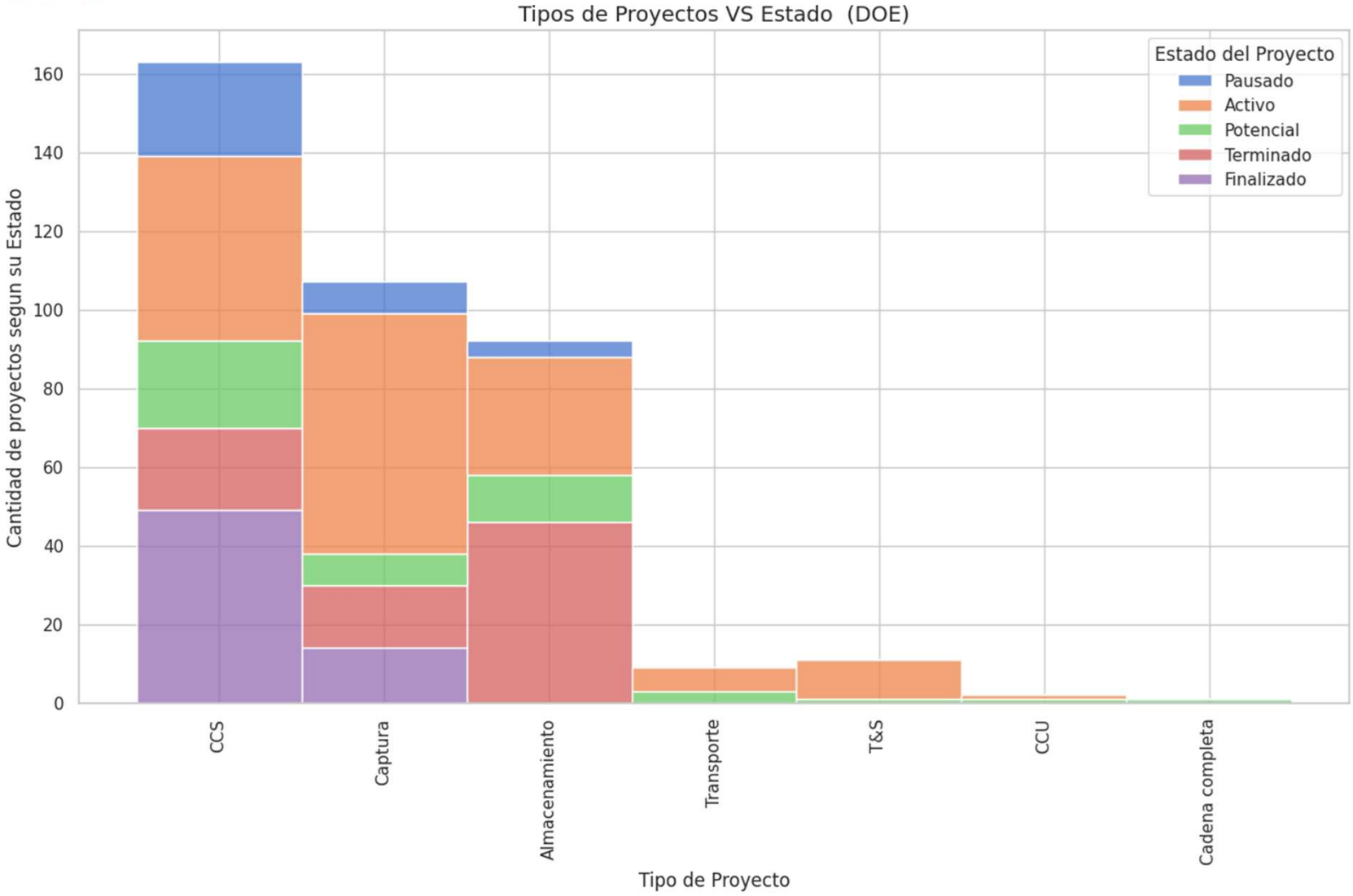
Tipo de los Proyecto de Captura de CO2 (DOE)



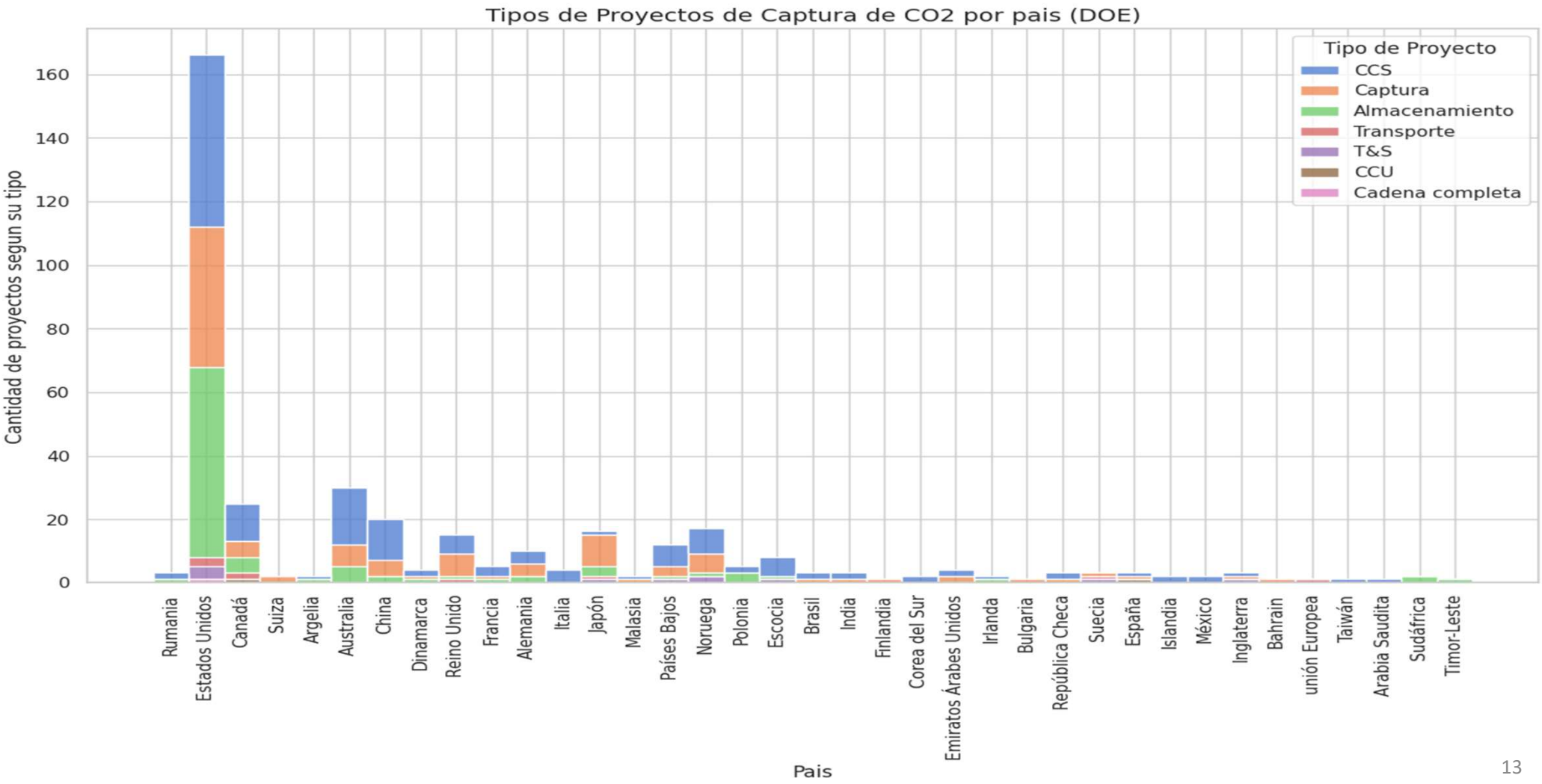
Estado de los Proyecto de Captura de CO2 (DOE)



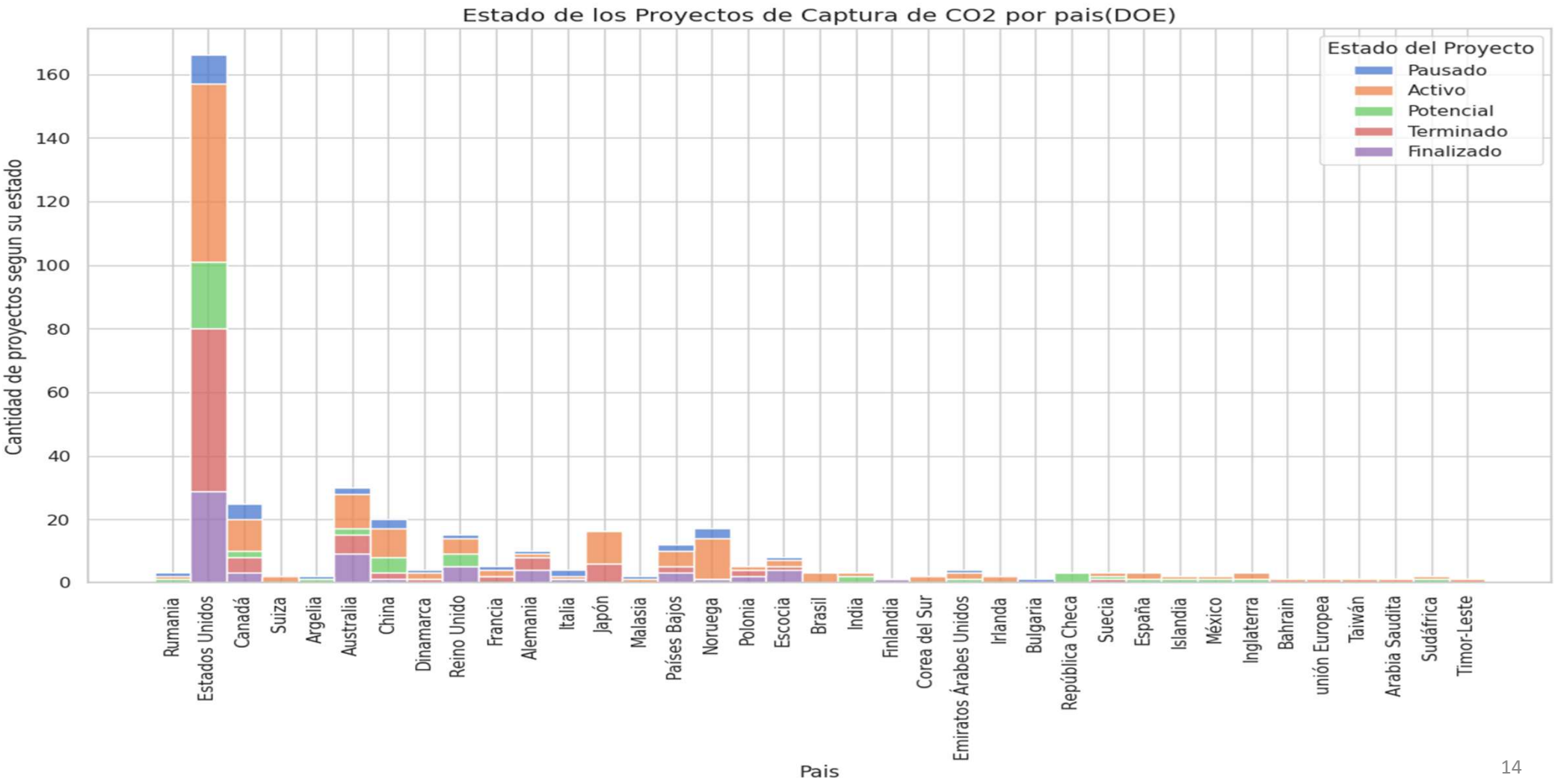
Data set 1


Tipos de Proyectos VS Estado (DOE)
Estado del Proyecto
Pausado
Activo
Potencial
Terminado
Finalizado
Cantidad de proyectos segun su Estado
0
20
40
60
80
100
120
140
160
CCS
Captura
Almacenamiento
Transporte
T&S
CCU
Cadena completa
Tipo de Proyecto

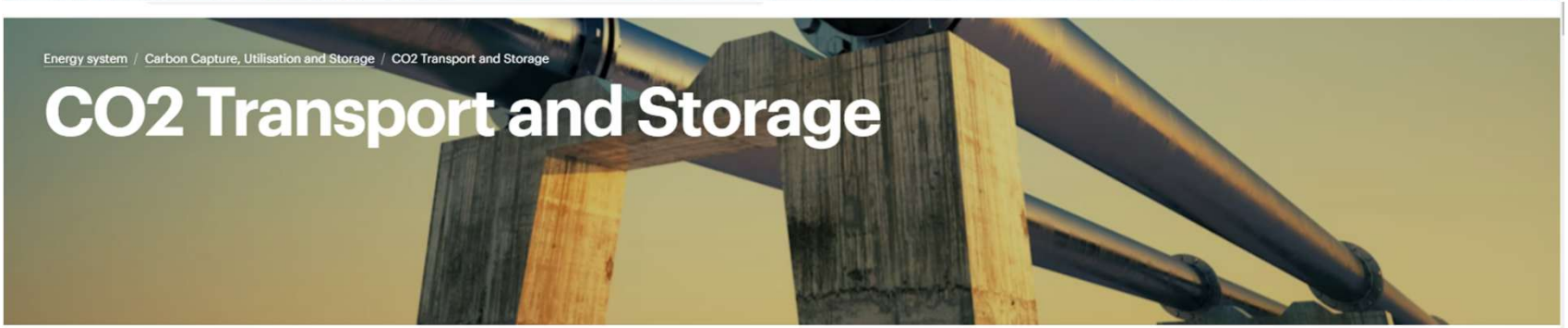
Data set 1


Tipos de Proyectos de Captura de CO2 por pais (DOE)
Tipo de Proyecto
CCS
Captura
Almacenamiento
Transporte
T&S
CCU
Cadena completa
Cantidad de proyectos segun su tipo
0
20
40
60
80
100
120
140
160
Rumania
Estados Unidos
Canadá
Suiza
Argelia
Australia
China
Dinamarca
Reino Unido
Francia
Alemania
Italia
Japón
Malasia
Países Bajos
Noruega
Polonia
Escocia
Brasil
India
Finlandia
Corea del Sur
Emiratos Árabes Unidos
Irlanda
Bulgaria
República Checa
Suecia
España
Islandia
México
Inglaterra
Bahrain
unión Europea
Taiwán
Arabia Saudita
Sudáfrica
Timor-Leste
Pais

Data set 1

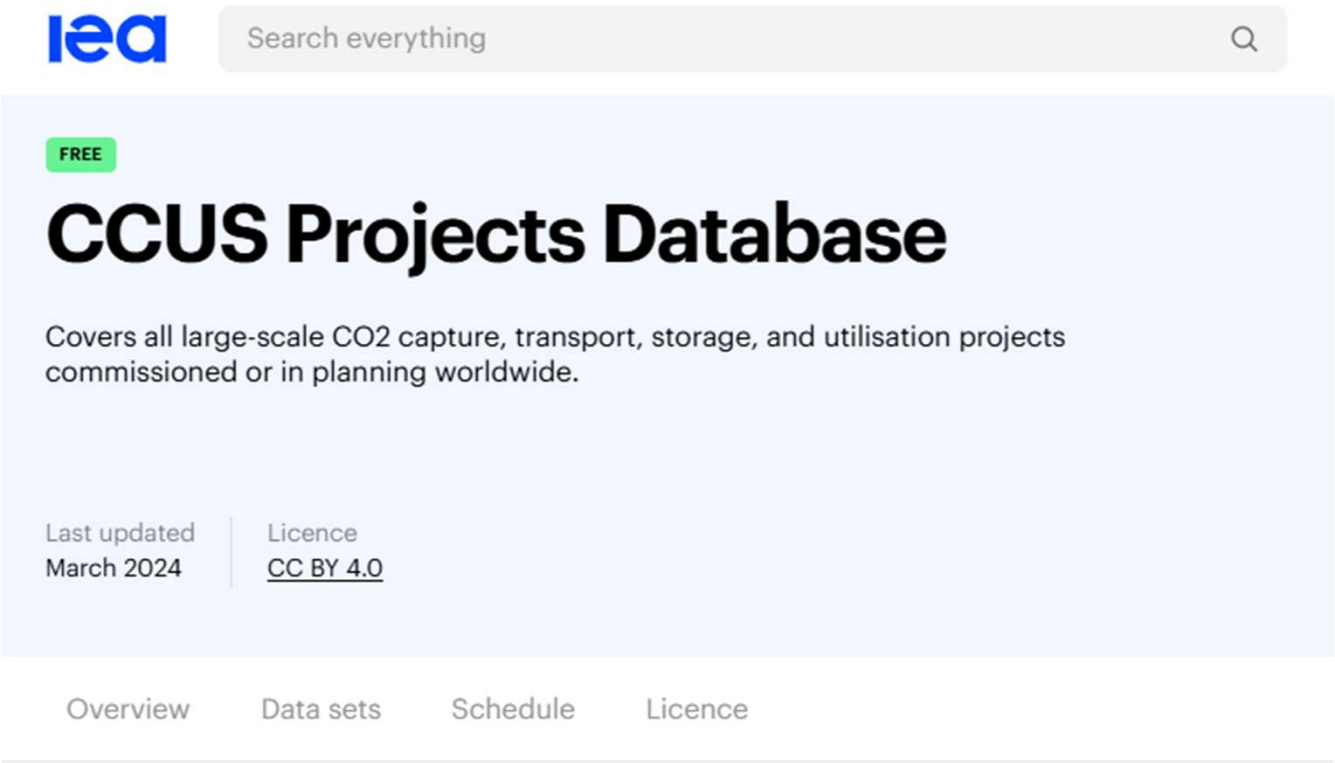

Estado de los Proyectos de Captura de CO2 por pais(DOE)
Estado del Proyecto
Pausado
Activo
Potencial
Terminado
Finalizado
Cantidad de proyectos segun su estado
0
20
40
60
80
100
120
140
160
Rumania
Estados Unidos
Canadá
Suiza
Argelia
Australia
China
Dinamarca
Reino Unido
Francia
Alemania
Italia
Japón
Malasia
Países Bajos
Noruega
Polonia
Escocia
Brasil
India
Finlandia
Corea del Sur
Emiratos Árabes Unidos
Irlanda
Bulgaria
República Checa
Suecia
España
Islandia
México
Inglaterra
Bahrain
unión Europea
Taiwán
Arabia Saudita
Sudáfrica
Timor-Leste
Pais

Data set2



https://www.iea.org/energy-system/carbon-capture-utilisation-and-storage/co2-capture-and-utilisation

Fuente del Data set 2



Data set 2 Visualización de datos

skimpy summary

Data Summary

Dataframe	Values
Number of rows	844
Number of columns	17

Data Types

Column Type	Count
string	11
datetime64	4
float64	2

number

column	NA	NA %	mean	sd	p0	p25	p50	p75	p100	hist
Announced capacity (Mt CO2/yr)	230	27.251184834123222	3.39	11.89	0	0.4	1.2	3	250	
Estimated capacity by IEA (Mt CO2/yr)	199	23.578199052132703	3.45	11.77	0.001	0.4	1.3	3	250	

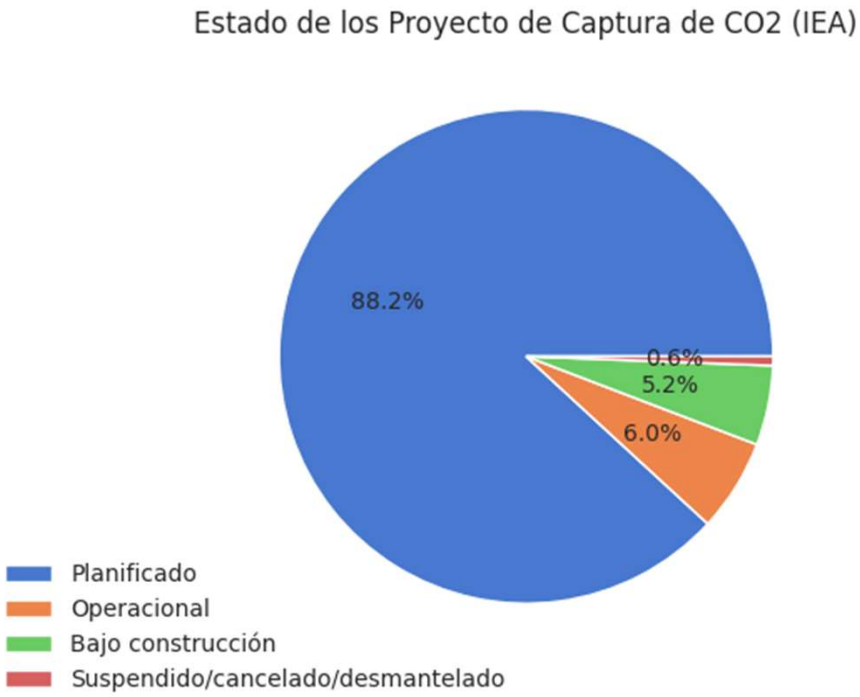
datetime

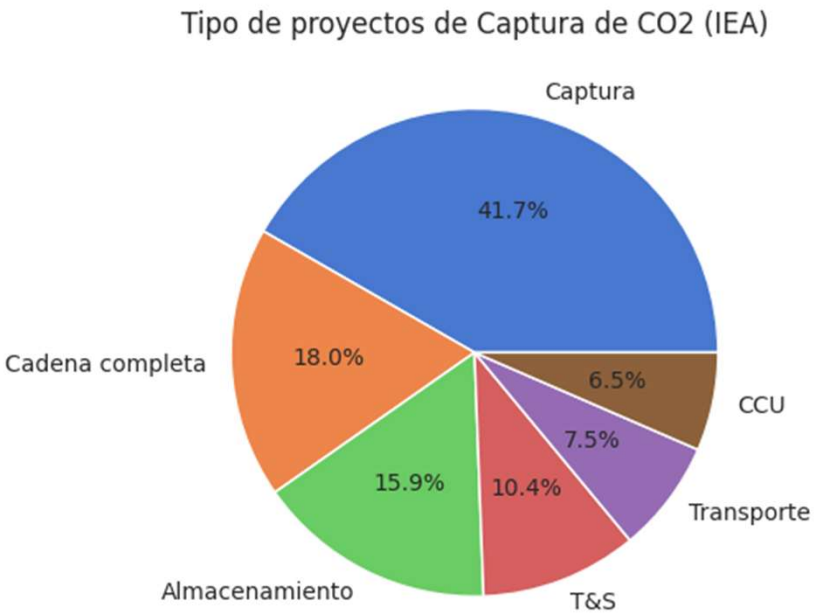
column	NA	NA %	first	last	frequency
Announcement	20	2.3696682464454977	1982-01-01	2024-01-01	None
FID	589	69.7867298578199	1984-01-01	2029-01-01	None
Operation	291	34.47867298578199	1972-01-01	2040-01-01	None
Suspension/decommissioning	838	99.28909952606635	2004-01-01	2026-01-01	None

Data set 2 Estadísticos descriptivos

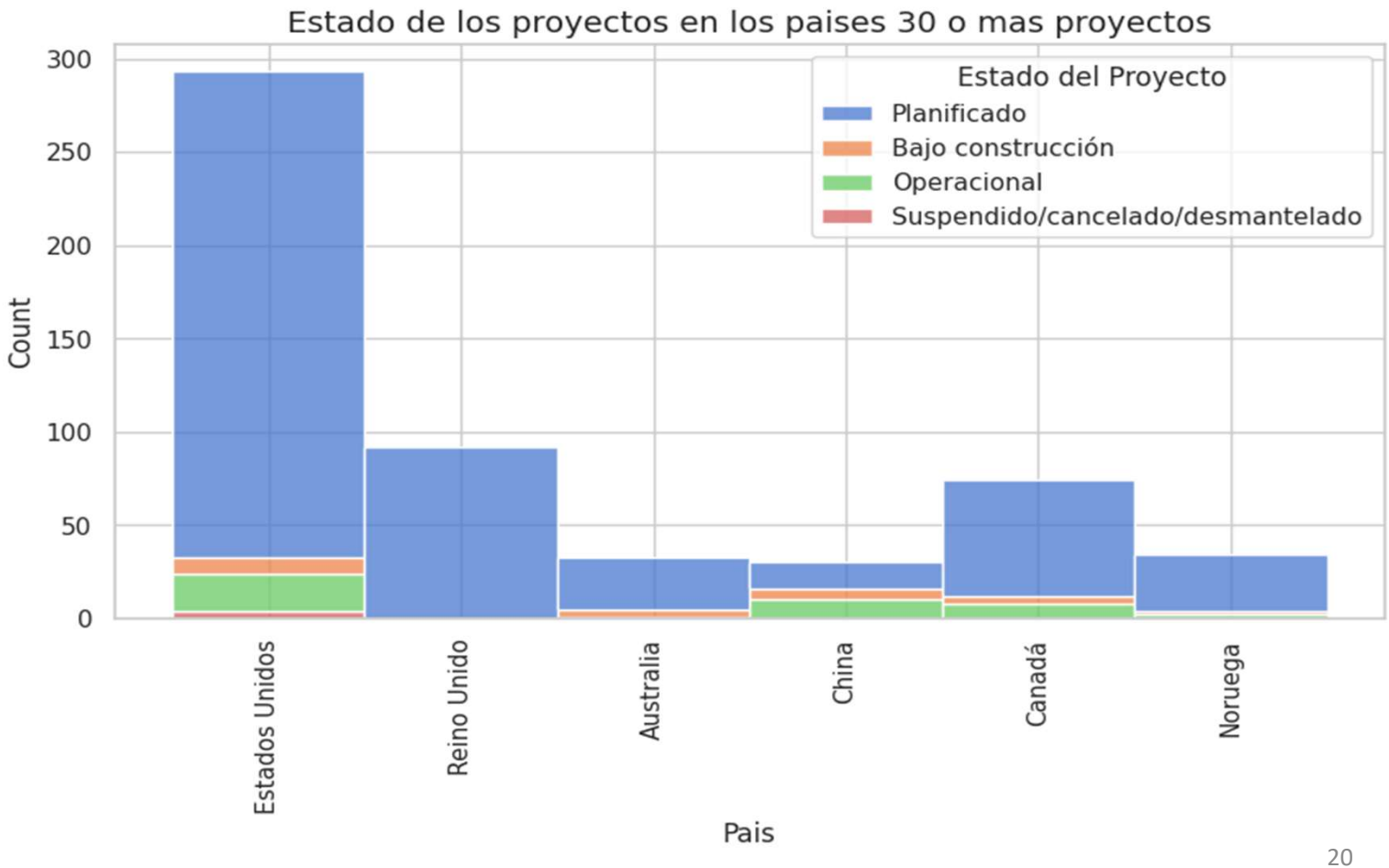
	Announcement	FID	Operation	Suspension/decommissioning	Announced capacity (Mt CO2/yr)	Estimated capacity by IEA (Mt CO2/yr)
count	824	255	553	6	614.000000	645.000000
mean	2020-10-27 23:21:33.203883776	2022-11-06 03:34:35.294117632	2026-03-05 22:47:05.316455680	2017-09-01 04:00:00	3.389782	3.449963
min	1982-01-01 00:00:00	1984-01-01 00:00:00	1972-01-01 00:00:00	2004-01-01 00:00:00	0.000000	0.001000
25%	2021-01-01 00:00:00	2023-01-01 00:00:00	2025-01-01 00:00:00	2012-07-02 00:00:00	0.400000	0.400000
50%	2022-01-01 00:00:00	2024-01-01 00:00:00	2027-01-01 00:00:00	2020-01-01 12:00:00	1.200000	1.300000
75%	2023-01-01 00:00:00	2024-01-01 00:00:00	2029-01-01 00:00:00	2024-07-02 06:00:00	3.000000	3.000000
max	2024-01-01 00:00:00	2029-01-01 00:00:00	2040-01-01 00:00:00	2026-01-01 00:00:00	250.000000	250.000000
std	NaN	NaN	NaN	NaN	11.885486	11.767173

Data set 2 Diagramas de distribución



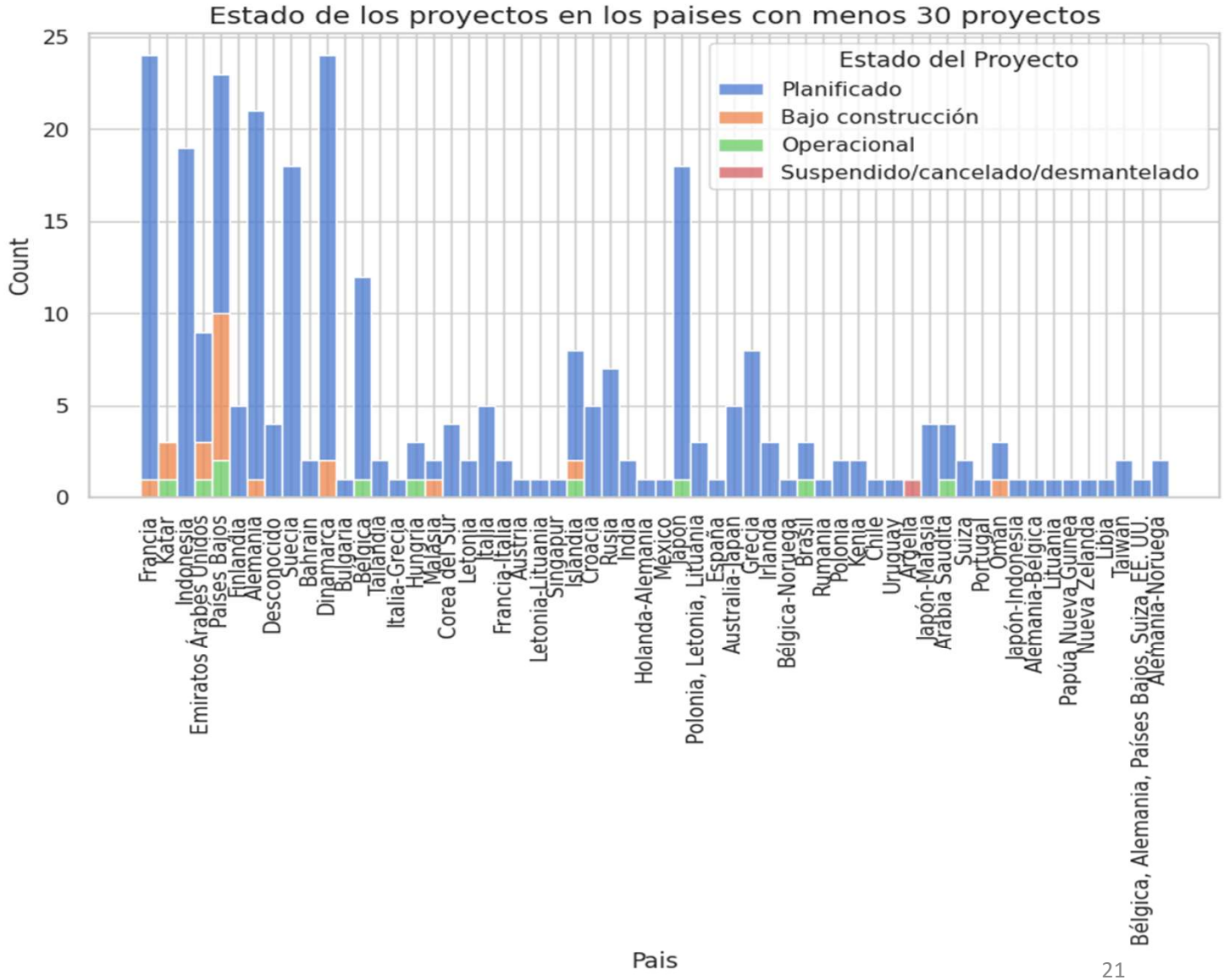


Data set 2


Estado de los proyectos en los paises 30 o mas proyectos
Estado del Proyecto
Planificado
Bajo construcción
Operacional
Suspendido/cancelado/desmantelado
Count
300
250
200
150
100
50
0
Estados Unidos
Reino Unido
Australia
China
Canadá
Noruega
Pais

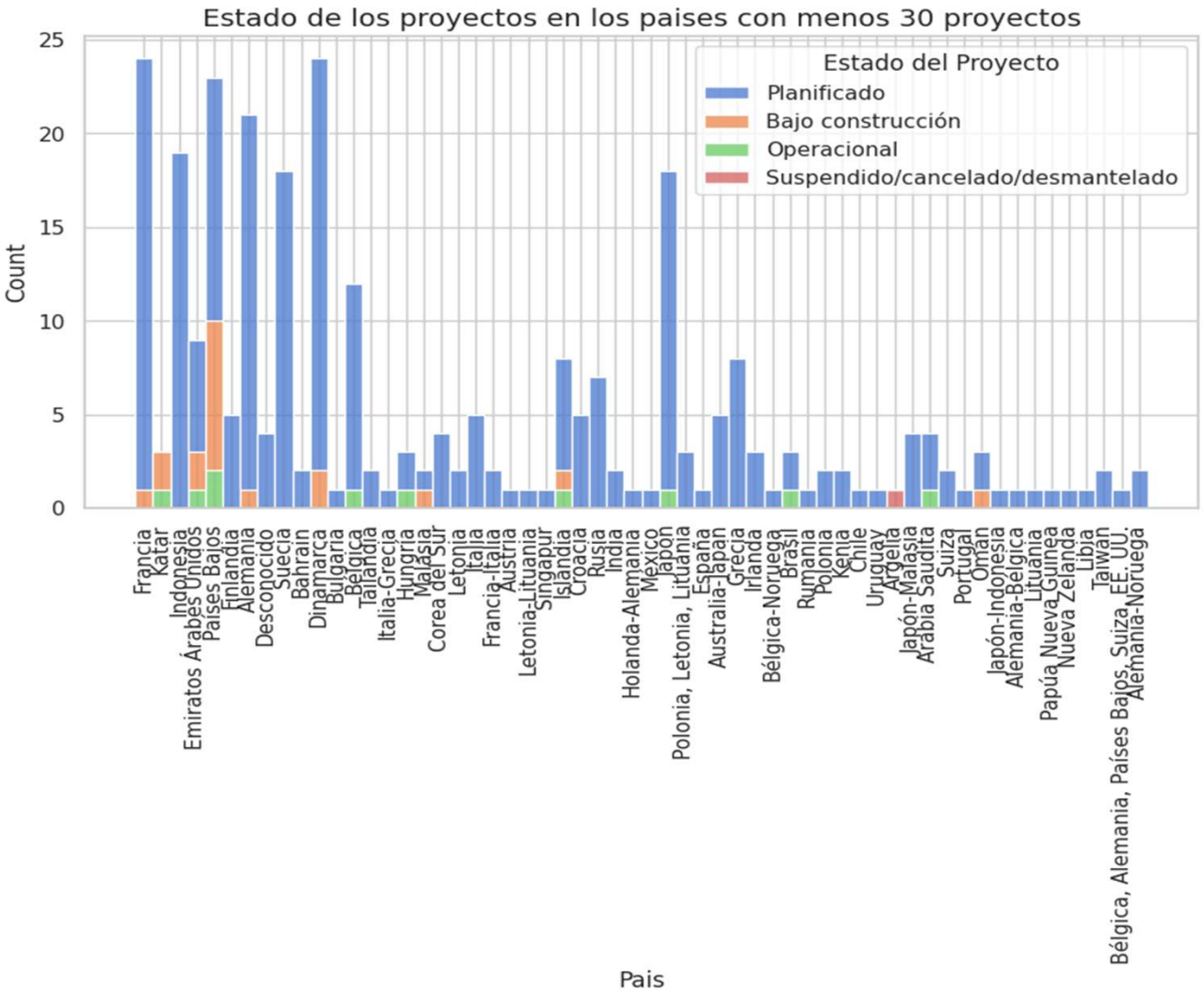
Data set 2

Francia
Katar
Indonesia
Emiratos Árabes Unidos
Países Bajos
Finlandia
Alemania
Desconocido
Suecia
Bahrain
Dinamarca
Bulgaria
Bélgica
Tailandia
Italia-Grecia
Hungría
Malasia
Corea del Sur
Letonia
Italia
Francia-Italia
Austria
Letonia-Lituania
Singapur
Islandia
Croacia
Rusia
India

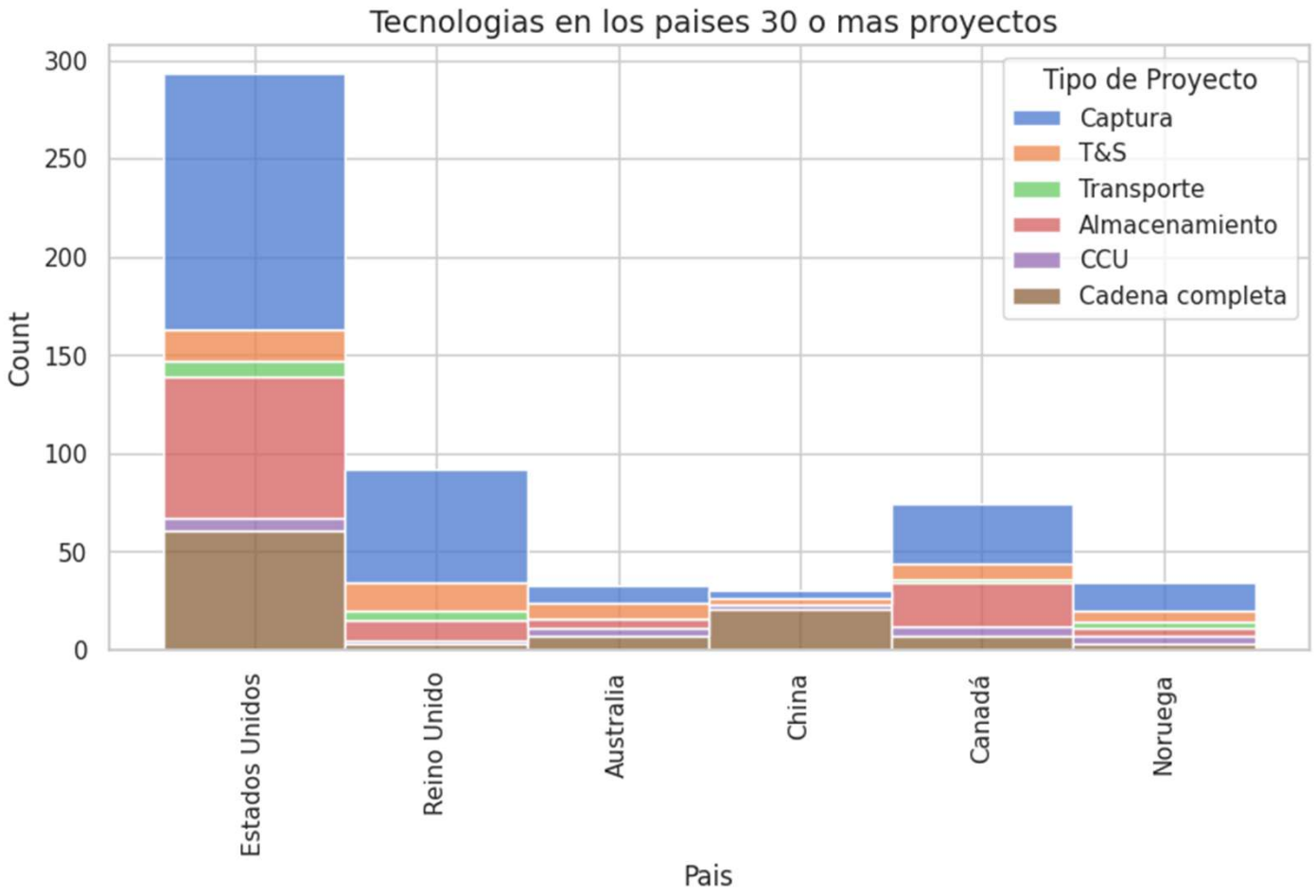

Estado de los proyectos en los paises con menos 30 proyectos
Estado del Proyecto
Planificado
Bajo construcción
Operacional
Suspendido/cancelado/desmantelado
Count
0
5
10
15
20
25
Francia
Katar
Indonesia
Emiratos Árabes Unidos
Países Bajos
Finlandia
Alemania
Desconocido
Suecia
Bahrain
Dinamarca
Bulgaria
Bélgica
Tailandia
Italia-Grecia
Hungría
Malasia
Corea del Sur
Letonia
Italia
Francia-Italia
Austria
Letonia-Lituania
Singapur
Islandia
Croacia
Rusia
India
Holanda-Alemania
Mexico
Japón
Polonia, Letonia, Lituania
España
Australia-Japan
Grecia
Irlanda
Bélgica-Noruega
Brasil
Rumania
Polonia
Kenia
Chile
Uruguay
Argelia
Japón-Malasia
Arabia Saudita
Suiza
Portugal
Oman
Japón-Indonesia
Alemania-Bélgica
Lituania
Papúa Nueva Guinea
Nueva Zelanda
Libia
Taiwan
Bélgica, Alemania, Países Bajos, Suiza, EE. UU.
Alemania-Noruega
Pais

Data set 2

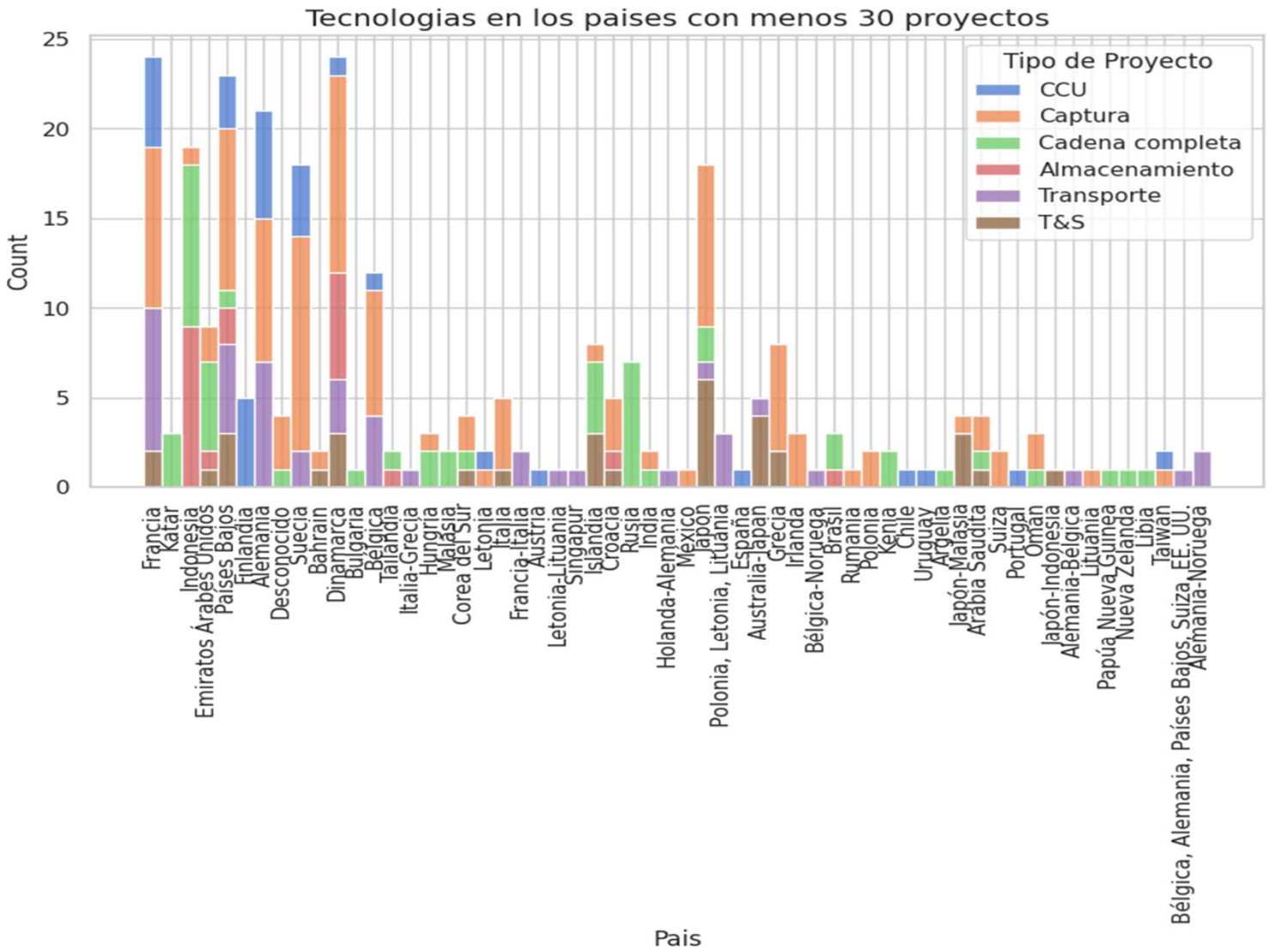
Holanda-Alemania
Mexico
Japón
Polonia, Letonia, Lituania
España
Australia-Japan
Grecia
Irlanda
Bélgica-Noruega
Brasil
Rumania
Polonia
Kenia
Chile
Uruguay
Argelia
Japón-Malasia
Arabia Saudita
Suiza
Portugal
Oman
Japón-Indonesia
Alemania-Bélgica
Lituania
Papúa Nueva Guinea
Nueva Zelanda
Libia
Taiwan
Bélgica, Alemania, Países Bajos, Suiza, EE. UU.
Alemania-Noruega


Estado de los proyectos en los paises con menos 30 proyectos
Estado del Proyecto
Planificado
Bajo construcción
Operacional
Suspendido/cancelado/desmantelado
Count
0
5
10
15
20
25
Francia
Katar
Indonesia
Emiratos Árabes Unidos
Países Bajos
Finlandia
Alemania
Desconocido
Suecia
Bahrain
Dinamarca
Bulgaria
Bélgica
Tailandia
Italia-Grecia
Hungría
Malasia
Corea del Sur
Letonia
Italia
Francia-Italia
Austria
Letonia-Lituania
Singapur
Islandia
Croacia
Rusia
India
Holanda-Alemania
Mexico
Japón
Polonia, Letonia, Lituania
España
Australia-Japan
Grecia
Irlanda
Bélgica-Noruega
Brasil
Rumania
Polonia
Kenia
Chile
Uruguay
Argelia
Japón-Malasia
Arabia Saudita
Suiza
Portugal
Oman
Japón-Indonesia
Alemania-Bélgica
Lituania
Papúa Nueva Guinea
Nueva Zelanda
Libia
Taiwan
Bélgica, Alemania, Países Bajos, Suiza, EE. UU.
Alemania-Noruega
Pais

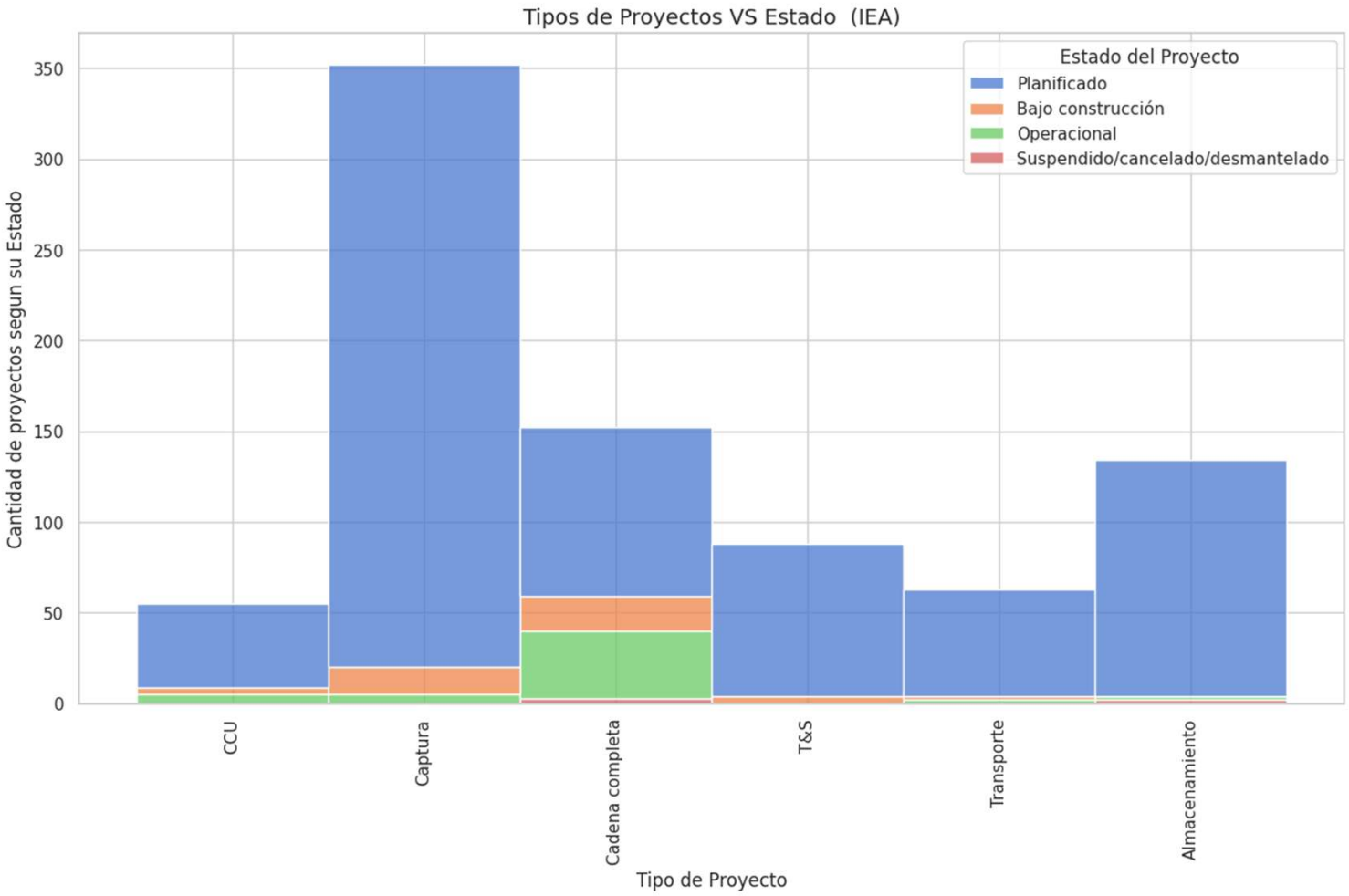
Data set 2


Tecnologias en los paises 30 o mas proyectos
Tipo de Proyecto
Captura
T&S
Transporte
Almacenamiento
CCU
Cadena completa
Count
300
250
200
150
100
50
0
Estados Unidos
Reino Unido
Australia
China
Canadá
Noruega
Pais

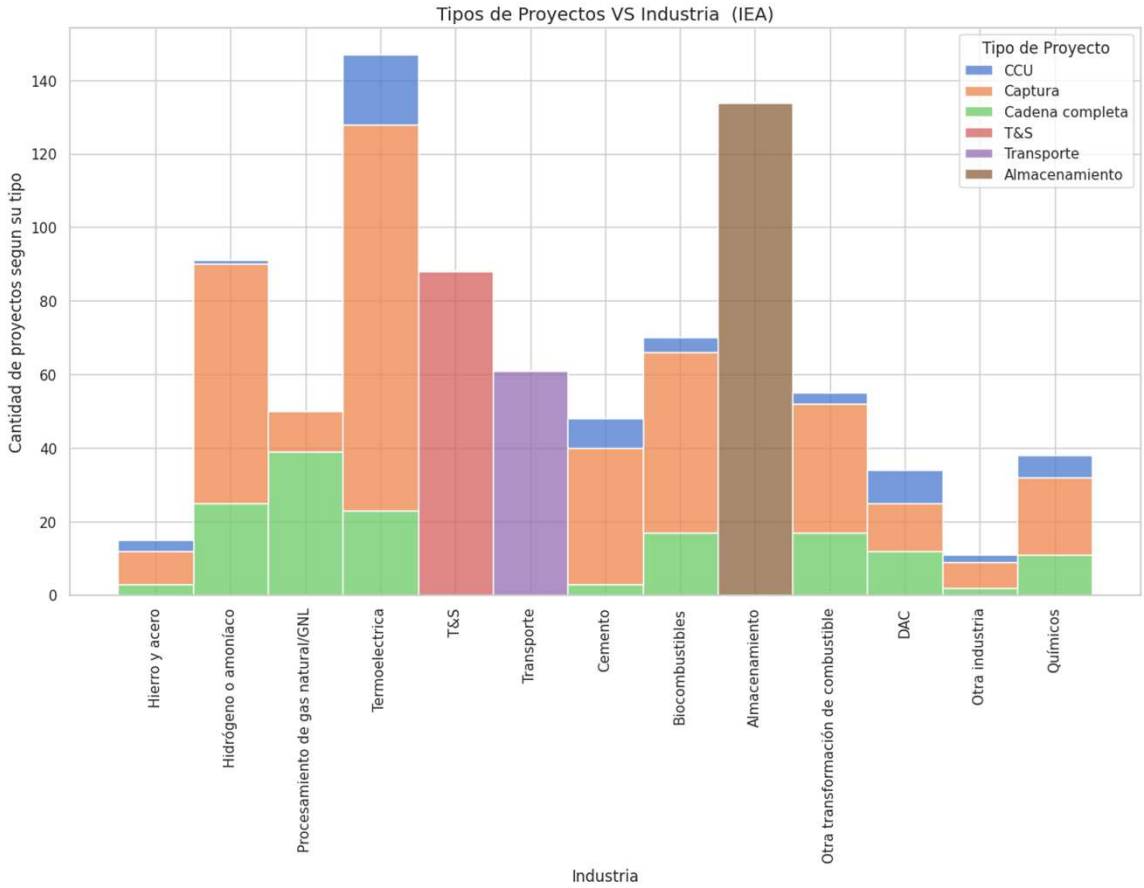
Data set 2


Tecnologias en los paises con menos 30 proyectos
Tipo de Proyecto
CCU
Captura
Cadena completa
Almacenamiento
Transporte
T&S
Count
0
5
10
15
20
25
Francia
Katar
Indonesia
Emiratos Árabes Unidos
Países Bajos
Finlandia
Alemania
Desconocido
Suecia
Bahrain
Dinamarca
Bulgaria
Bélgica
Tailandia
Italia-Grecia
Hungría
Malasia
Corea del Sur
Letonia
Italia
Francia-Italia
Austria
Letonia-Lituania
Singapur
Islandia
Croacia
Rusia
India
Holanda-Alemania
México
Japón
Polonia, Letonia, Lituania
España
Australia-Japan
Grecia
Irlanda
Bélgica-Noruega
Brasil
Rumania
Polonia
Kenia
Chile
Uruguay
Argelia
Japón-Malasia
Arabia Saudita
Suiza
Portugal
Oman
Japón-Indonesia
Alemania-Bélgica
Lituania
Papúa Nueva Guinea
Nueva Zelanda
Libia
Taiwan
Bélgica, Alemania, Países Bajos, Suiza, EE. UU.
Alemania-Noruega
Pais

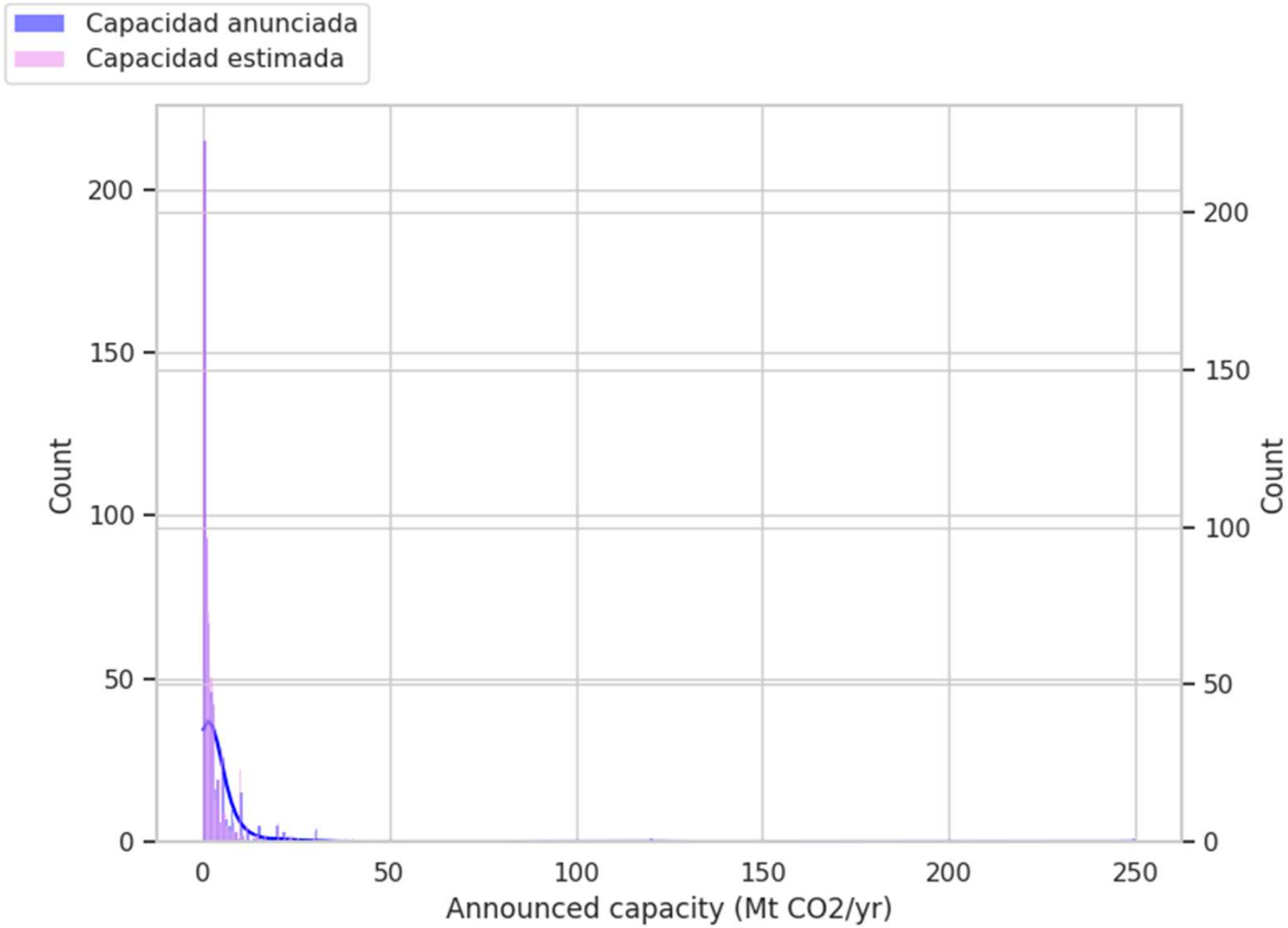
Data set 2


Tipos de Proyectos VS Estado (IEA)
Estado del Proyecto
Planificado
Bajo construcción
Operacional
Suspendido/cancelado/desmantelado
Cantidad de proyectos segun su Estado
0
50
100
150
200
250
300
350
CCU
Captura
Cadena completa
T&S
Transporte
Almacenamiento
Tipo de Proyecto

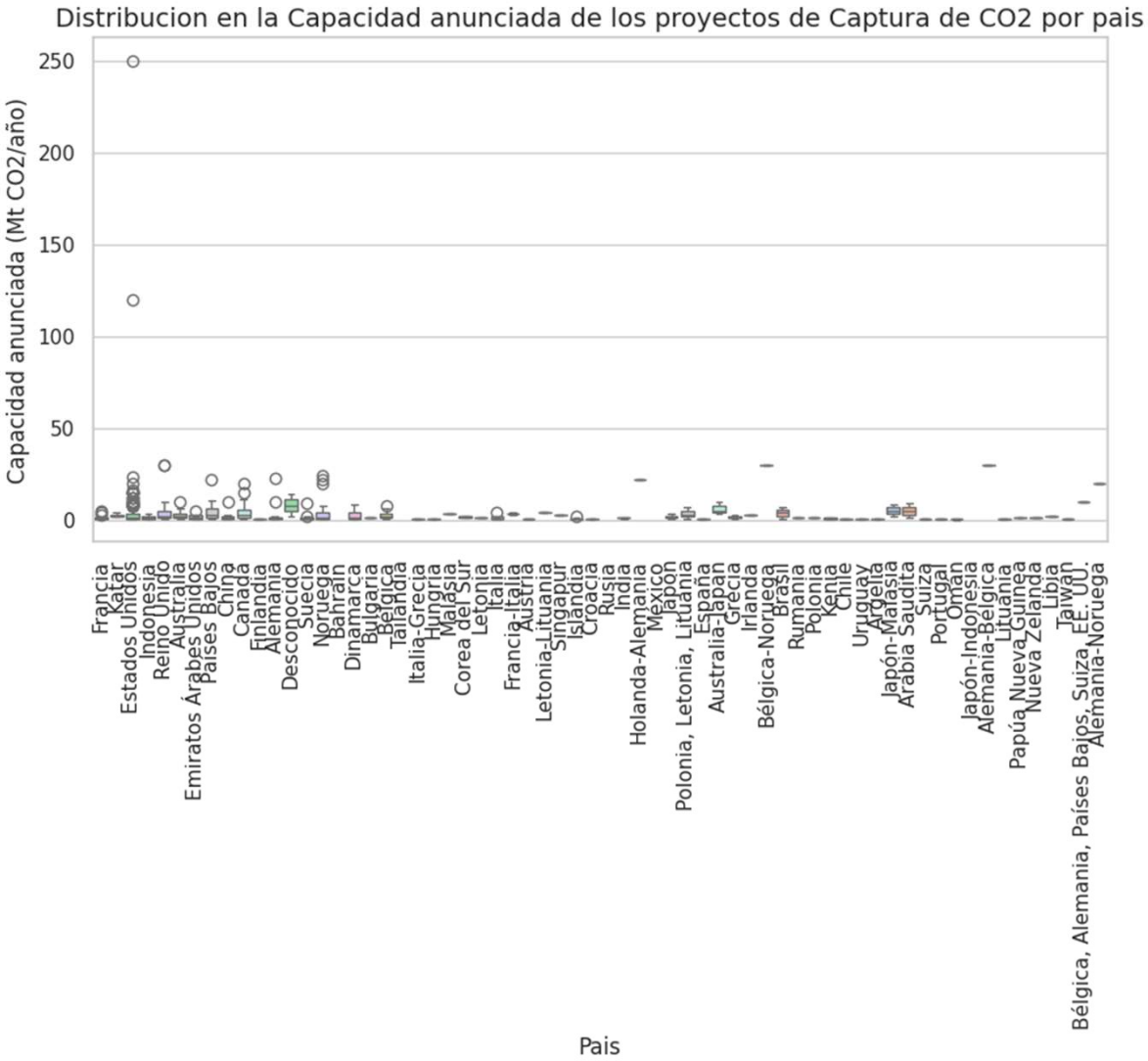
Data set 2


Tipos de Proyectos VS Industria (IEA)
Tipo de Proyecto
CCU
Captura
Cadena completa
T&S
Transporte
Almacenamiento
Cantidad de proyectos segun su tipo
0
20
40
60
80
100
120
140
Hierro y acero
Hidrógeno o amoníaco
Procesamiento de gas natural/GNL
Termoelectrica
T&S
Transporte
Cemento
Biocombustibles
Almacenamiento
Otra transformación de combustible
DAC
Otra industria
Químicos
Industria

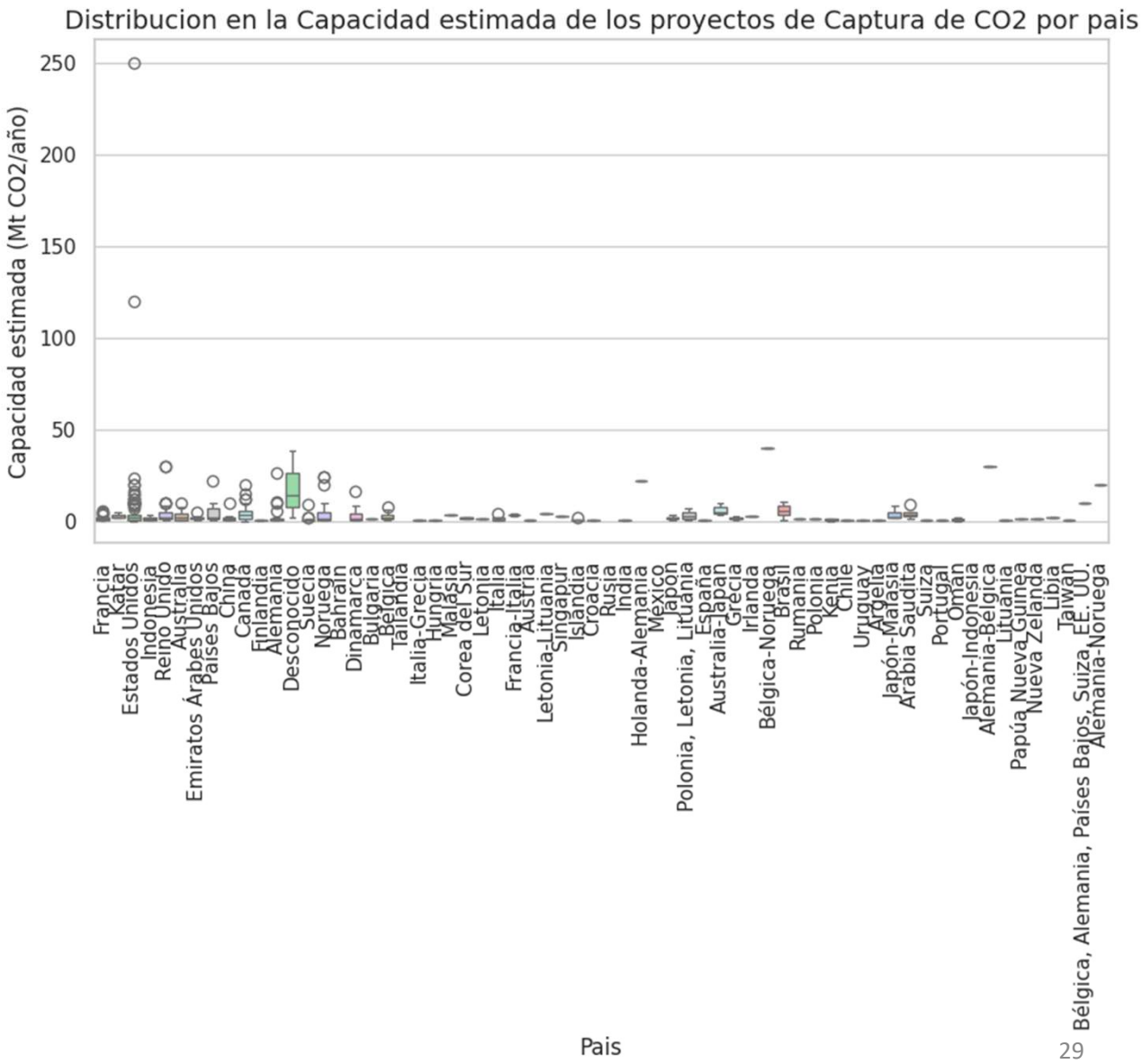
Data set 2
Histograma



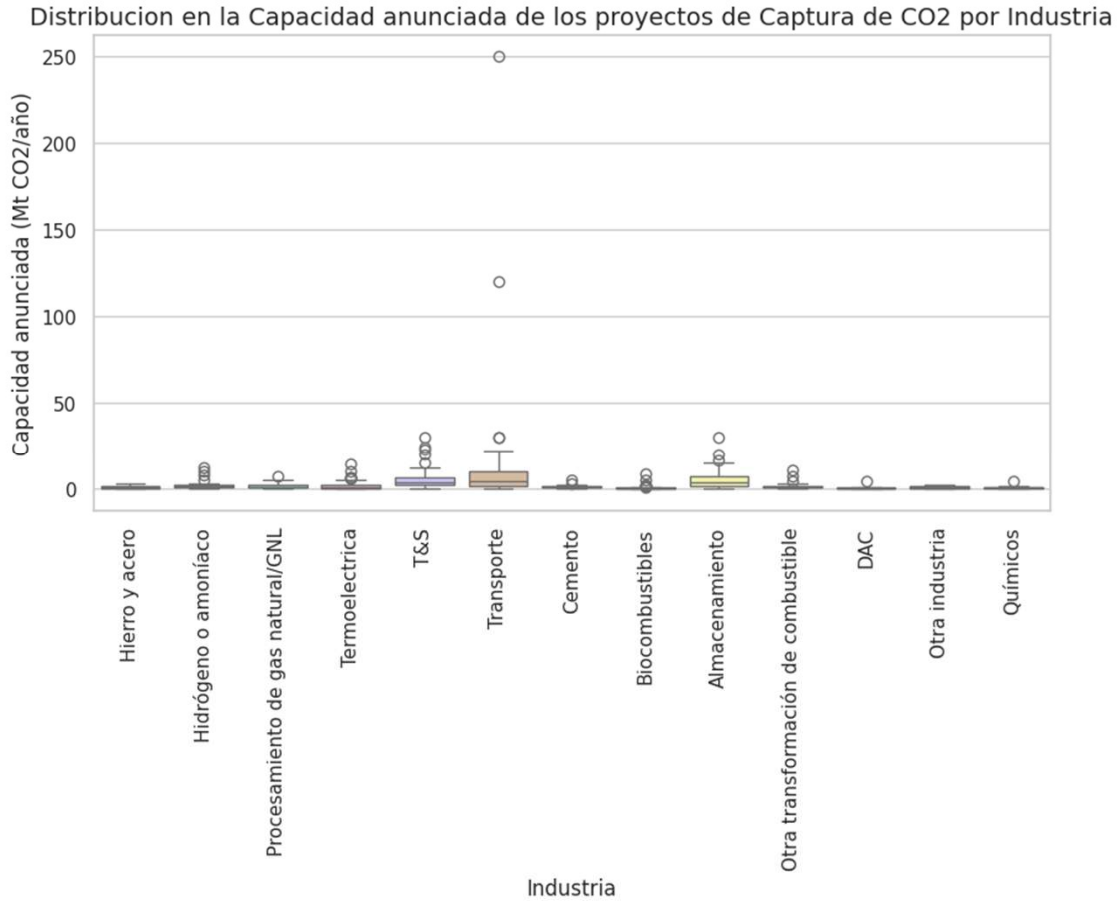
Data set 2
Box plot



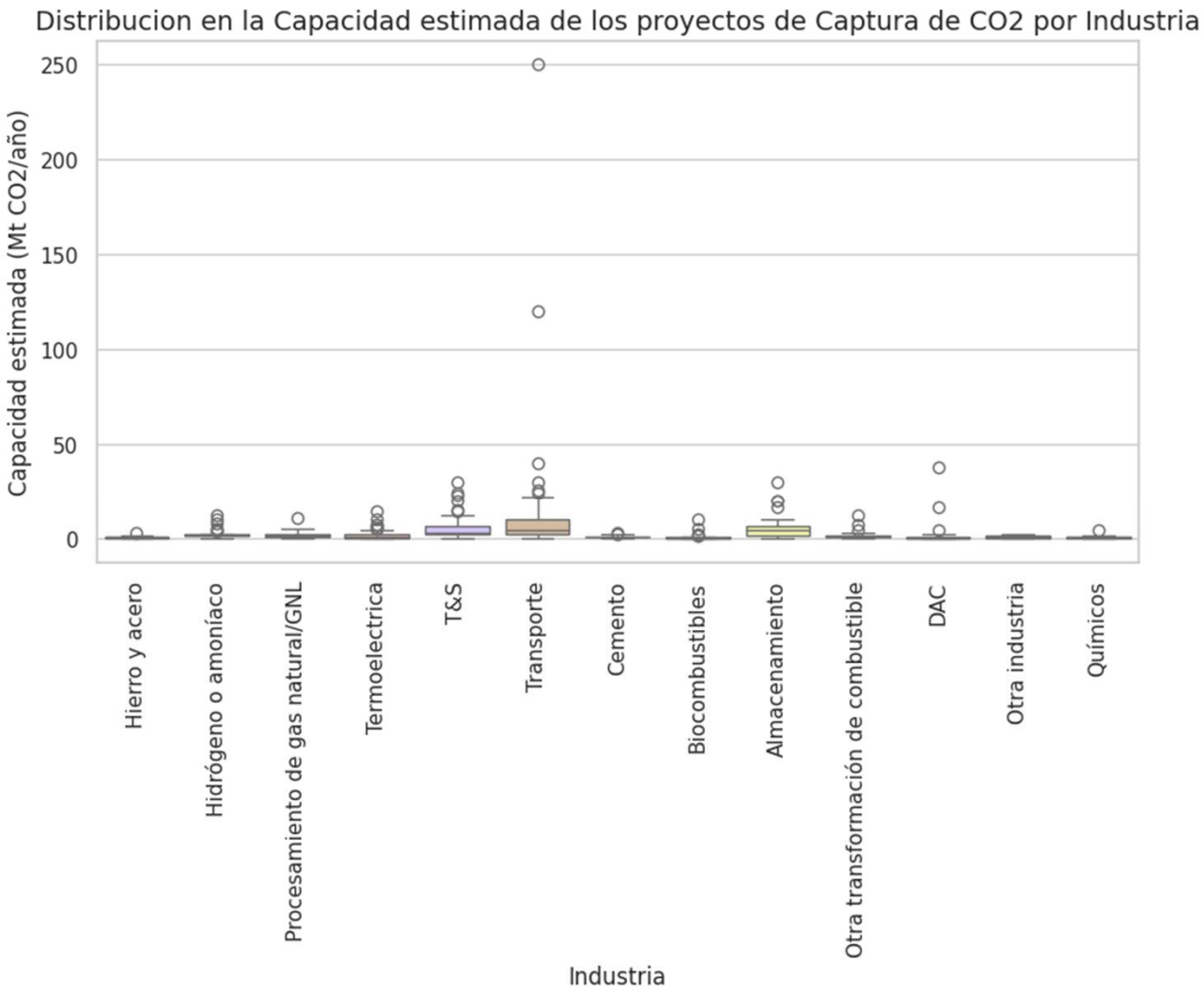
Data set 2
Box plot



Data set 2
Box plot



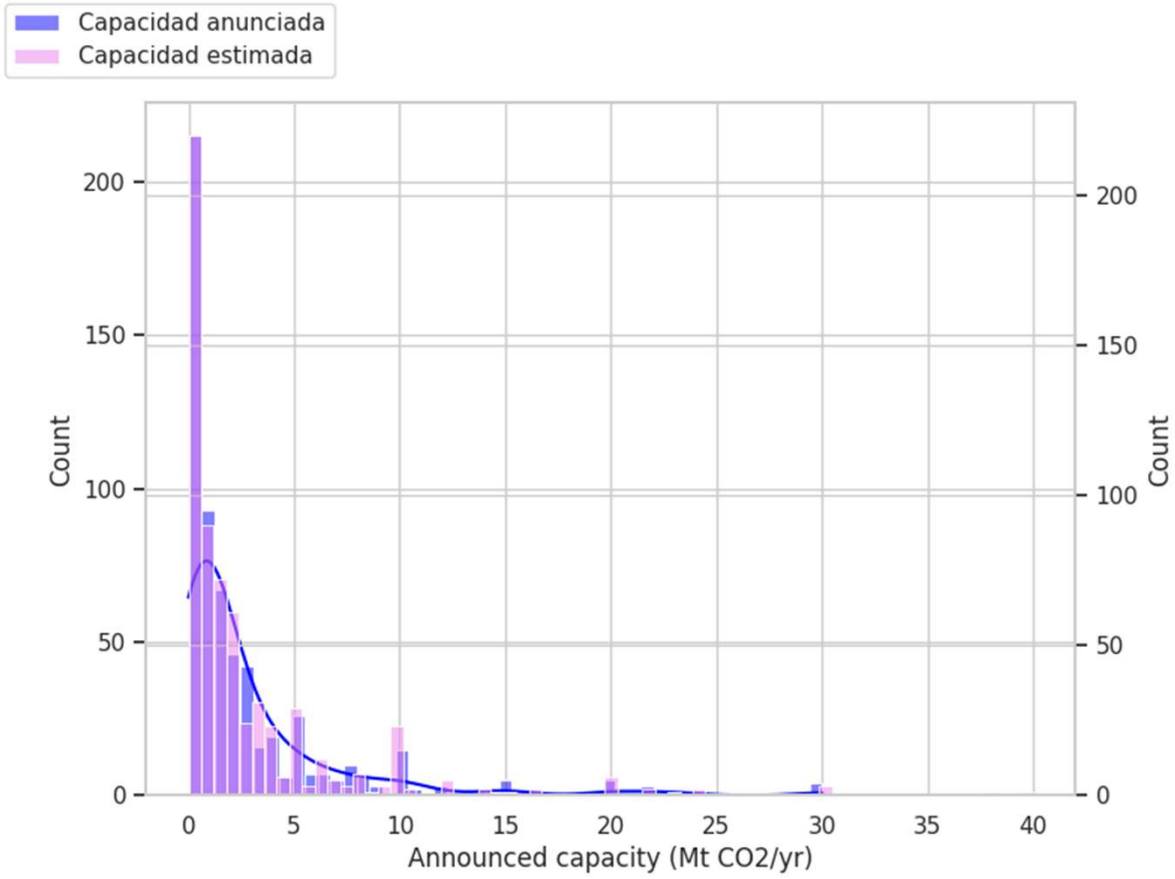
Data set 2 box plot



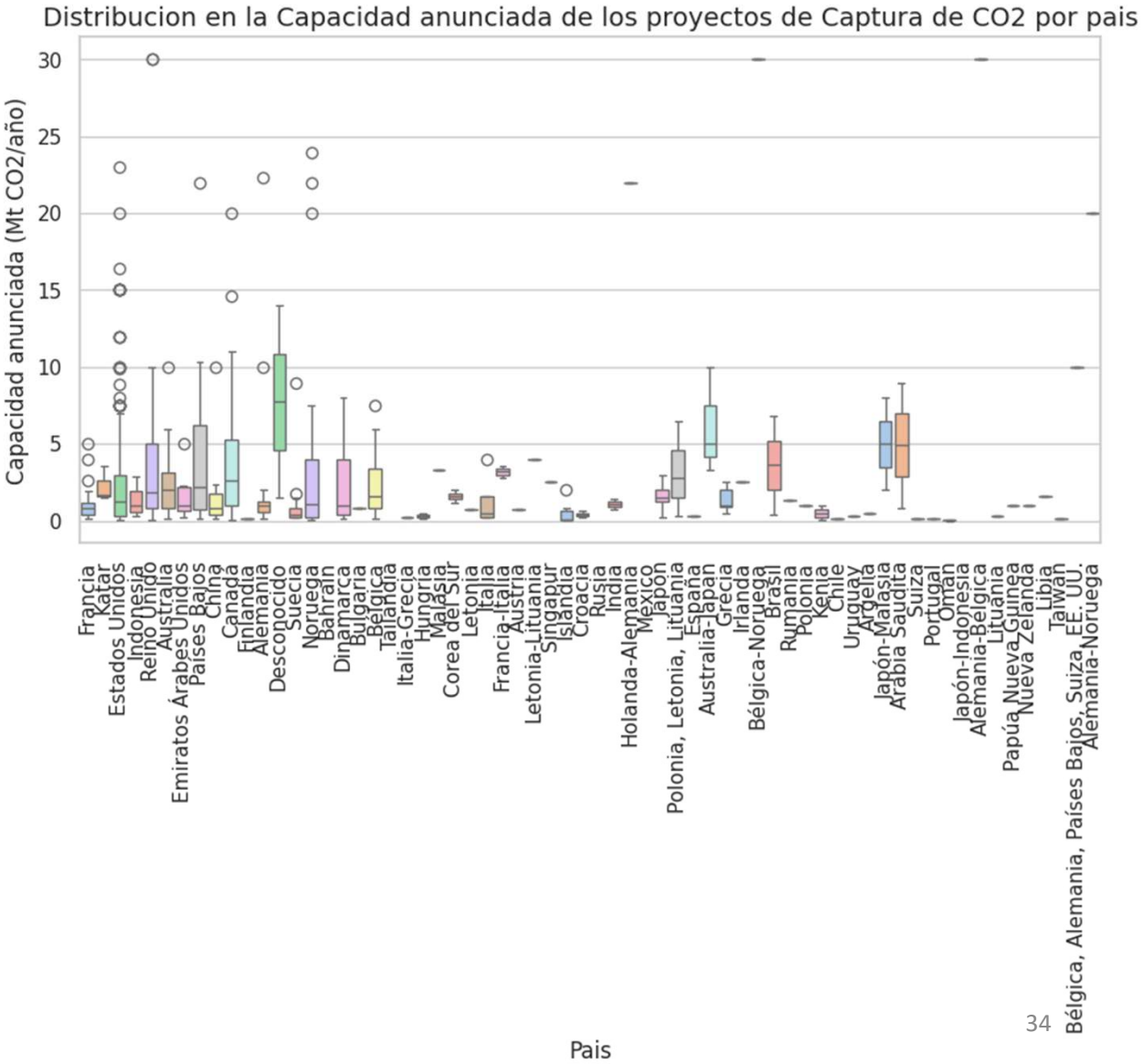
Data set 2 Estadísticos descriptivos sin valores atípicos

	Announcement	FID	Operation	Suspension/decommissioning	Project phase	Announced capacity (Mt CO2/yr)	Estimated capacity by IEA (Mt CO2/yr)
count	822	255	553	6	180.000000	612.000000	643.000000
mean	2020-10-26 00:56:03.503649536	2022-11-06 03:34:35.294117632	2026-03-05 22:47:05.316455680	2017-09-01 04:00:00	1.605556	2.796284	2.885266
min	1982-01-01 00:00:00	1984-01-01 00:00:00	1972-01-01 00:00:00	2004-01-01 00:00:00	1.000000	0.000000	0.001000
25%	2021-01-01 00:00:00	2023-01-01 00:00:00	2025-01-01 00:00:00	2012-07-02 00:00:00	1.000000	0.400000	0.400000
50%	2022-01-01 00:00:00	2024-01-01 00:00:00	2027-01-01 00:00:00	2020-01-01 12:00:00	1.000000	1.200000	1.270000
75%	2023-01-01 00:00:00	2024-01-01 00:00:00	2029-01-01 00:00:00	2024-07-02 06:00:00	2.000000	3.000000	3.000000
max	2024-01-01 00:00:00	2029-01-01 00:00:00	2040-01-01 00:00:00	2026-01-01 00:00:00	4.000000	30.000000	40.000000
std	NaN	NaN	NaN	NaN	0.705107	4.425264	4.768221

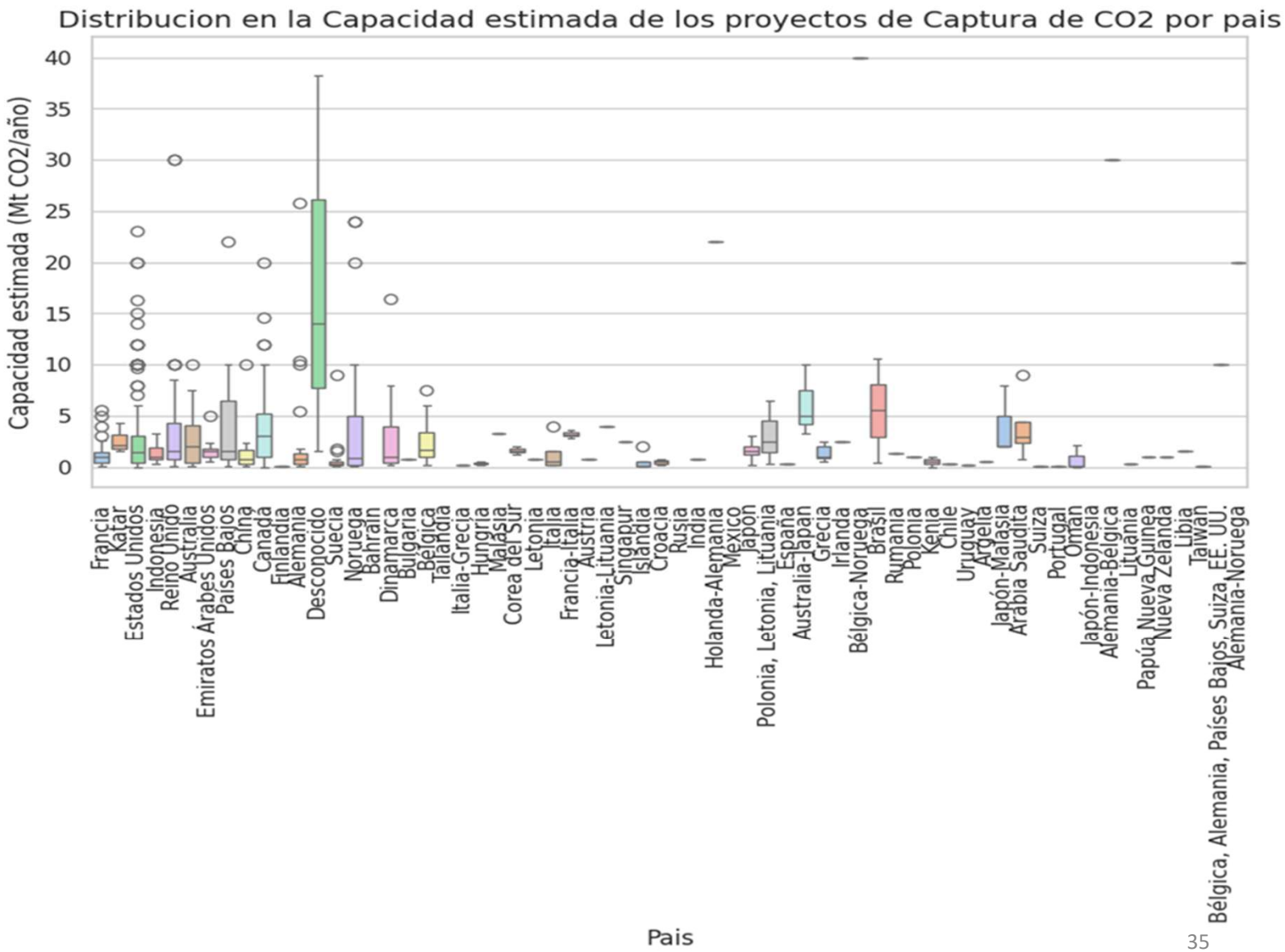
Data set 2
Histograma sin valores atípicos



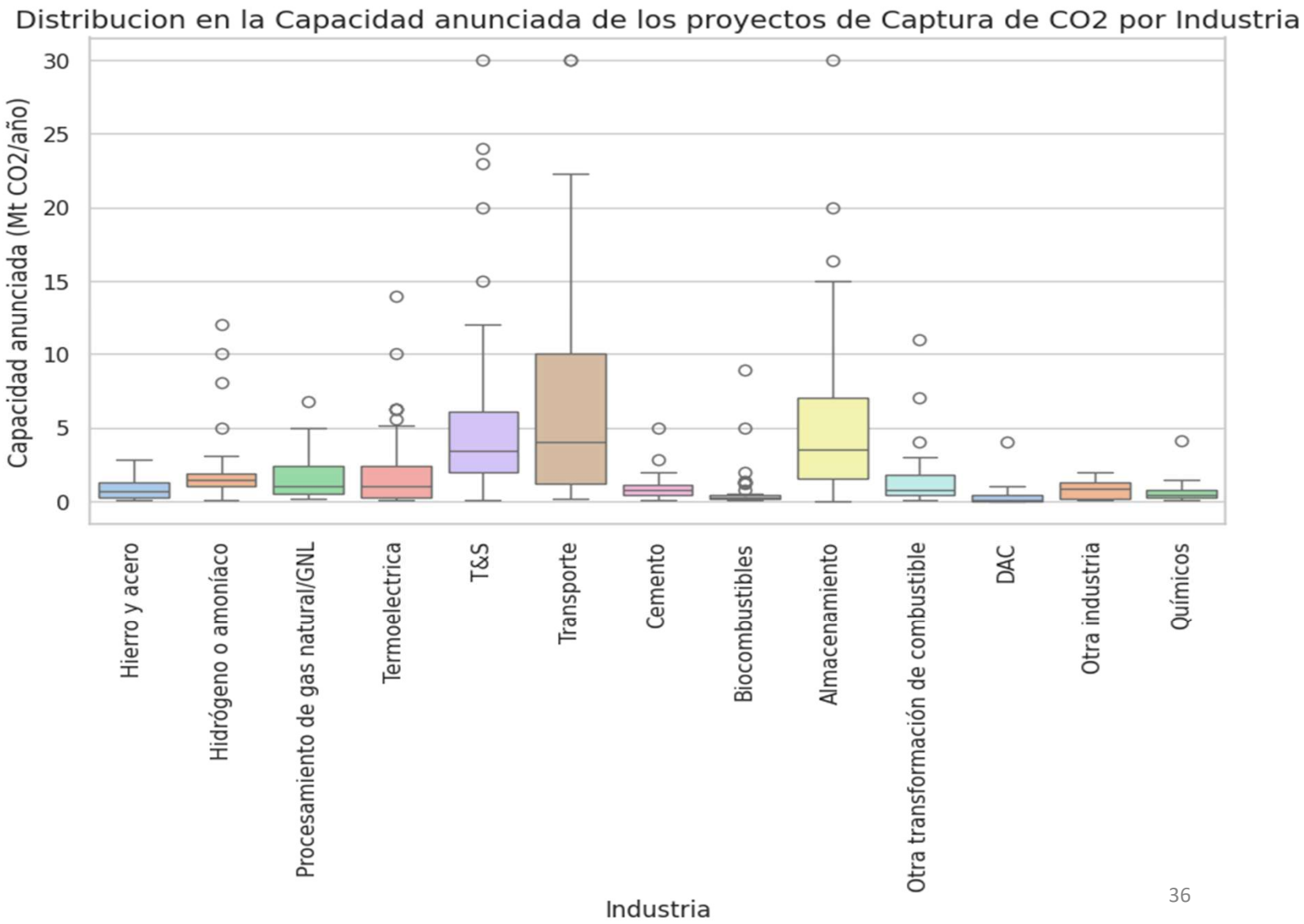
Data set 2
Box Plot sin valores atípicos



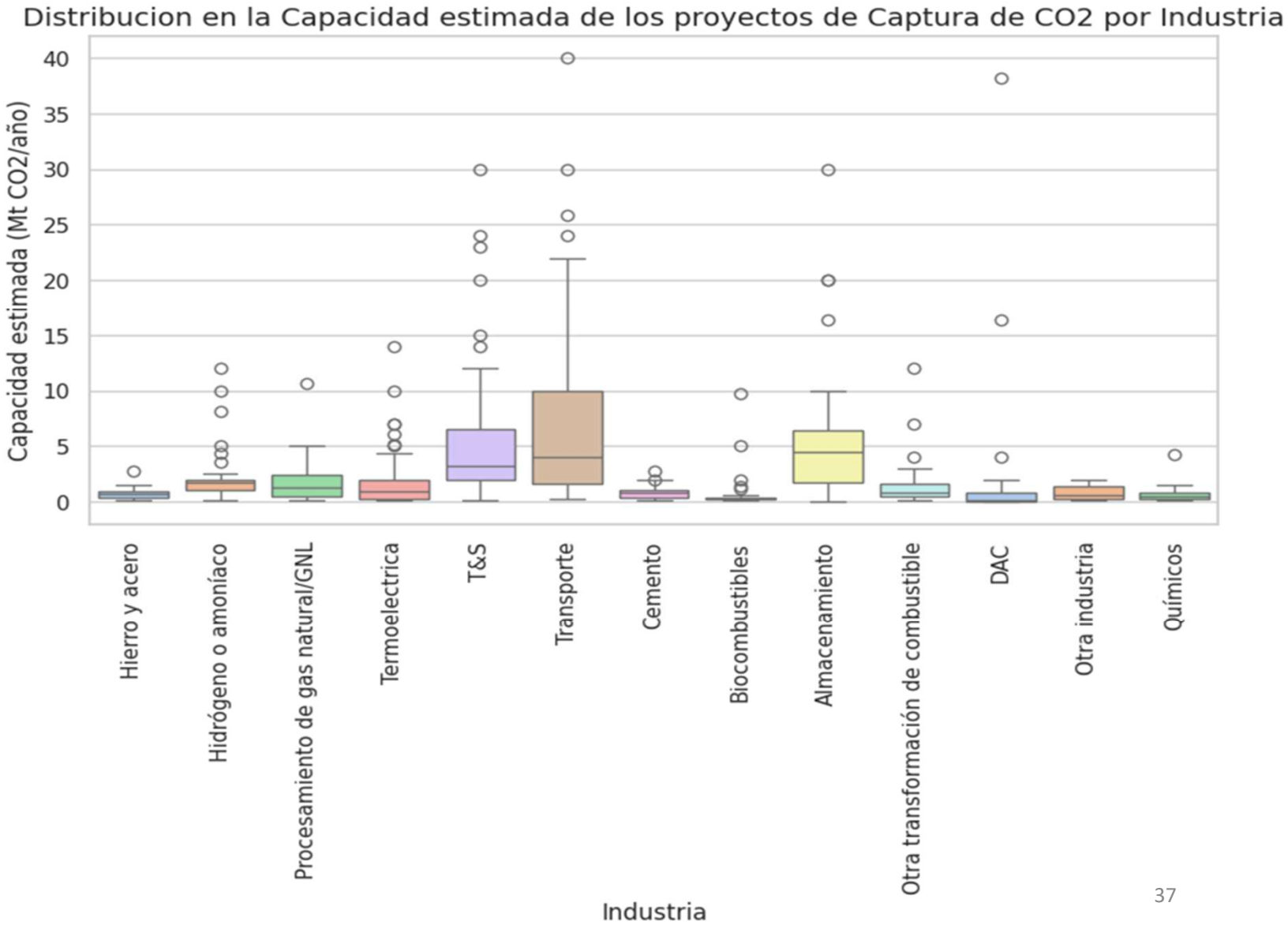
Data set 2 Box Plot sin valores atípicos



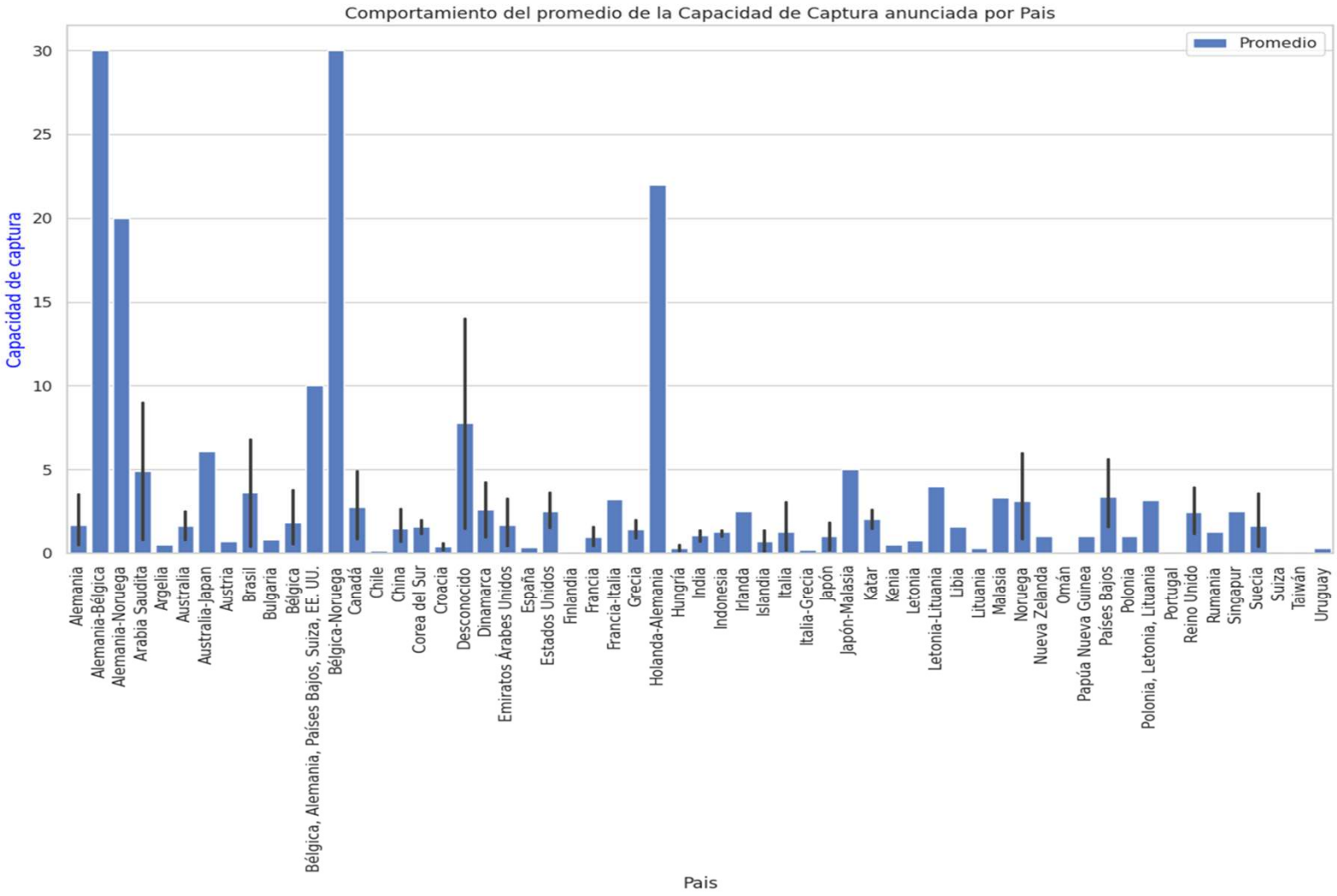
Data set 2 Box Plot sin valores atípicos



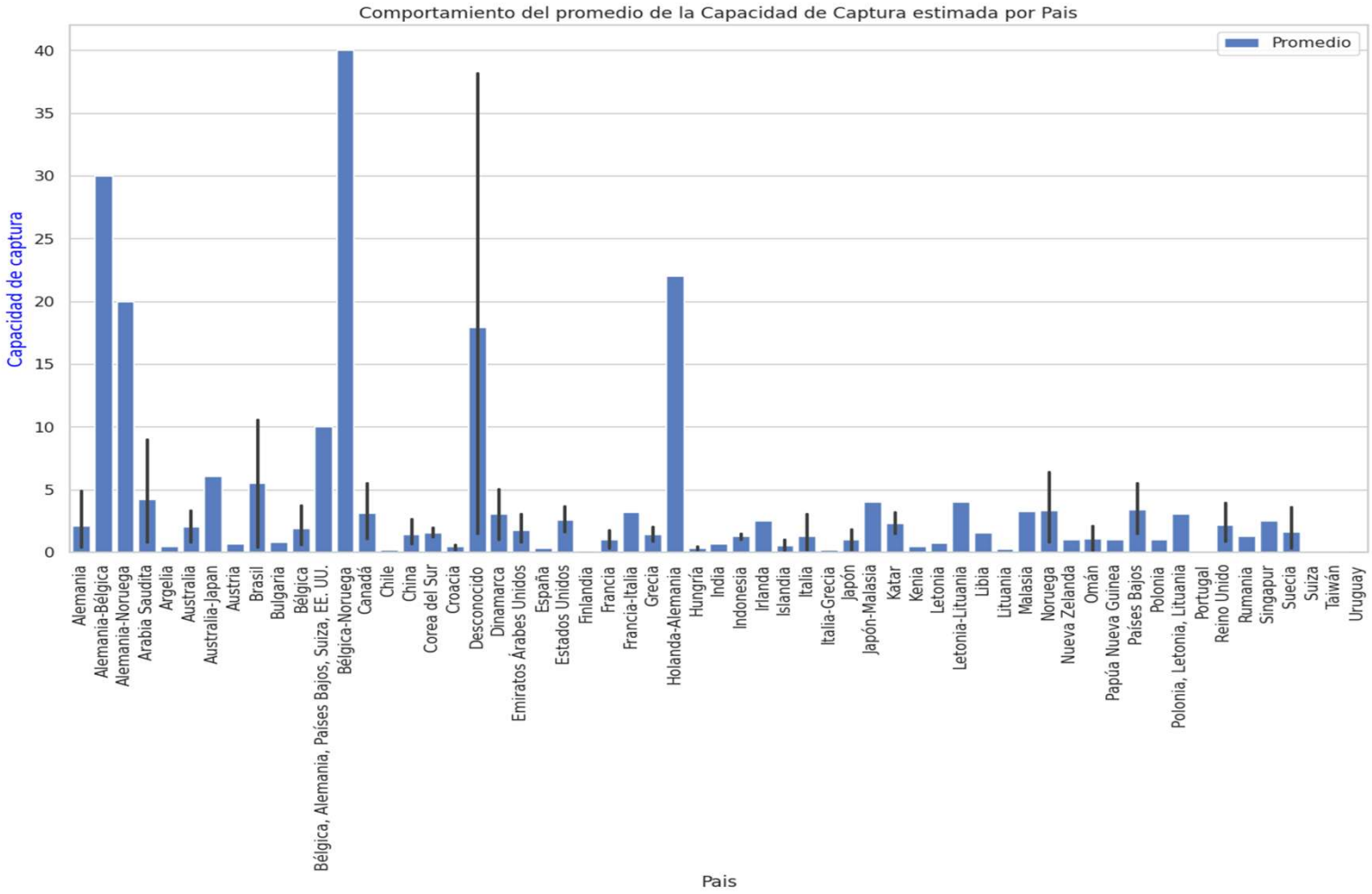
Data set 2 Box Plot sin valores atípicos



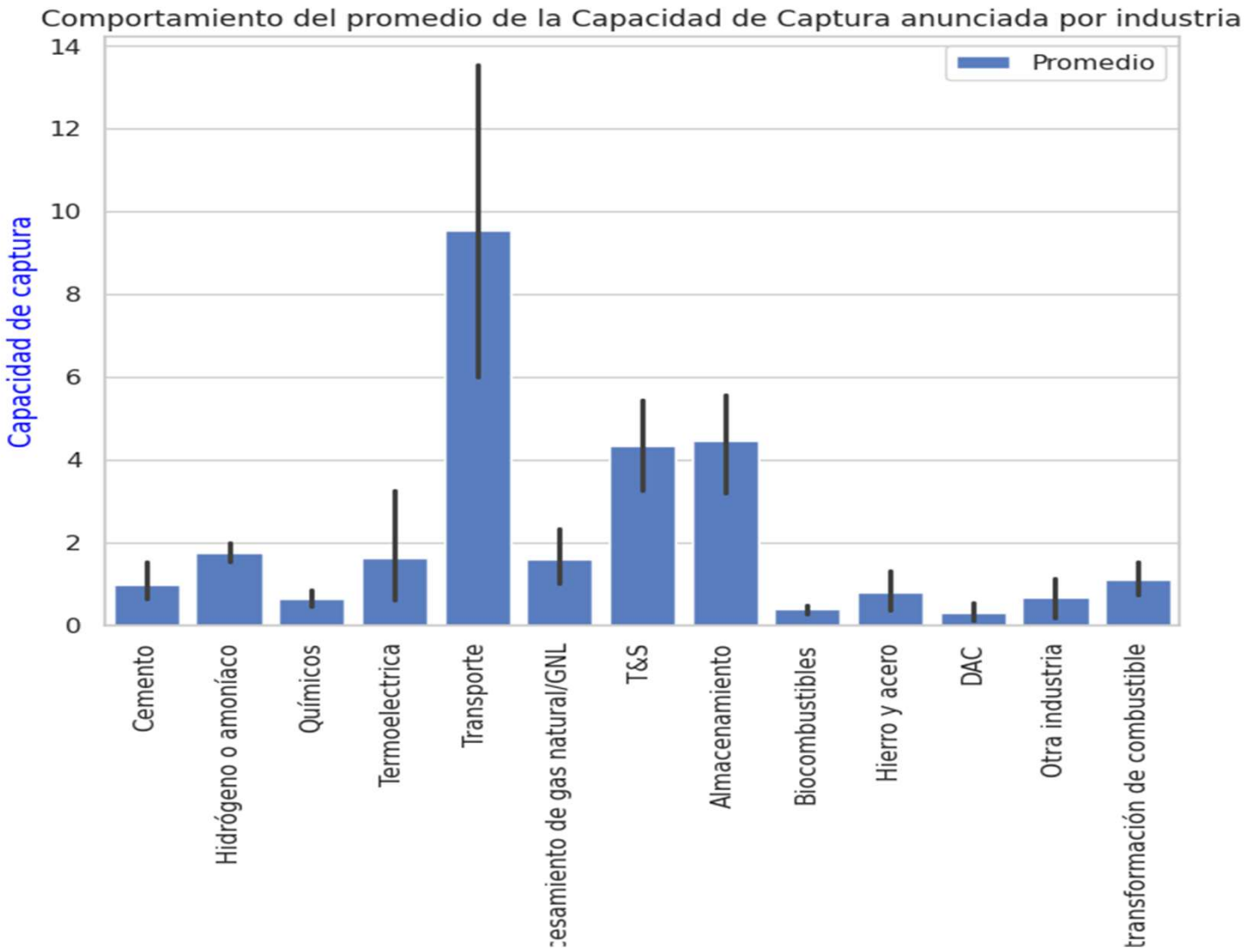
Data set 2


Comportamiento del promedio de la Capacidad de Captura anunciada por Pais
Promedio
Capacidad de captura
0
5
10
15
20
25
30
Alemania
Alemania-Bélgica
Alemania-Noruega
Arabia Saudita
Argelia
Australia
Australia-Japan
Austria
Brasil
Bulgaria
Bélgica
Bélgica, Alemania, Países Bajos, Suiza, EE. UU.
Bélgica-Noruega
Canadá
Chile
China
Corea del Sur
Croacia
Desconocido
Dinamarca
Emiratos Árabes Unidos
España
Estados Unidos
Finlandia
Francia
Francia-Italia
Grecia
Holanda-Alemania
Hungría
India
Indonesia
Irlanda
Islandia
Italia
Italia-Grecia
Japón
Japón-Malasia
Katar
Kenia
Letonia
Letonia-Lituania
Libia
Lituania
Malasia
Noruega
Nueva Zelanda
Omán
Papúa Nueva Guinea
Países Bajos
Polonia
Polonia, Letonia, Lituania
Portugal
Reino Unido
Rumania
Singapur
Suecia
Suiza
Taiwán
Uruguay
Pais

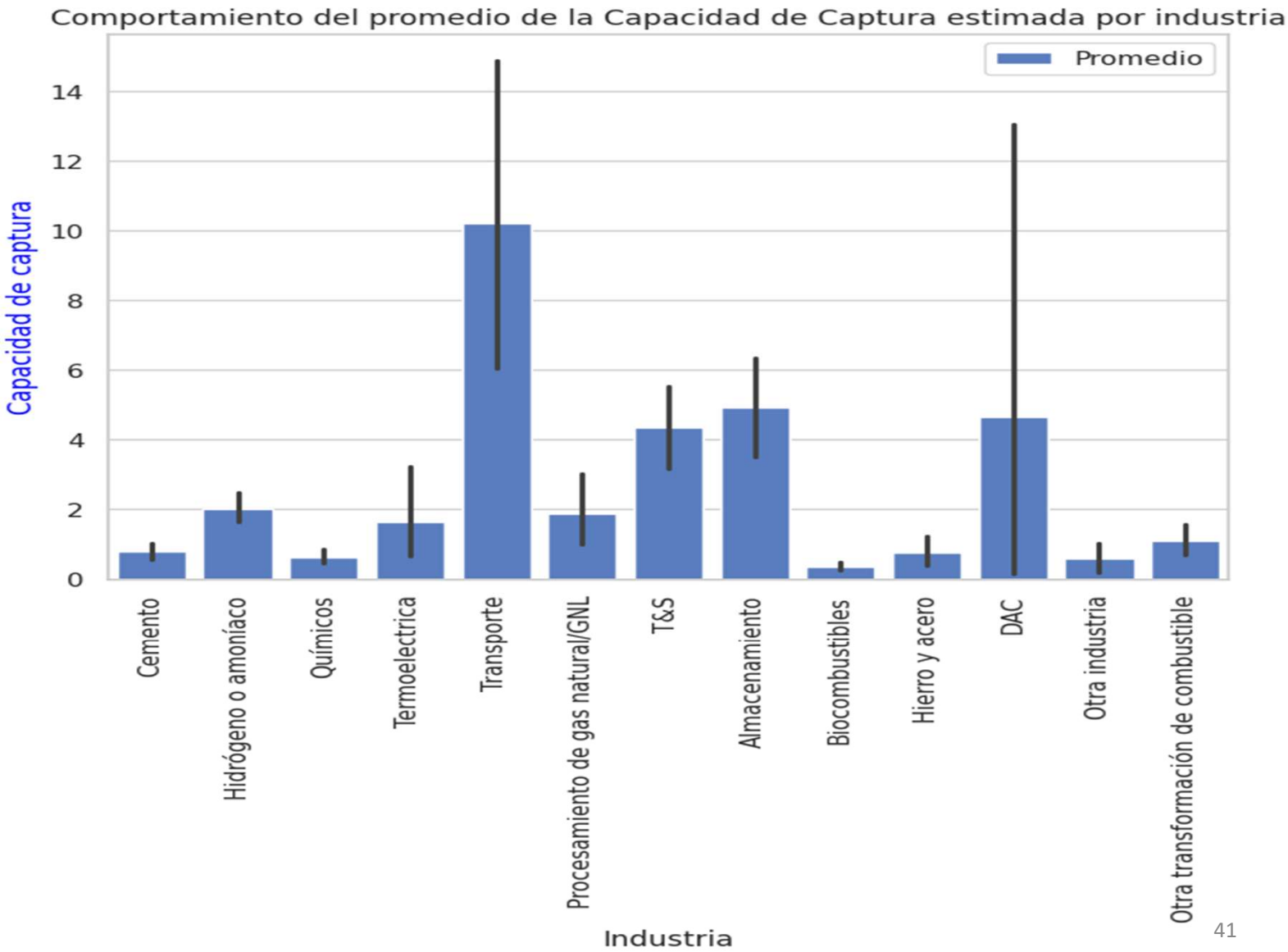
Data set 2


Comportamiento del promedio de la Capacidad de Captura estimada por Pais
Promedio
Capacidad de captura
0
5
10
15
20
25
30
35
40
Alemania
Alemania-Bélgica
Alemania-Noruega
Arabia Saudita
Argelia
Australia
Australia-Japan
Austria
Brasil
Bulgaria
Bélgica
Bélgica, Alemania, Países Bajos, Suiza, EE. UU.
Bélgica-Noruega
Canadá
Chile
China
Corea del Sur
Croacia
Desconocido
Dinamarca
Emiratos Árabes Unidos
España
Estados Unidos
Finlandia
Francia
Francia-Italia
Grecia
Holanda-Alemania
Hungría
India
Indonesia
Irlanda
Islandia
Italia
Italia-Grecia
Japón
Japón-Malasia
Katar
Kenia
Letonia
Letonia-Lituania
Libia
Lituania
Malasia
Noruega
Nueva Zelanda
Omán
Papúa Nueva Guinea
Países Bajos
Polonia
Polonia, Letonia, Lituania
Portugal
Reino Unido
Rumania
Singapur
Suecia
Suiza
Taiwán
Uruguay
Pais

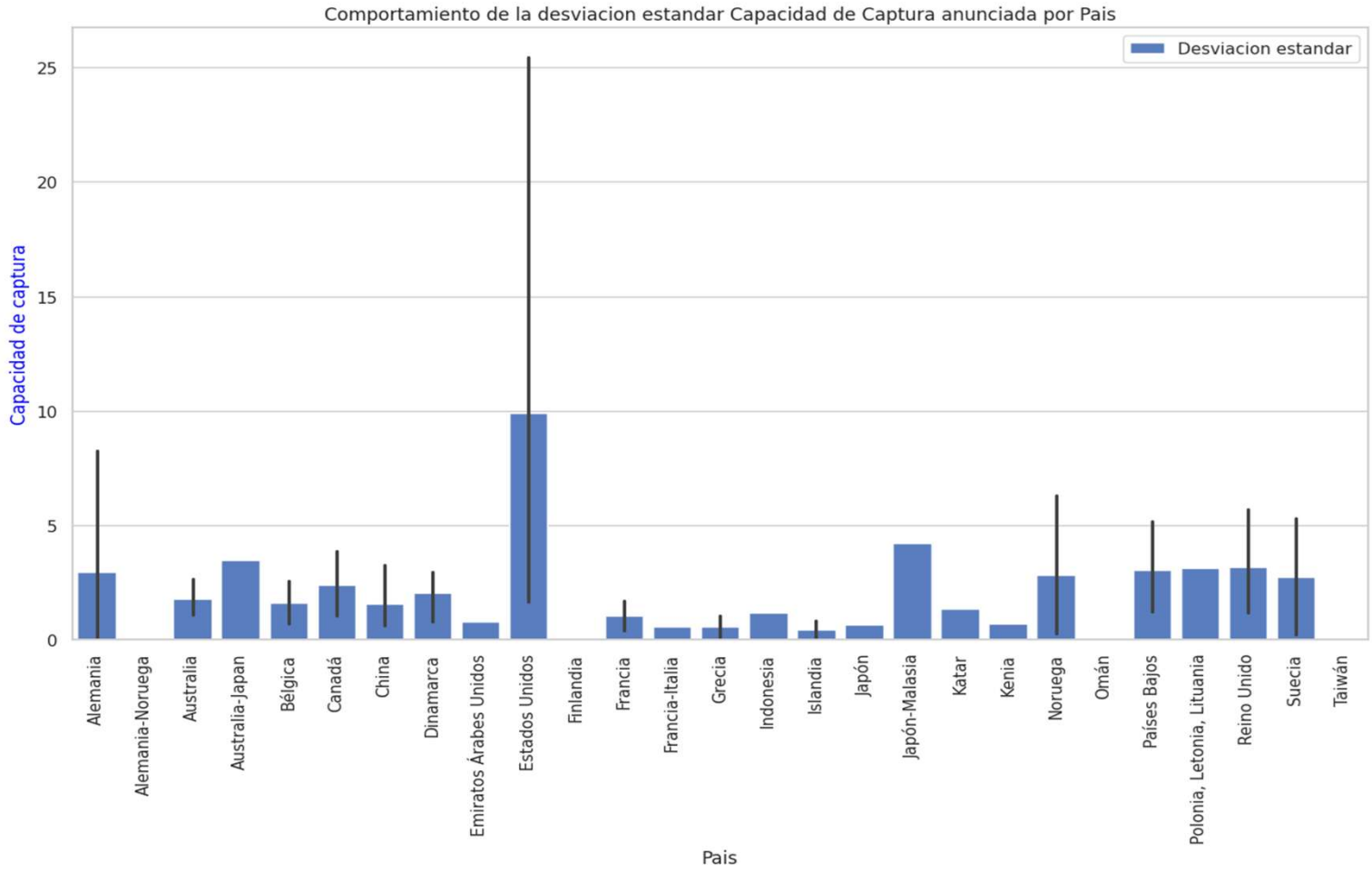
Data Set 2


Comportamiento del promedio de la Capacidad de Captura anunciada por industria
Promedio
Capacidad de captura
0
2
4
6
8
10
12
14
Cemento
Hidrógeno o amoníaco
Químicos
Termoelectrica
Transporte
esamiento de gas natural/GNL
T&S
Almacenamiento
Biocombustibles
Hierro y acero
DAC
Otra industria
transformación de combustible

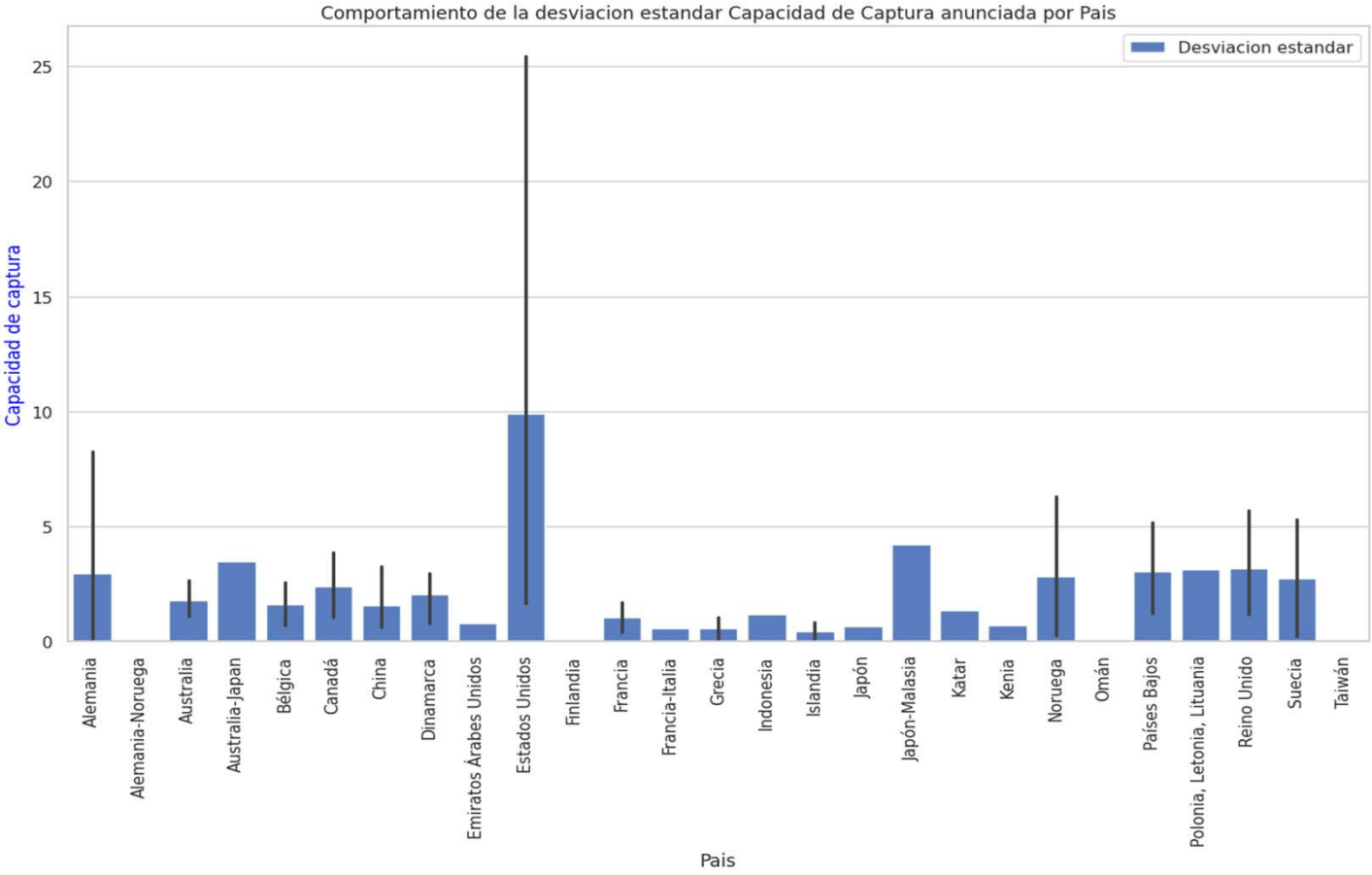
Data set 2


Comportamiento del promedio de la Capacidad de Captura estimada por industria
Promedio
Capacidad de captura
0
2
4
6
8
10
12
14
Cemento
Hidrógeno o amoníaco
Químicos
Termoelectrica
Transporte
Procesamiento de gas natural/GNL
T&S
Almacenamiento
Biocombustibles
Hierro y acero
DAC
Otra industria
Otra transformación de combustible
Industria

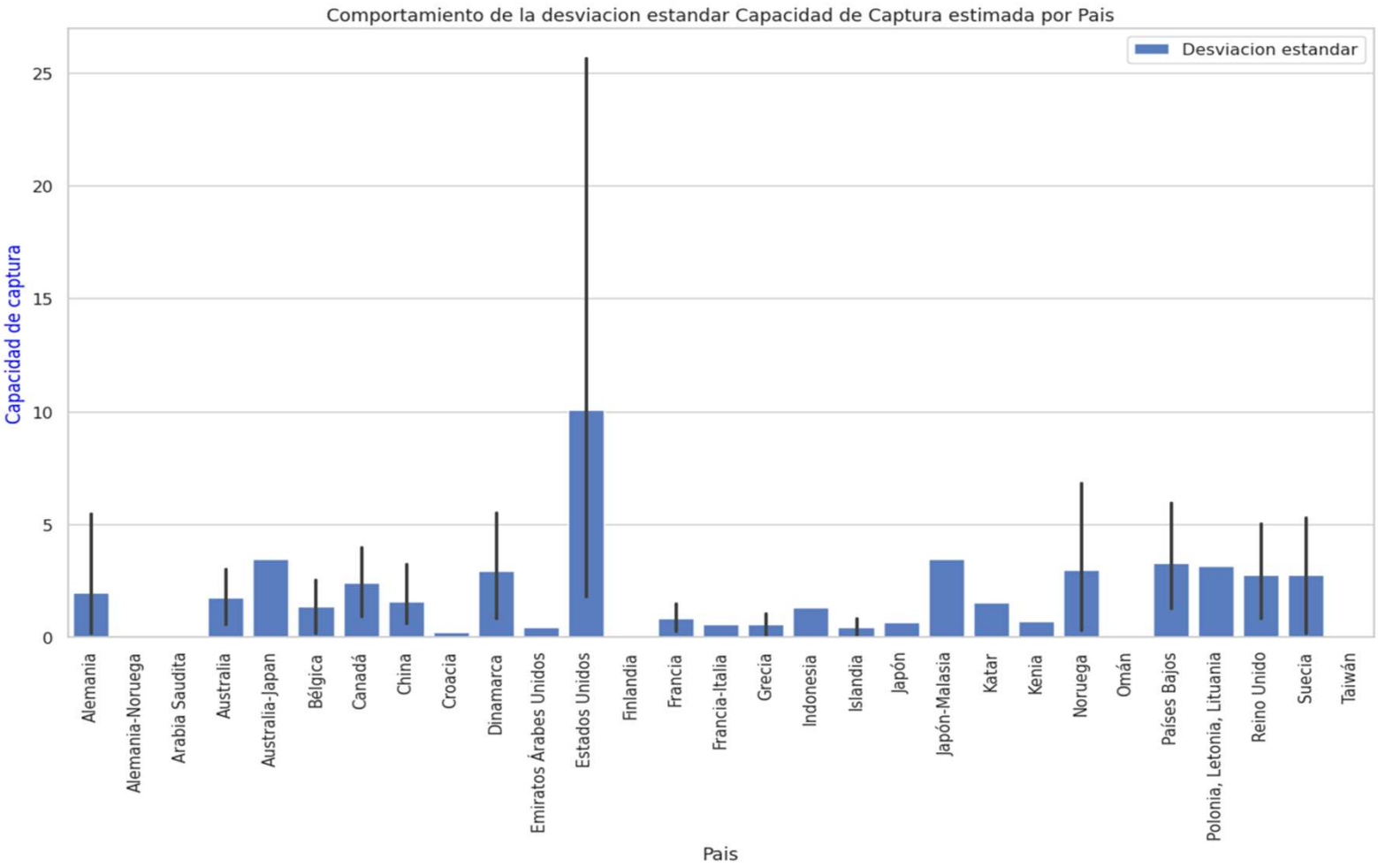
Data set 2


Comportamiento de la desviacion estandar Capacidad de Captura anunciada por Pais
Desviacion estandar
Capacidad de captura
0
5
10
15
20
25
Alemania
Alemania-Noruega
Australia
Australia-Japan
Bélgica
Canadá
China
Dinamarca
Emiratos Árabes Unidos
Estados Unidos
Finlandia
Francia
Francia-Italia
Grecia
Indonesia
Islandia
Japón
Japón-Malasia
Katar
Kenia
Noruega
Omán
Países Bajos
Polonia, Letonia, Lituania
Reino Unido
Suecia
Taiwán
Pais

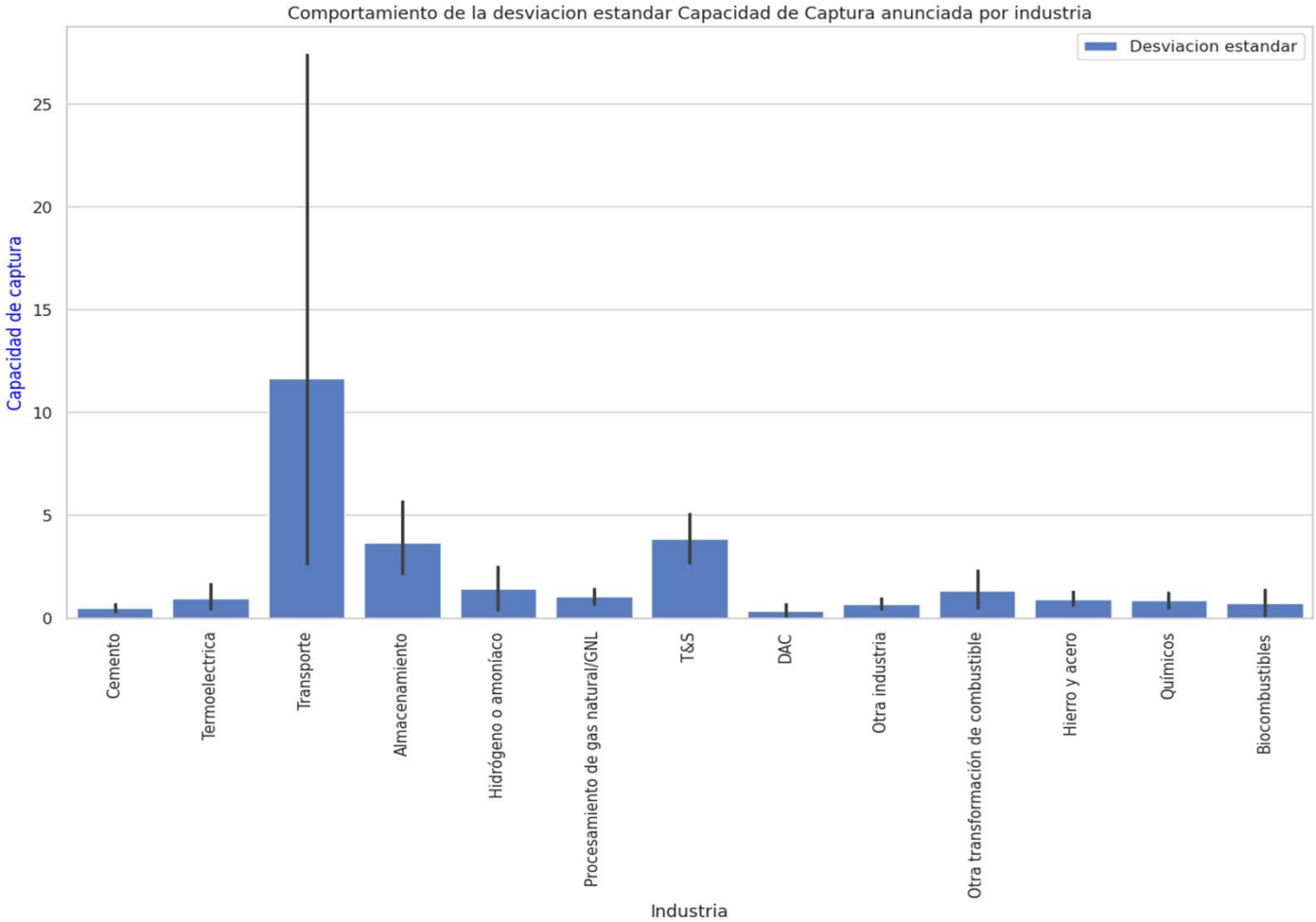
Data set 2


Comportamiento de la desviacion estandar Capacidad de Captura anunciada por Pais
Desviacion estandar
Capacidad de captura
0
5
10
15
20
25
Alemania
Alemania-Noruega
Australia
Australia-Japan
Bélgica
Canadá
China
Dinamarca
Emiratos Árabes Unidos
Estados Unidos
Finlandia
Francia
Francia-Italia
Grecia
Indonesia
Islandia
Japón
Japón-Malasia
Katar
Kenia
Noruega
Omán
Países Bajos
Polonia, Letonia, Lituania
Reino Unido
Suecia
Taiwán
Pais

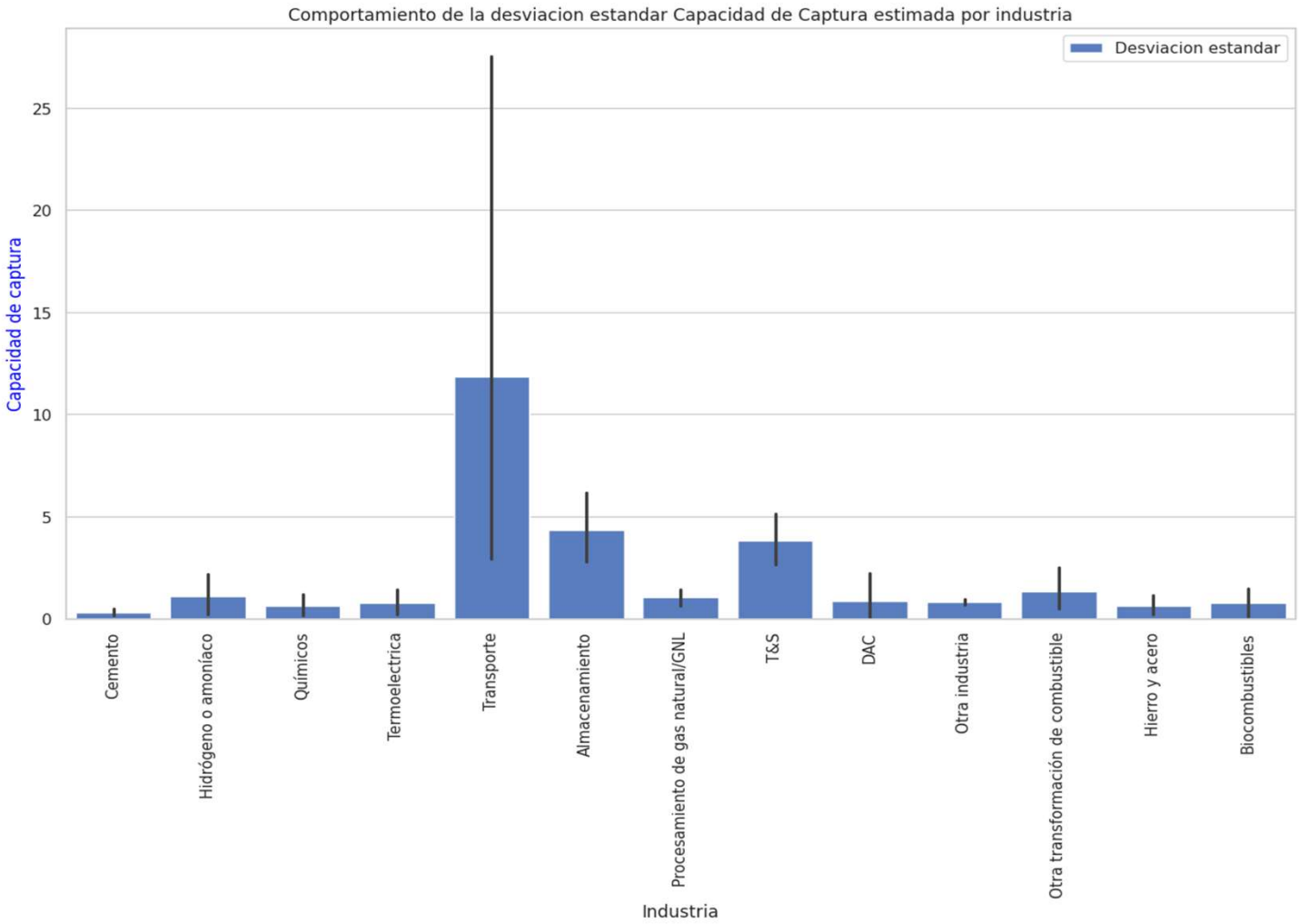
Data set 2


Comportamiento de la desviacion estandar Capacidad de Captura estimada por Pais
Desviacion estandar
Capacidad de captura
0
5
10
15
20
25
Alemania
Alemania-Noruega
Arabia Saudita
Australia
Australia-Japan
Bélgica
Canadá
China
Croacia
Dinamarca
Emiratos Árabes Unidos
Estados Unidos
Finlandia
Francia
Francia-Italia
Grecia
Indonesia
Islandia
Japón
Japón-Malasia
Katar
Kenia
Noruega
Omán
Países Bajos
Polonia, Letonia, Lituania
Reino Unido
Suecia
Taiwán
Pais

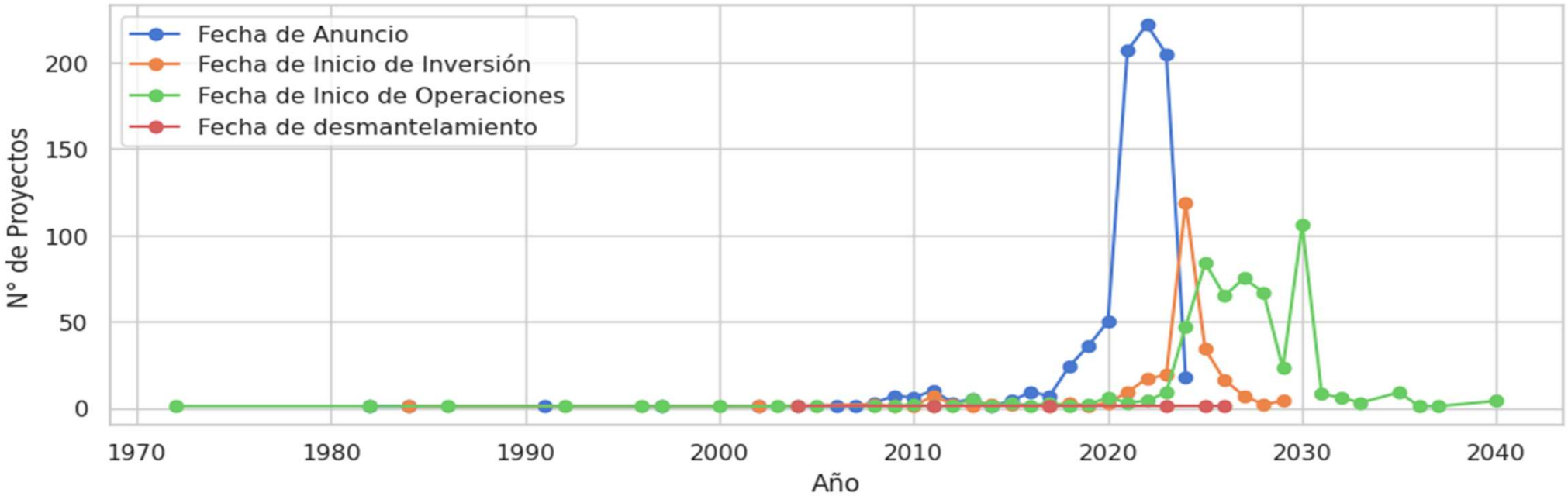
Data set 2 Diagrama de barras con barras de error



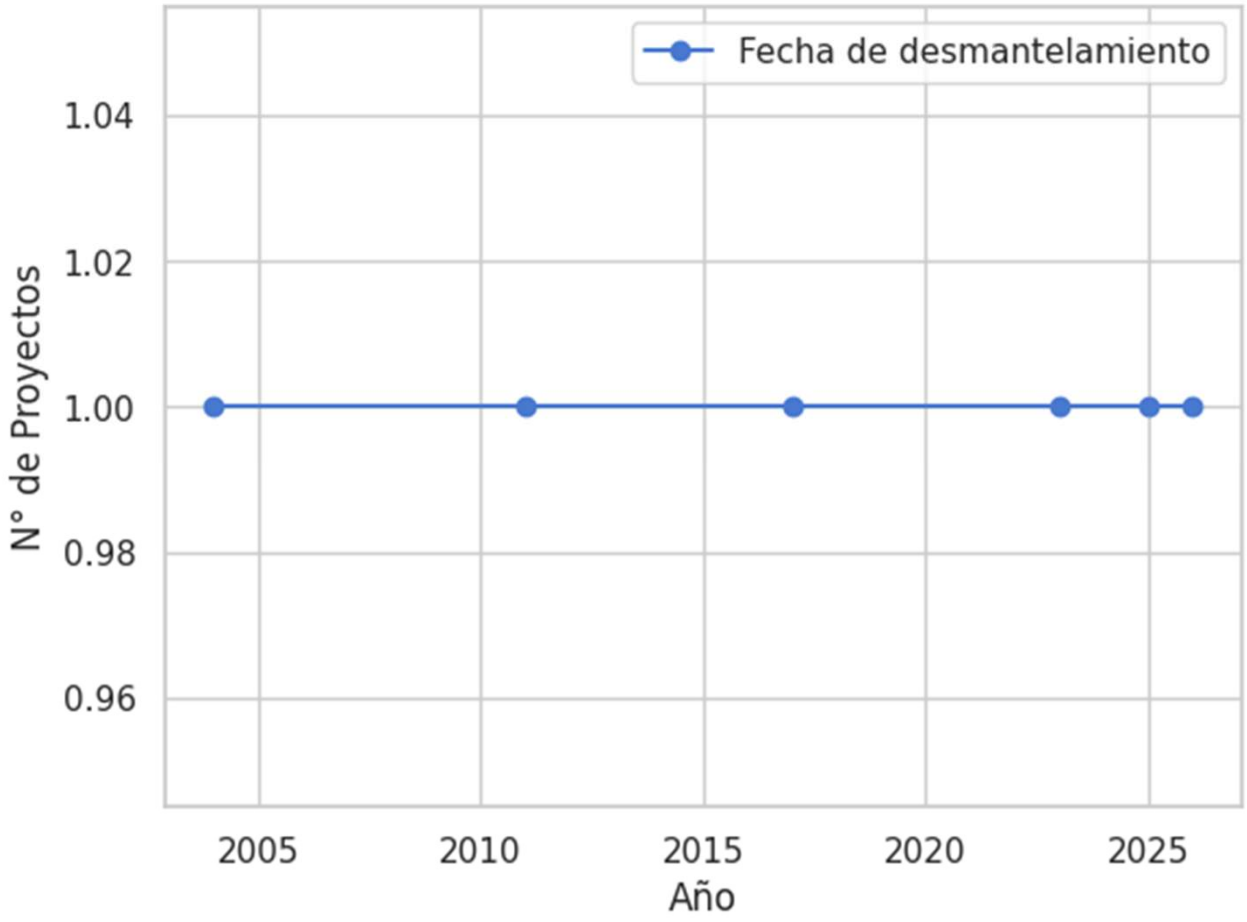
Data set 2 Diagrama de barras con barras de error



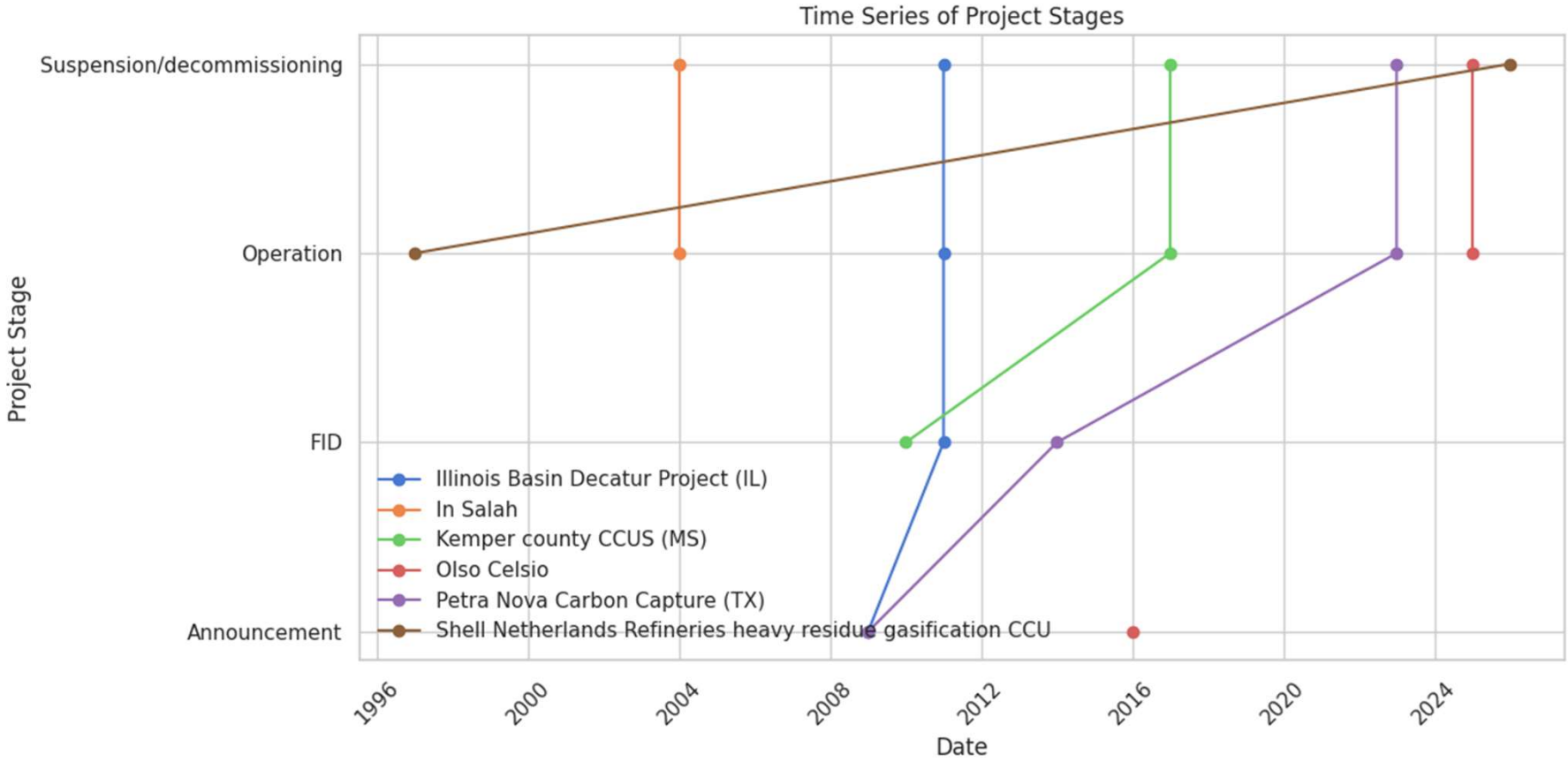
Data set 2 Serie de tiempo



Data set 2


Fecha de desmantelamiento
1.04
1.02
1.00
0.98
0.96
N° de Proyectos
2005
2010
2015
2020
2025
Año

Data set 2


Time Series of Project Stages
Suspension/decommissioning
Operation
FID
Announcement
Project Stage
Illinois Basin Decatur Project (IL)
In Salah
Kemper county CCUS (MS)
Olso Celsio
Petra Nova Carbon Capture (TX)
Shell Netherlands Refineries heavy residue gasification CCU
1996
2000
2004
2008
2012
2016
2020
2024
Date

Data set 2
Exploración para el modelamiento del PCA

Antes de la limpieza de valores atípicos

```
promedio de Capacidad anunciada, Varianza de capacidad anunciada; [np.float64(3.3897817589576547), 141.26476983483792]
promedio de Capacidad estimada, Varianza de Capacidad estimada:  [np.float64(3.4499627906976746), 138.46636432470035]
```

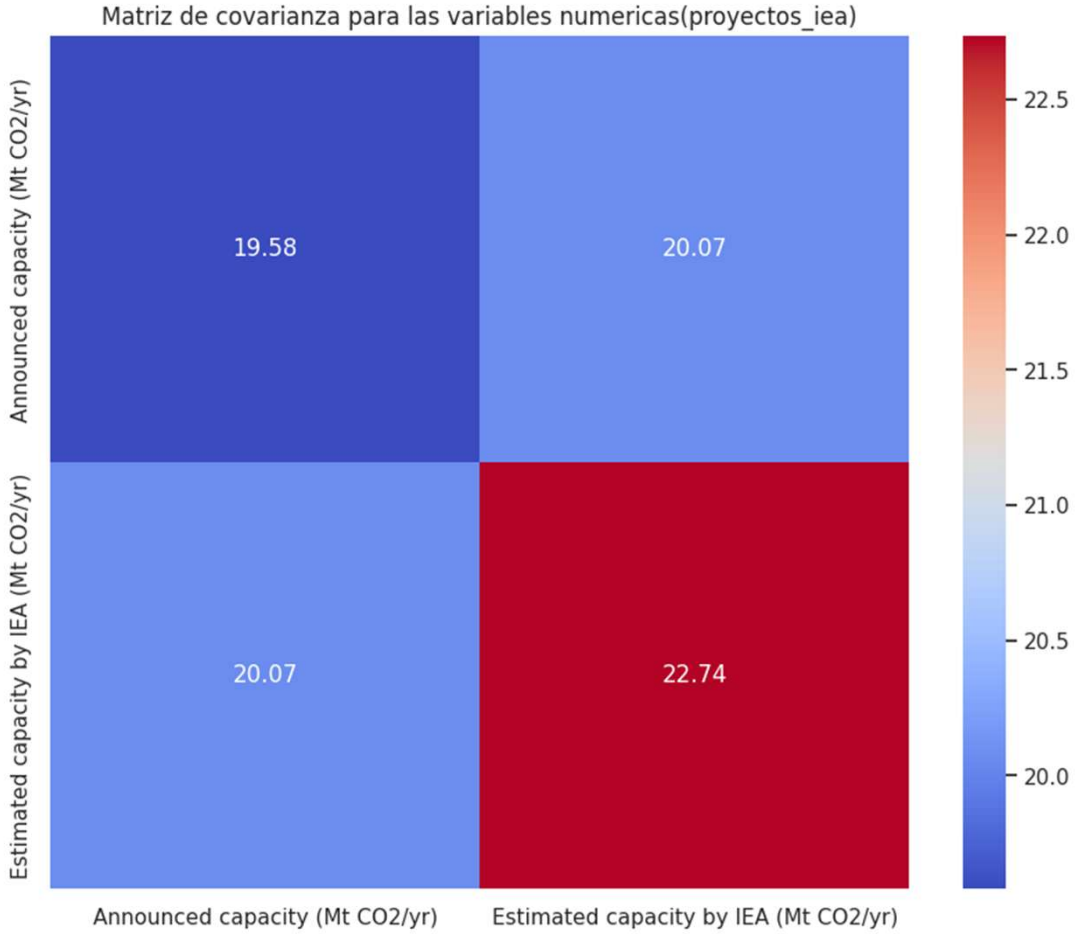
Luego de la limpieza de valores atípicos

```
promedio de Capacidad anunciada, Varianza de capacidad anunciada; [np.float64(2.79628431372549), 19.58296361461439]
promedio de Capacidad estimada, Varianza de Capacidad estimada:  [np.float64(2.885265940902022), 22.735927002373984]
```

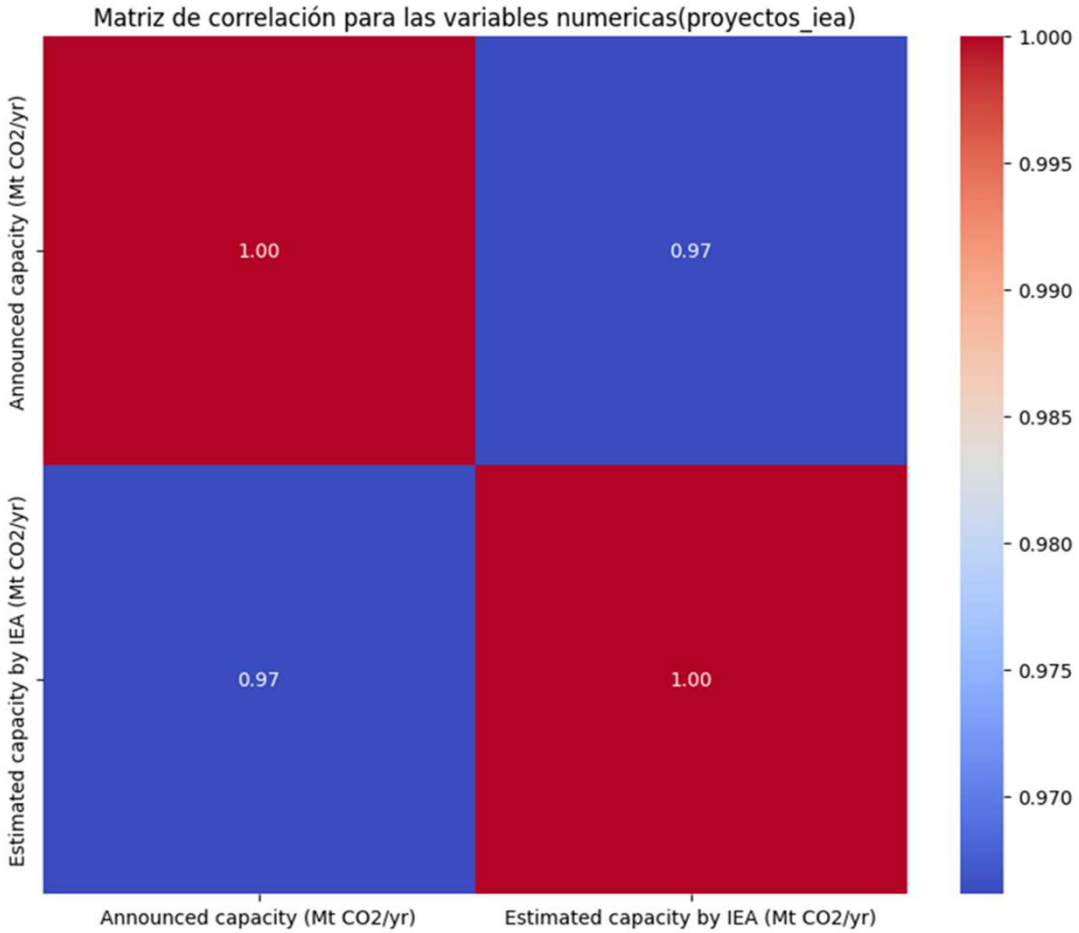
Data set 2
Matriz de Covarianza

	Announced capacity (Mt CO2/yr)	Estimated capacity by IEA (Mt CO2/yr)
Announced capacity (Mt CO2/yr)	19.582964	20.067968
Estimated capacity by IEA (Mt CO2/yr)	20.067968	22.735927

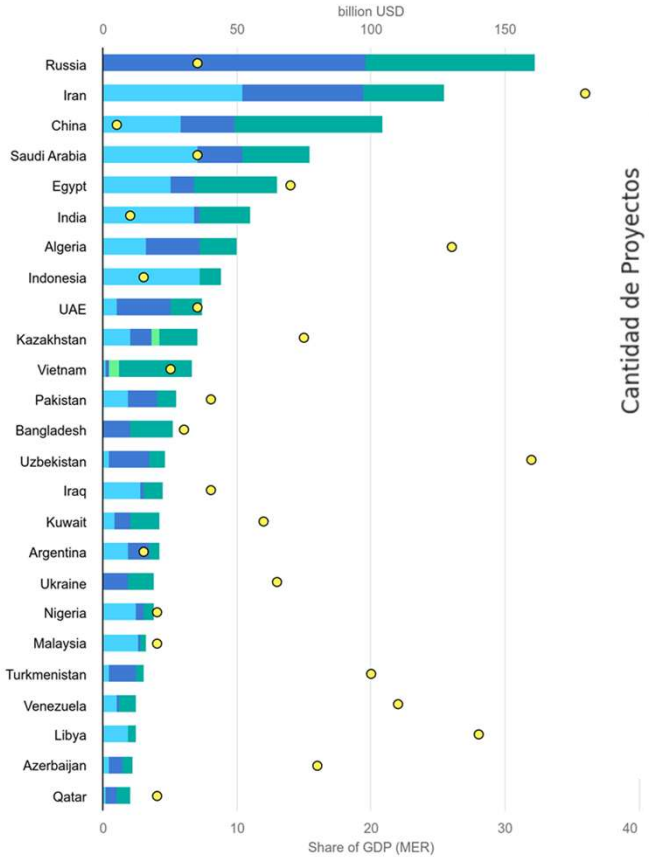
Data set 2
Matriz de Covarianza-Mapa de Calor

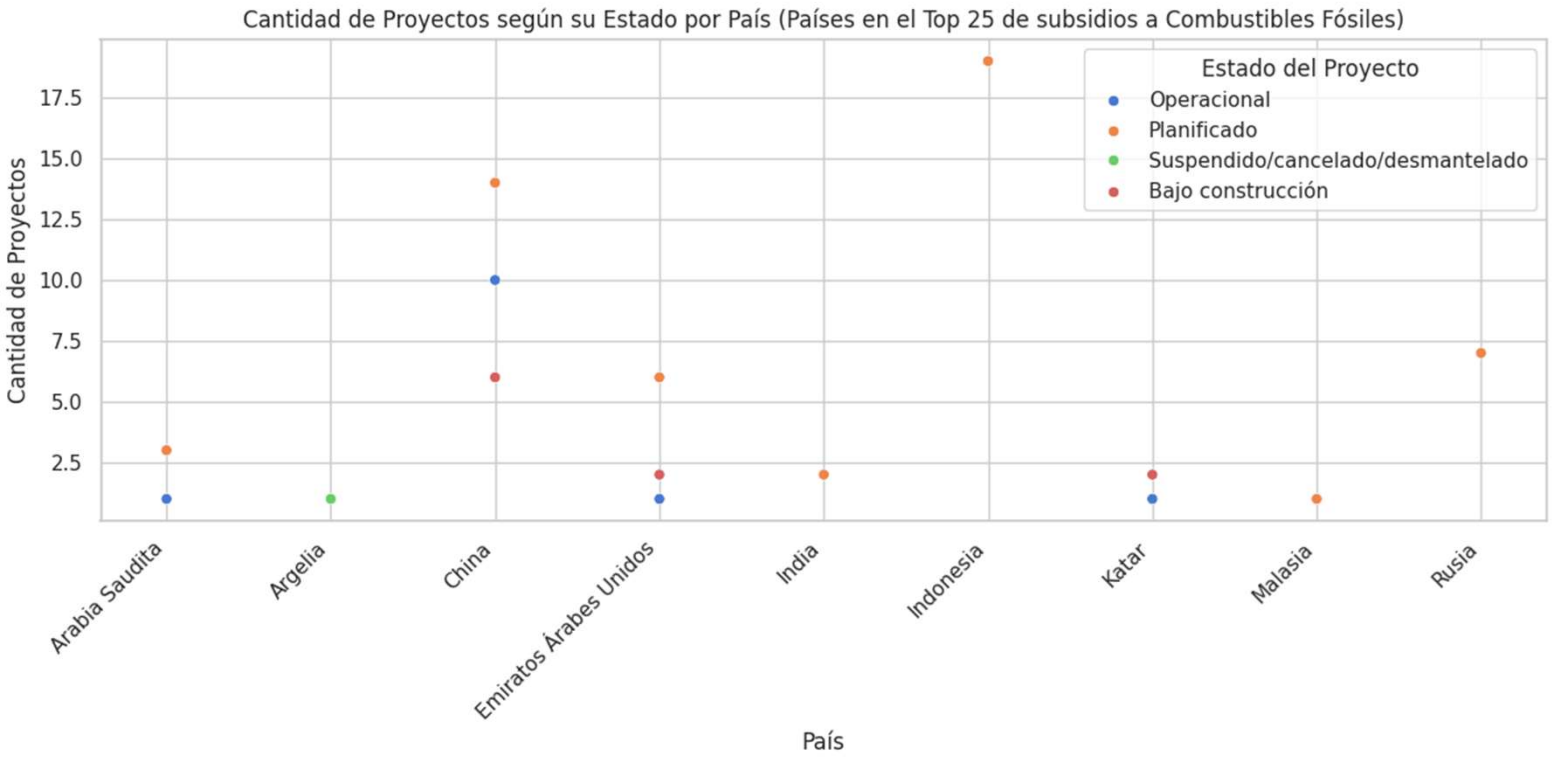


Data set 2
Matriz de Correlación-Mapa de Calor



Top 25 de Países con subsidios de Combustibles Fósiles





Value of fossil-fuel subsidies by fuel in the top 25 countries, 2022

Data set 3 : China

https://www.youtube.com/watch?v=LYSNER4Dnjs

Fuente del Data set 3

View PDF Download full issue

ELSEVIER

Journal of Cleaner Production
Volume 454, 15 May 2024, 142280

A preliminary assessment of CO_2 capture, transport, and storage network for China's steel sector

Yuancheng Lin [a b], Tubagus Aryandi Gunawan [c], Cecelia Isaac [c], Hongxi Luo [c], Fangwei Cheng [c], Eric D. Larson [c], Chris Greig [c], Linwei Ma [a b], Zheng Li [a b]

Show more

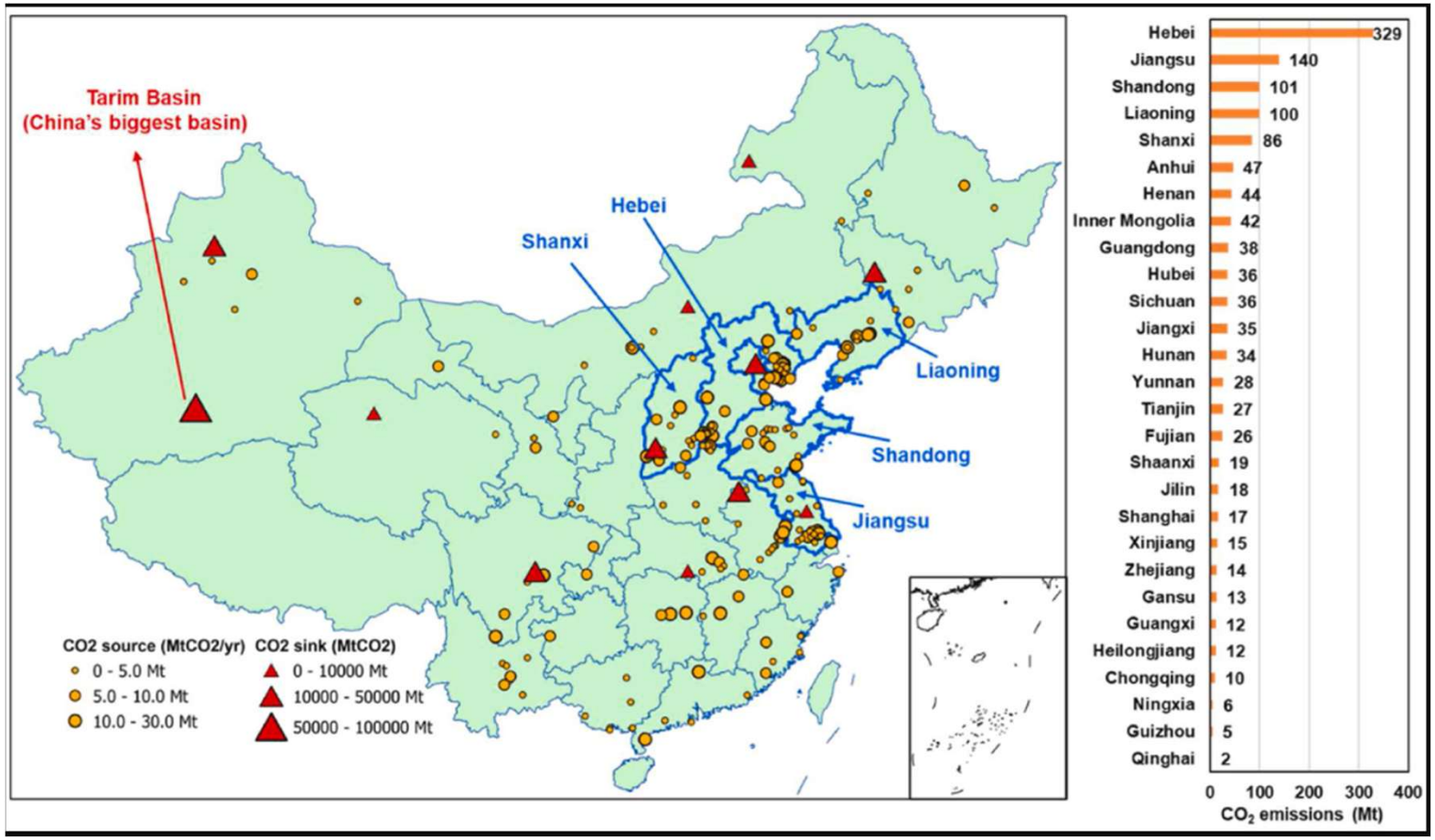
+ Add to Mendeley Share Cite

https://doi.org/10.1016/j.jclepro.2024.142280

Get rights and content

Full text access

China Data set 3



Data set 3

```
[4]  1 ## Descripcion de las variables numéricas
     2 df_China.describe()
```

	Captured CO2/Mt	Captured cost/($/t)
count	227.000000	227.000000
mean	5.690749	54.897357
std	4.946819	10.671026
min	0.200000	34.700000
25%	2.450000	46.800000
50%	4.000000	54.200000
75%	7.700000	60.750000
max	28.300000	104.200000

Data set 3

```
# Libreria Skimpy permite realizar estadistiscos descriptivos de variables categóricas
skim(df_China)
```

skimpy summary

Data Summary

Dataframe	Values
Number of rows	227
Number of columns	3

Data Types

Column Type	Count
float64	2
string	1

number

column	NA	NA %	mean	sd	p0	p25	p50	p75	p100	hist
Captured CO2/Mt	0	0	5.691	4.947	0.2	2.45	4	7.7	28.3	
Captured cost/($/t)	0	0	54.9	10.67	34.7	46.8	54.2	60.75	104.2	

string

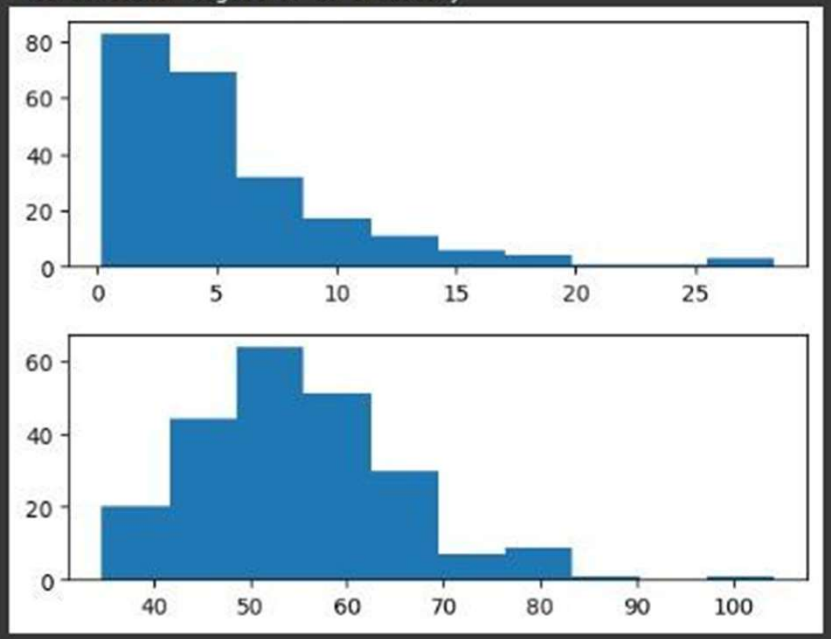
column	NA	NA %	shortest	longest	min	max	chars per row	words per row	total words
Province	0	0	Anhui	Inner Mongolia	Anhui	Zhejiang	6.67	1	236

End

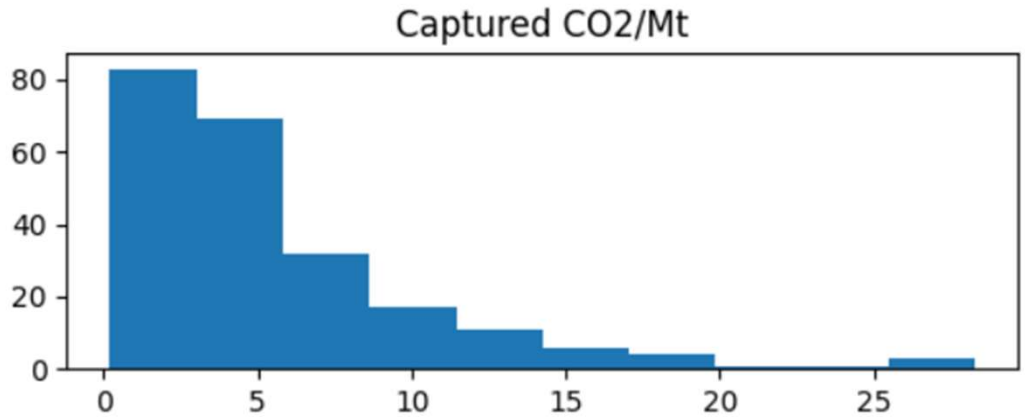
Data set 3

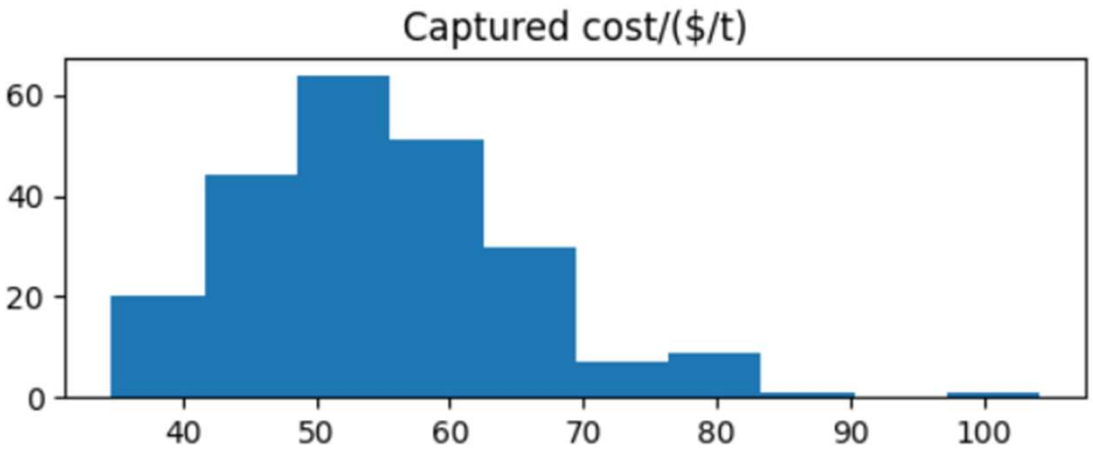
```
[34]
# Analisis estadistico - Medidas de Tendencia Central
plt.figure(figsize=(6,2))            ## Configuración del tamaño del grafico
plt.hist(df_China["Captured CO2/Mt"])  ## Histograma de Captured CO2/Mt
plt.figure(figsize=(6,2))            ## Configuración del tamaño del grafico
plt.hist(df_China["Captured cost/($/t)"])  ## Histograma de Captured CO2/Mt
```

```
(array([20., 44., 64., 51., 30.,  7.,  9.,  1.,  0.,  1.]),
 array([ 34.7 ,  41.65,  48.6 ,  55.55,  62.5 ,  69.45,  76.4 ,  83.35,
         90.3 ,  97.25, 104.2 ]),
 <BarContainer object of 10 artists>)
```

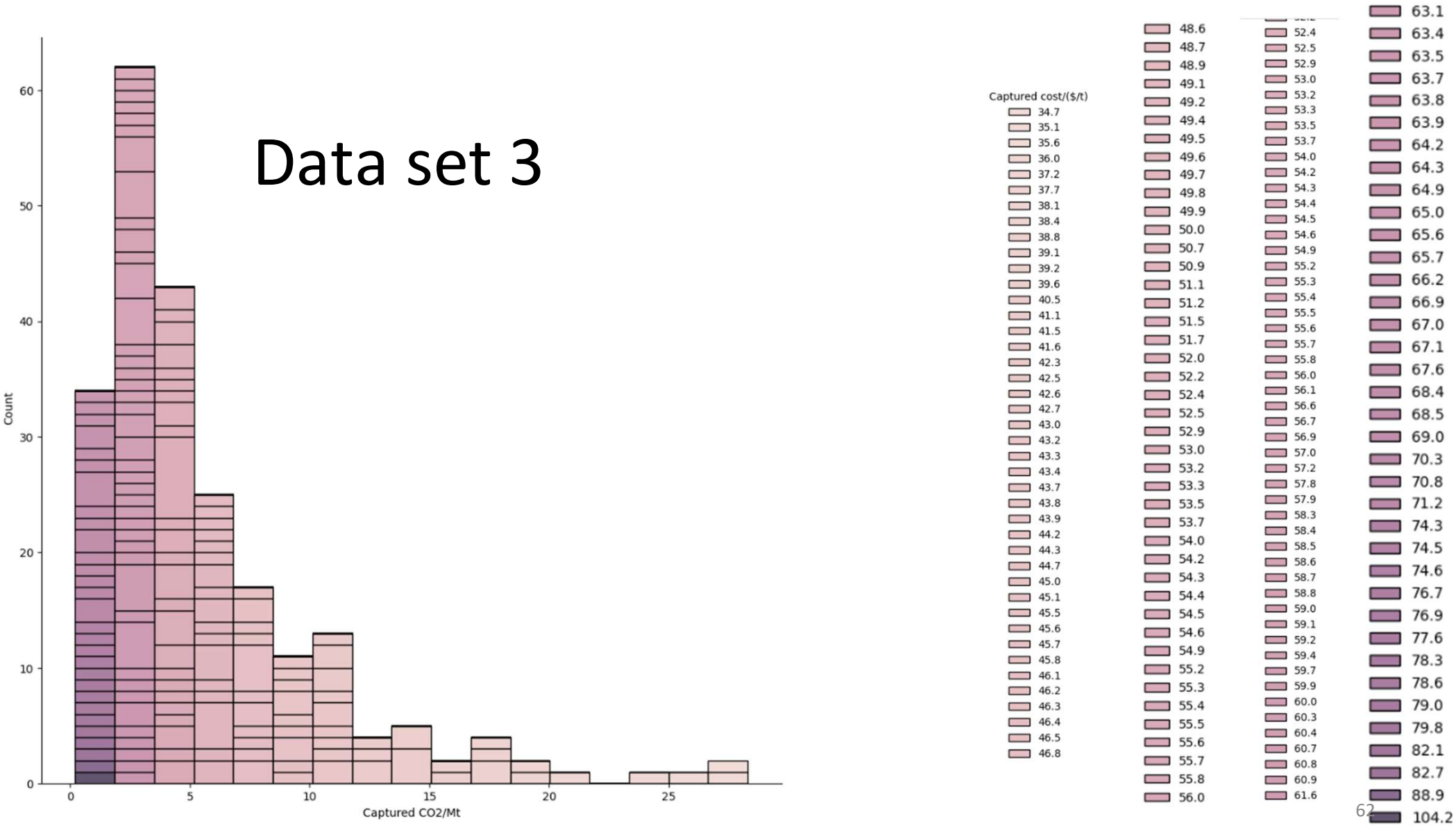


Data set 3


Captured CO2/Mt
80
60
40
20
0
0
5
10
15
20
25

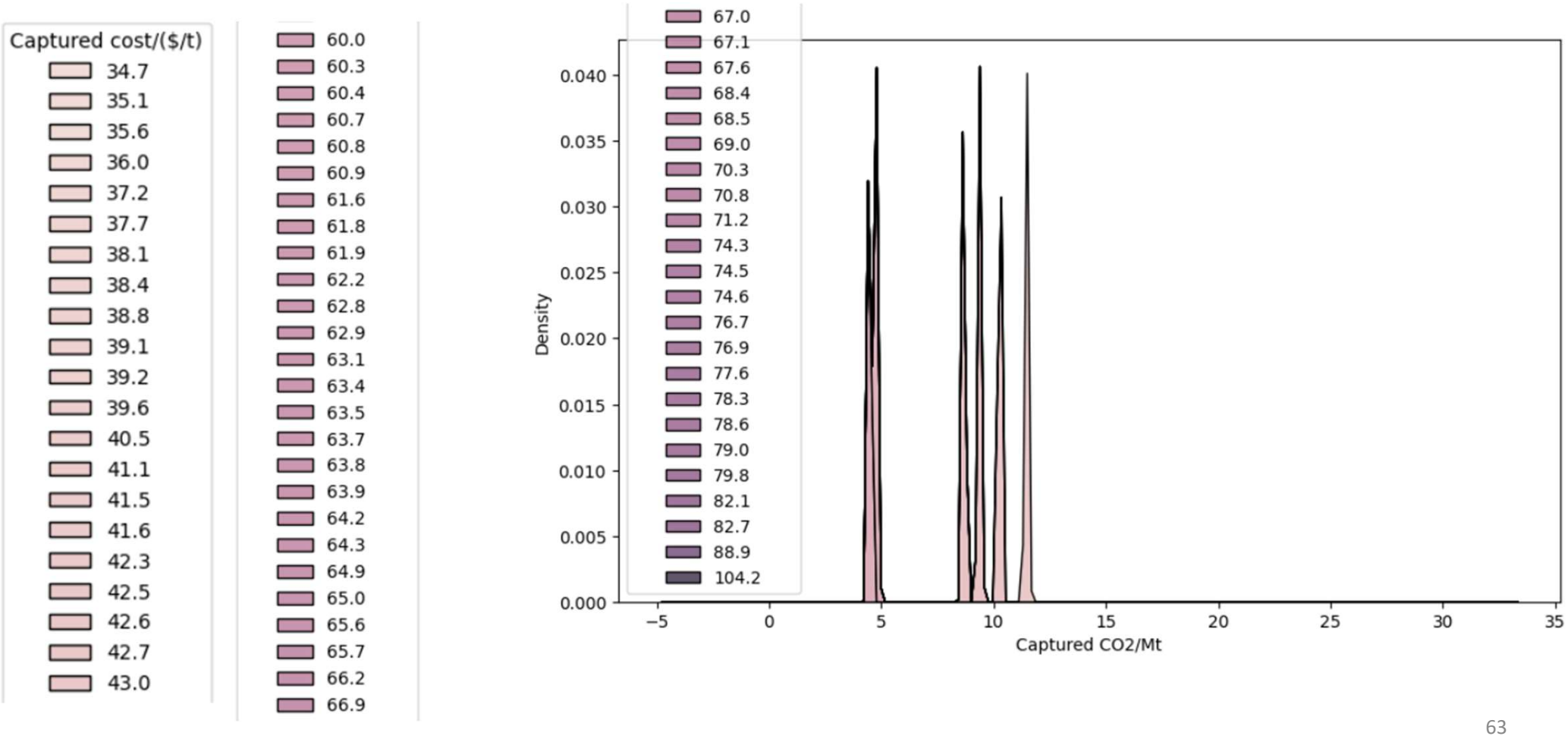

Captured cost/($/t)
60
40
20
0
40
50
60
70
80
90
100

Data set 3

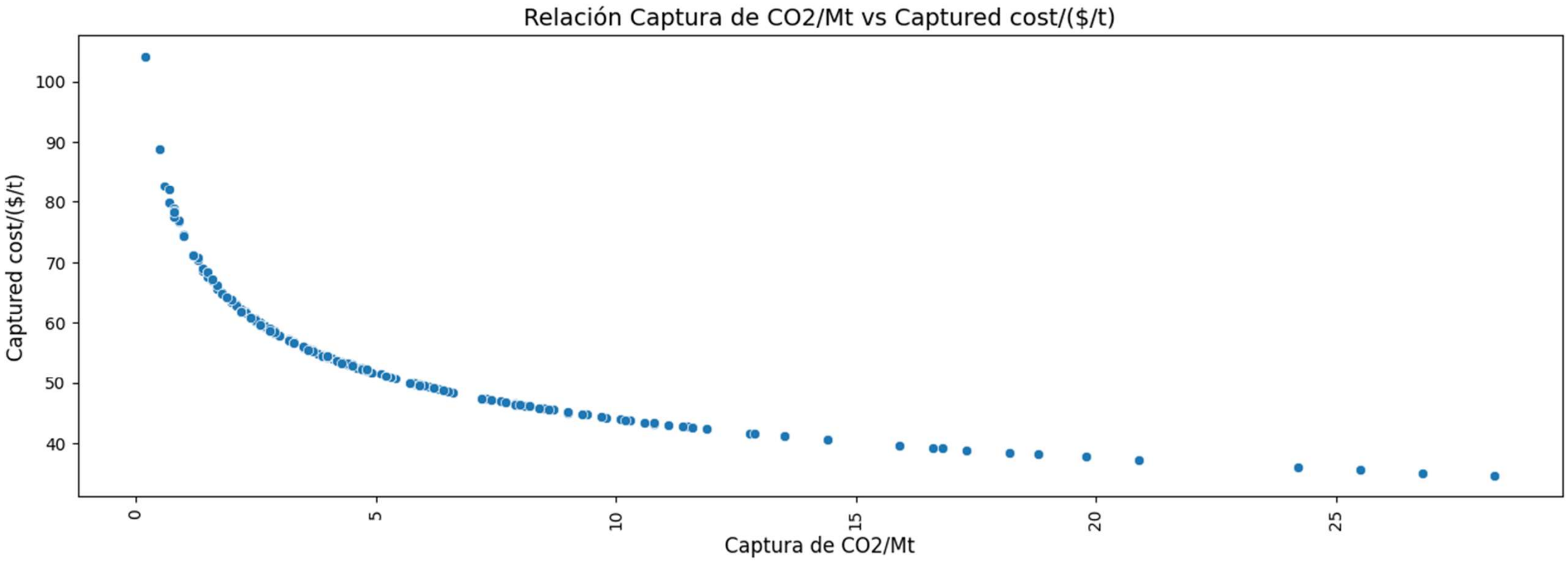

Count
Captured CO2/Mt
0
10
20
30
40
50
60
0
5
10
15
20
25
Captured cost/($/t)
34.7
35.1
35.6
36.0
37.2
37.7
38.1
38.4
38.8
39.1
39.2
39.6
40.5
41.1
41.5
41.6
42.3
42.5
42.6
42.7
43.0
43.2
43.3
43.4
43.7
43.8
43.9
44.2
44.3
44.7
45.0
45.1
45.5
45.6
45.7
45.8
46.1
46.2
46.3
46.4
46.5
46.8
48.6
48.7
48.9
49.1
49.2
49.4
49.5
49.6
49.7
49.8
49.9
50.0
50.7
50.9
51.1
51.2
51.5
51.7
52.0
52.2
52.4
52.5
52.9
53.0
53.2
53.3
53.5
53.7
54.0
54.2
54.3
54.4
54.5
54.6
54.9
55.2
55.3
55.4
55.5
55.6
55.7
55.8
56.0
52.4
52.5
52.9
53.0
53.2
53.3
53.5
53.7
54.0
54.2
54.3
54.4
54.5
54.6
54.9
55.2
55.3
55.4
55.5
55.6
55.7
55.8
56.0
56.1
56.6
56.7
56.9
57.0
57.2
57.8
57.9
58.3
58.4
58.5
58.6
58.7
58.8
59.0
59.1
59.2
59.4
59.7
59.9
60.0
60.3
60.4
60.7
60.8
60.9
61.6
63.1
63.4
63.5
63.7
63.8
63.9
64.2
64.3
64.9
65.0
65.6
65.7
66.2
66.9
67.0
67.1
67.6
68.4
68.5
69.0
70.3
70.8
71.2
74.3
74.5
74.6
76.7
76.9
77.6
78.3
78.6
79.0
79.8
82.1
82.7
88.9
104.2

Data set 3 KDE

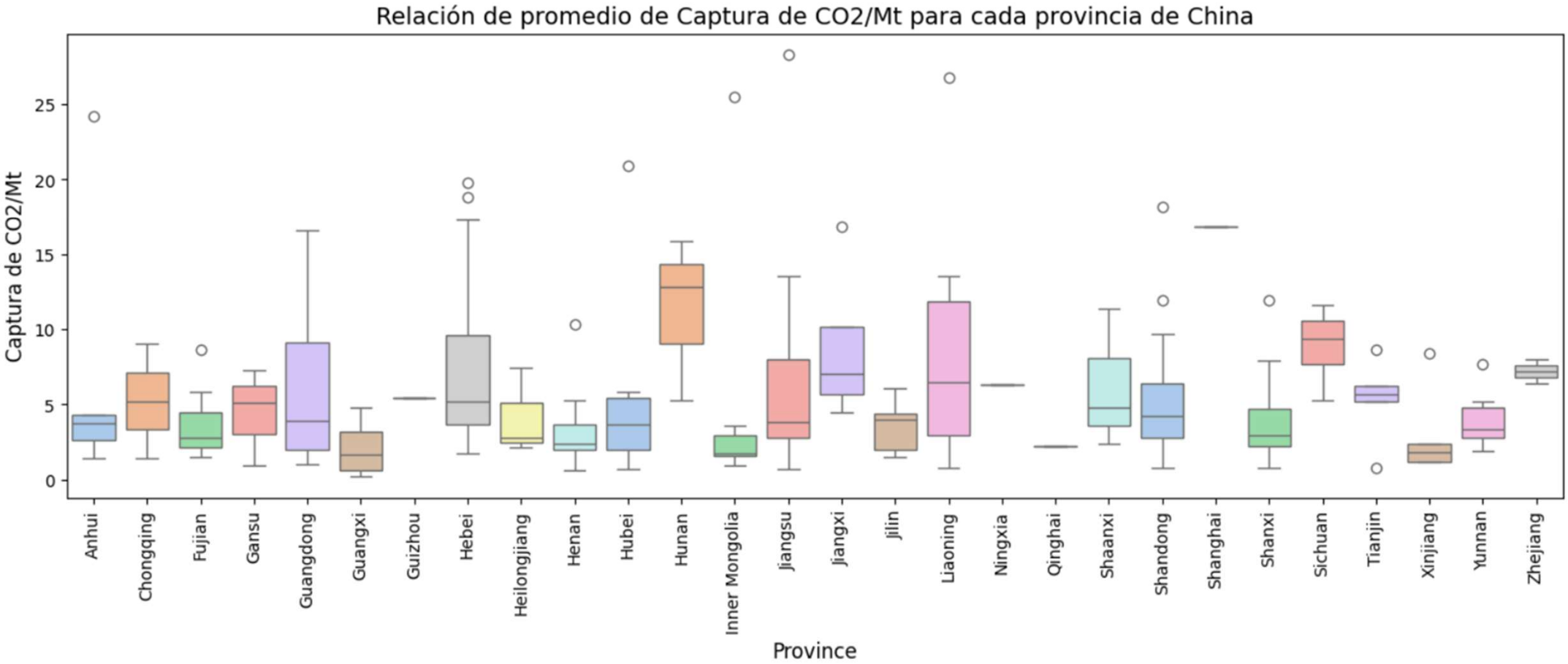
kernel density estimate (KDE)

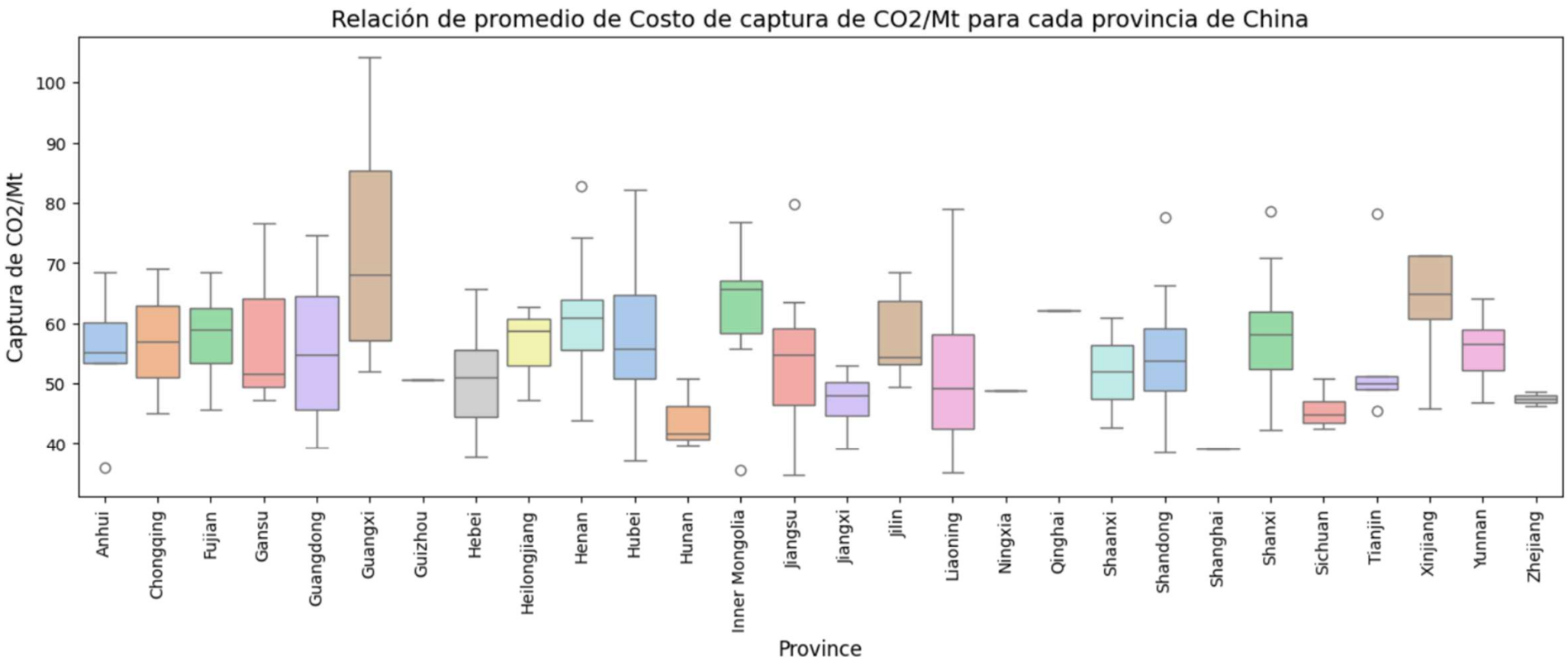


Data set 3 *Scatterplot*

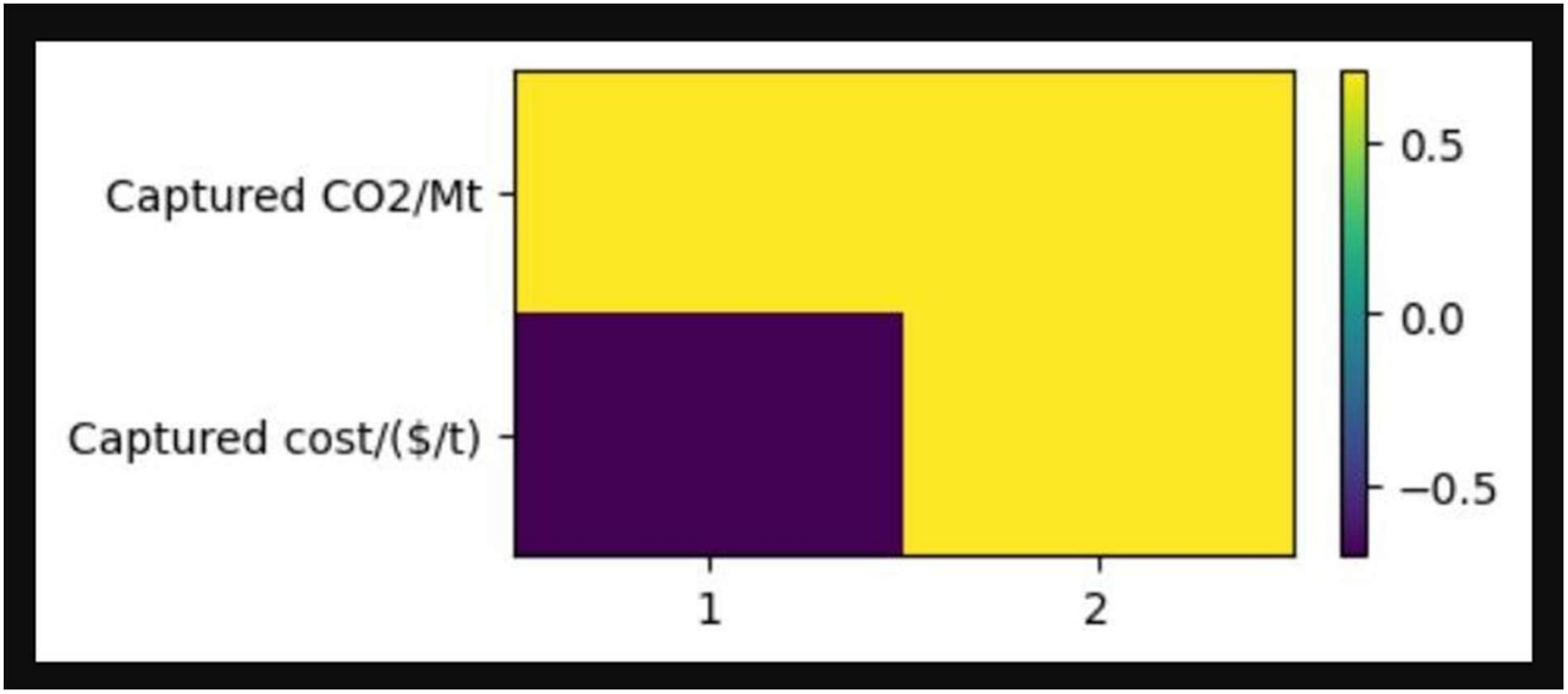


Data set 3

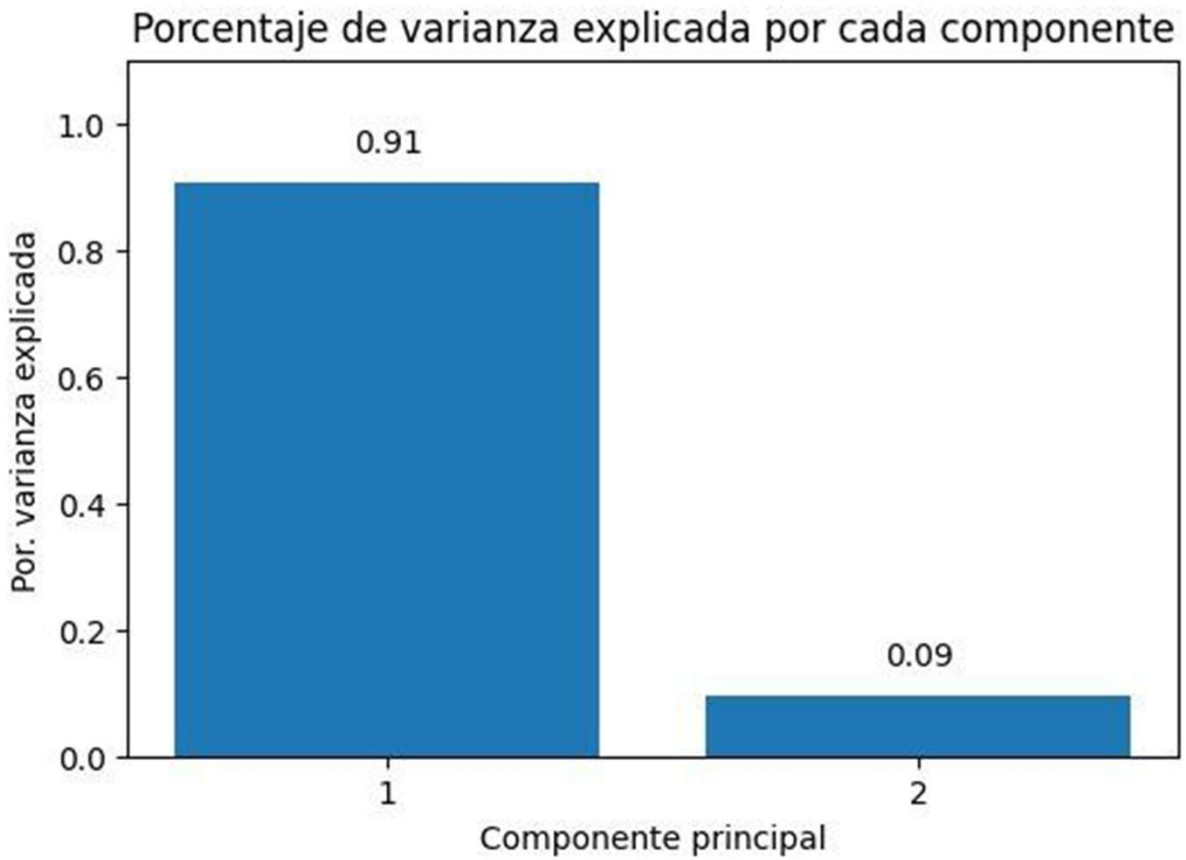

Relación de promedio de Captura de CO2/Mt para cada provincia de China
Captura de CO2/Mt
0
5
10
15
20
25
Anhui
Chongqing
Fujian
Gansu
Guangdong
Guangxi
Guizhou
Hebei
Heilongjiang
Henan
Hubei
Hunan
Inner Mongolia
Jiangsu
Jiangxi
Jilin
Liaoning
Ningxia
Qinghai
Shaanxi
Shandong
Shanghai
Shanxi
Sichuan
Tianjin
Xinjiang
Yunnan
Zhejiang
Province


Relación de promedio de Costo de captura de CO2/Mt para cada provincia de China
Captura de CO2/Mt
100
90
80
70
60
50
40
Anhui
Chongqing
Fujian
Gansu
Guangdong
Guangxi
Guizhou
Hebei
Heilongjiang
Henan
Hubei
Hunan
Inner Mongolia
Jiangsu
Jiangxi
Jilin
Liaoning
Ningxia
Qinghai
Shaanxi
Shandong
Shanghai
Shanxi
Sichuan
Tianjin
Xinjiang
Yunnan
Zhejiang
Province

Data set 3 Modelado PCA - Mapa de Calor



Data set 3 Modelado PCA



Data set 3

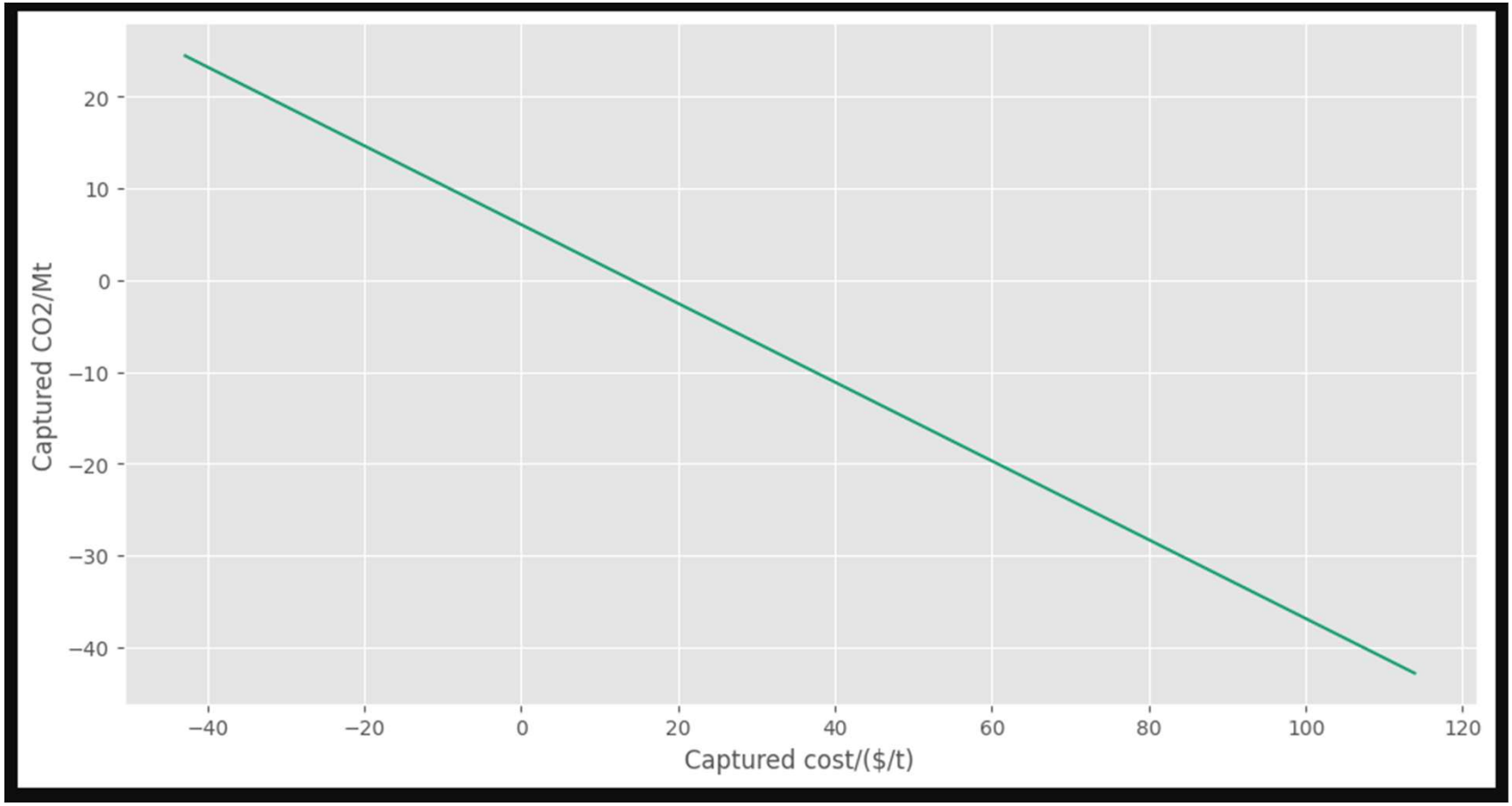
	variable_1	variable_2	r	abs_r
1	Captured CO2/Mt	Captured cost/($/t)	-0.811049	0.811049
2	Captured cost/($/t)	Captured CO2/Mt	-0.811049	0.811049

Matriz de Covarianza

```python
#Matriz de Covarianza
# Selecting only numerical columns for covariance calculation
numerical_df = df_China.select_dtypes(include=['number'])
numerical_df.cov(min_periods=None, ddof=1)
```

	Captured CO2/Mt	Captured cost/($/t)
Captured CO2/Mt	24.471020	-42.813343
Captured cost/($/t)	-42.813343	113.870789

Matriz de Covarianza



```python
# Correlación entre columnas numéricas
def tidy_corr_matrix(corr_mat):
    '''
    Función para convertir una matriz de correlación de pandas en formato tidy
    '''
    corr_mat = corr_mat.stack().reset_index()
    corr_mat.columns = ['variable_1','variable_2','r']
    corr_mat = corr_mat.loc[corr_mat['variable_1'] != corr_mat['variable_2'], :]
    corr_mat['abs_r'] = np.abs(corr_mat['r'])
    corr_mat = corr_mat.sort_values('abs_r', ascending=False)

    return(corr_mat)

corr_matrix = df_China.select_dtypes(include=['float64', 'int']) \
                .corr(method='pearson')
display(tidy_corr_matrix(corr_matrix).head(5))
```

	variable_1	variable_2	r	abs_r
1	Captured CO2/Mt	Captured cost/($/t)	-0.811049	0.811049
2	Captured cost/($/t)	Captured CO2/Mt	-0.811049	0.811049

Journal of CO2 Utilization 89 (2024) 102941

Contents lists available at ScienceDirect

Journal of CO2 Utilization

journal homepage: www.elsevier.com/locate/jcou

Review article

Machine learning of metal-organic framework design for carbon dioxide capture and utilization

Yang Jeong Park [a,b], Sungroh Yoon [b,c,*], Sung Eun Jerng [d,**]

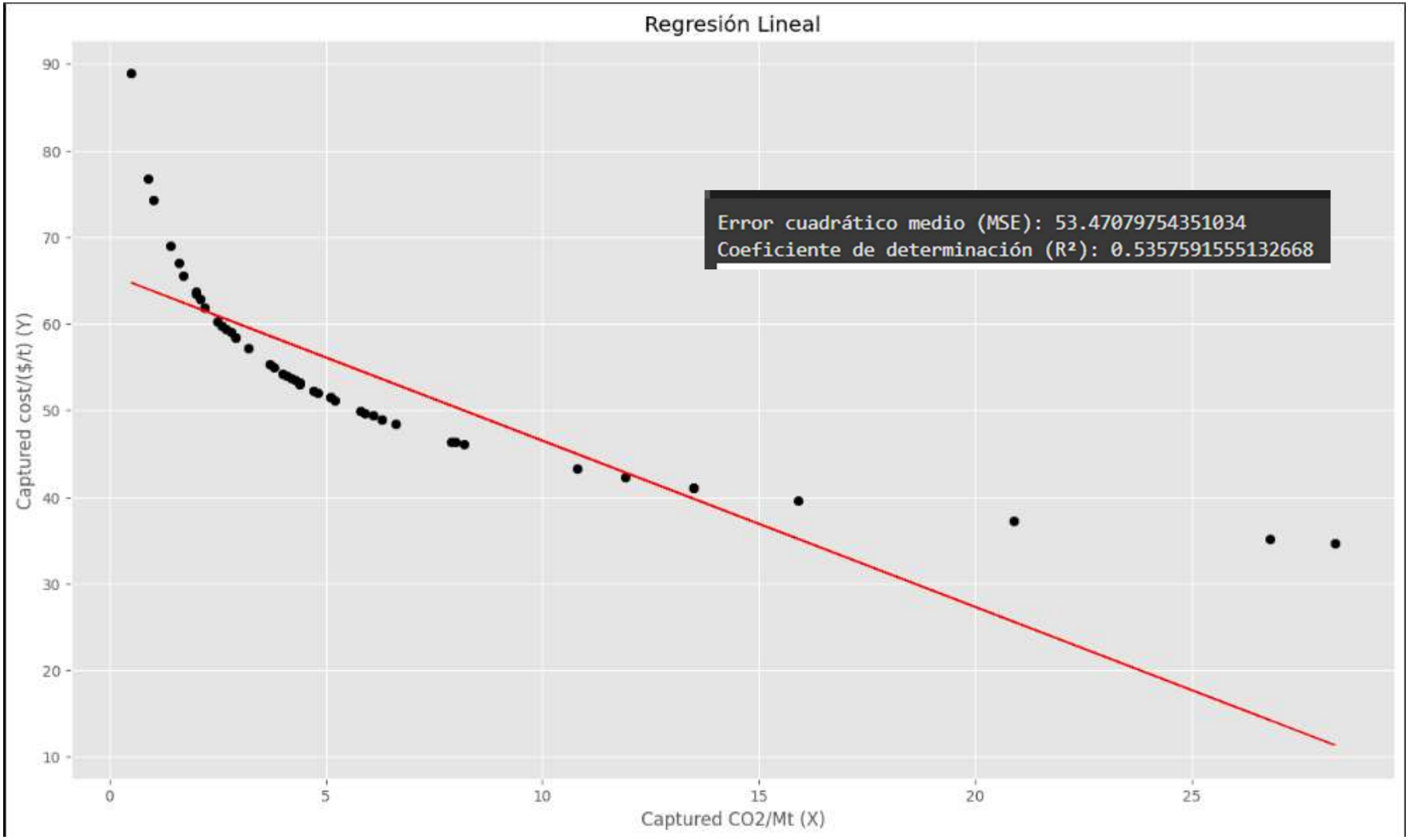
[a] Department of Materials Science and Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, 02139, USA
[b] Department of Electrical and Computer Engineering, Seoul National University, Seoul, 08826, Republic of Korea
[c] Interdisciplinary Program in Artificial Intelligence, Seoul National University, Seoul, 08826, Republic of Korea
[d] Department of Environmental Energy Engineering, The University of Suwon, 17, Wauan-gil, Bongdam-eup, Hwaseong-si, Gyeonggi-do, 18323, Republic of Korea

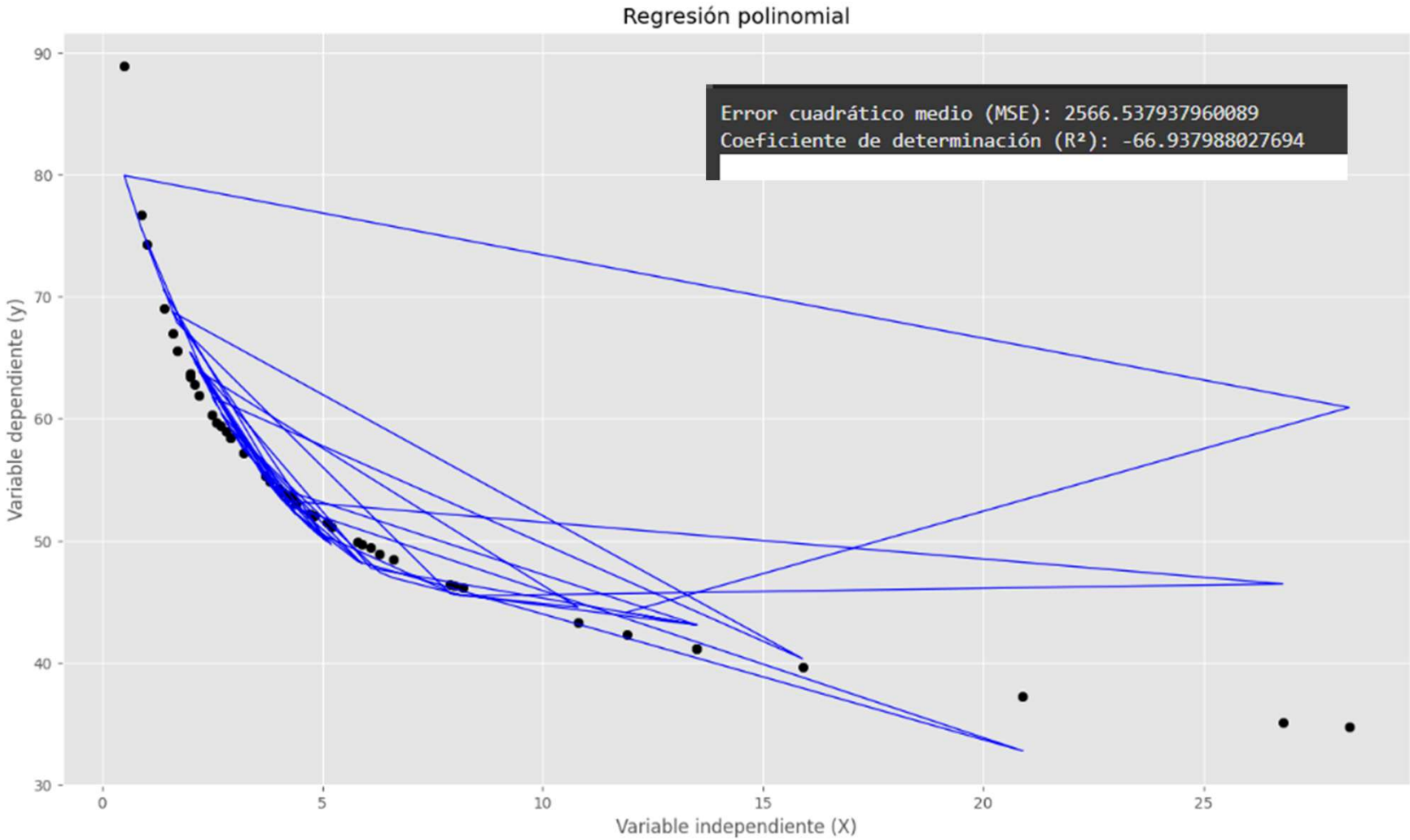
3.2.1. Root Mean Square Error (RMSE)

RMSE measures the square root of the average squared differences between predicted values ($\hat{y}_i$) and actual values (y_i). It is sensitive to large errors, making it useful when larger errors are particularly undesirable. The formula for RMSE is:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}$$

where n is the number of observations.


Regresión Lineal
Error cuadrático medio (MSE): 53.47079754351034
Coeficiente de determinación (R²): 0.5357591555132668
Captured cost/($/t) (Y)
Captured CO2/Mt (X)
90
80
70
60
50
40
30
20
10
0
5
10
15
20
25


Regresión polinomial
Error cuadrático medio (MSE): 2566.537937960089
Coeficiente de determinación (R²): -66.937988027694
Variable dependiente (y)
Variable independiente (X)
90
80
70
60
50
40
30
0
5
10
15
20
25

Dashboard

- Enlace:
- https://104d-35-232-177-255.ngrok-free.app/

Referencias

- Data set 1 : https://netl.doe.gov/carbon-management/carbon-storage/worldwide-ccs-database
- Data set 2: https://www.iea.org/data-and-statistics/data-product/ccus-projects-database
- Data set 3: https://www.sciencedirect.com/science/article/pii/S0959652624017281?via%3Dihub#appsec1

The different modules involved in the CO2 shipping process chain

https://www.sciencedirect.com/science/article/pii/S1364032123000710