

**Universidad Americana**  
**Facultad de Ingeniería y Arquitectura**



**Inteligencia de Negocios**

**“Informe Exploración y Análisis de datos Ambientales”**

**Elaborado por:**  
Alejandro Jaime

**Docente:** Arlen López

**Fecha de entrega:** 07/10/2024

<b>Abstracto</b>	<b>3</b>
<b>Introducción</b>	<b>3</b>
<b>Recopilación y preprocesamiento de datos</b>	<b>3</b>
<b>Metodología</b>	<b>6</b>
<b>Resultados</b>	<b>6</b>
<b>Discusión</b>	<b>11</b>
<b>Conclusiones</b>	<b>12</b>
<b>Referencias</b>	<b>12</b>

## **Abstracto**

Desde el siglo XIX el aumento de los niveles de emisión de CO<sub>2</sub> ha influido de gran manera en el cambio climático, por eso es necesario identificar cuales son los países que emiten más CO<sub>2</sub> y los factores que están detrás de estas emisiones para así poder formular políticas ambientales efectivas. A través de un análisis y exploración que se realizó al conjunto de datos titulado “CO<sub>2</sub> Emission by countries Year wise (1750-2022)” el cual cuenta con datos relevantes de las emisiones de CO<sub>2</sub> de todos los países durante el período de 1750 a 2022 fue posible encontrar conclusiones que contextualizan el posible aumento de emisiones de CO<sub>2</sub> durante los últimos años. Mediante el uso de herramientas de visualización como gráfico de barra, histograma, gráfico lineal, gráfico de dispersión y mapa de calor se lograron analizar los datos y encontrar que el país con mayor emisión de CO<sub>2</sub> es Estados Unidos seguido por China, a pesar que China es el país más poblado del mundo no necesariamente es el que emite mayor cantidad de CO<sub>2</sub>. Los resultados obtenidos demostraron que eventos como la Revolución Industrial y la Segunda Guerra mundial dieron pauta a una mayor industrialización de los países lo cual trajo consigo un mayor uso de combustibles fósiles que producen una mayor emisión de CO<sub>2</sub> y por ende una mayor contaminación de la atmósfera.

## **Introducción**

Uno de los principales desafíos que ha tenido el mundo en este siglo XXI es el cambio climático, una de las causas de este fenómeno son las emisiones de dióxido de carbono(CO<sub>2</sub>). Mientras más concentración de CO<sub>2</sub> haya en la atmósfera, las consecuencias del calentamiento global serán de mayor impacto, como lo son el crecimiento del nivel del mar, incendios forestales, sequías, aumento en las temperaturas, entre otras. Es necesario identificar los principales emisores de CO<sub>2</sub> y comprender las variables que impulsan estas emisiones, para así poder llevar a cabo acciones efectivas que mitiguen el impacto ambiental.

El objetivo de este análisis es estudiar las emisiones de CO<sub>2</sub> de todos los países durante el periodo de 1750 a 2020, identificando patrones o tendencias notables en el conjunto de datos mediante la creación de distintos gráficos que permitan visualizar y explorar estos datos.

## Recopilación y preprocesamiento de datos

Para este análisis, se utilizó el conjunto de datos “CO2 Emission by countries Year wise (1750-2022)” sacado de Kaggle. Para poder acceder a la información de este dataset se almacenó en un repositorio público de github y utilizando el enlace de este repositorio se cargaron dichos datos dentro de un cuaderno de Jupyter.

```
#2. CARGAR EL CONJUNTO DE DATOS

#URL del dataset en GITHUB
url = 'https://github.com/Ajaime21/ajaime/raw/refs/heads/main/CO2%20emission%20by%20countries.csv'

#Cargar el dataset desde la URL
df = pd.read_csv(url, encoding='unicode_escape')

#Crear una copia del dataset original para modificar
df_modificado = df.copy()

#Imprimir las primeras 5 líneas del dataset
print(df_modificado.head())
```

	Country	Code	Calling Code	Year	CO2 emission (Tons)	Population(2022)	\
0	Afghanistan	AF	93	1750	0.0	41128771.0	
1	Afghanistan	AF	93	1751	0.0	41128771.0	
2	Afghanistan	AF	93	1752	0.0	41128771.0	
3	Afghanistan	AF	93	1753	0.0	41128771.0	
4	Afghanistan	AF	93	1754	0.0	41128771.0	

	Area	% of World	Density(km2)
0	652230.0	0.40%	63/km²
1	652230.0	0.40%	63/km²
2	652230.0	0.40%	63/km²
3	652230.0	0.40%	63/km²
4	652230.0	0.40%	63/km²

Se realizó una exploración inicial de los datos

```
#3. EXPLORACION INICIAL

#Mostrar número de filas y columnas
print(f"Filas: {df.shape[0]}, Columnas: {df.shape[1]}")
```

El dataset cuenta con un total de 59620 registros y 9 columnas

```
Filas: 59620, Columnas: 9
```

```
# Ver el tipo de dato de cada columna
print("")
print(df_modificado.dtypes)
```

Dentro de las 9 variables tenemos que las columnas Country, Code, Calling Code, %of the World y Density(km2) son de tipo caracter, en cambio las columnas Year, CO2 emission (Tons), Population(2022) y Area son de tipo numéricas

```
Country      object
Code         object
Calling Code  object
Year         int64
CO2 emission (Tons) float64
Population(2022) float64
Area         float64
% of World   object
Density(km2) object
```

```
# Ver descripción de las variables numéricas
print("")
print(df_modificado.describe())
```

Se calcularon estadísticas descriptivas básicas para las variables numéricas como la media, desviación estándar, el mínimo y el máximo

	Year	CO2 emission (Tons)	Population(2022)	Area
count	59620.000000	5.962000e+04	5.311600e+04	5.528400e+04
mean	1885.000000	1.034774e+09	3.992260e+07	6.522073e+05
std	78.231085	1.041652e+10	1.482365e+08	1.865483e+06
min	1750.000000	0.000000e+00	1.131200e+04	2.100000e+01
25%	1817.000000	0.000000e+00	1.770414e+06	1.770450e+04
50%	1885.000000	0.000000e+00	8.673095e+06	1.103815e+05
75%	1953.000000	8.715092e+06	2.862920e+07	4.925730e+05
max	2020.000000	4.170000e+11	1.425887e+09	1.709824e+07

```
# Ver descripción de las variables categóricas
print("")
print(df_modificado.describe(include=['object']))
```

Una descripción de las variables categóricas

	Country	Code	Calling Code	% of World	Density(km2)
count	59620	57452	56097	55284	53116
unique	220	212	203	23	145
top	Afghanistan	AF	1	0.00%	4/km²
freq	271	271	542	24119	1626

## Metodología

La metodología que se utilizó en este análisis consiste en una combinación de técnicas estadísticas y visualizaciones de datos para explorar las relaciones entre las diferentes columnas/variables del dataset y así identificar patrones o tendencias notables en el conjunto de datos.

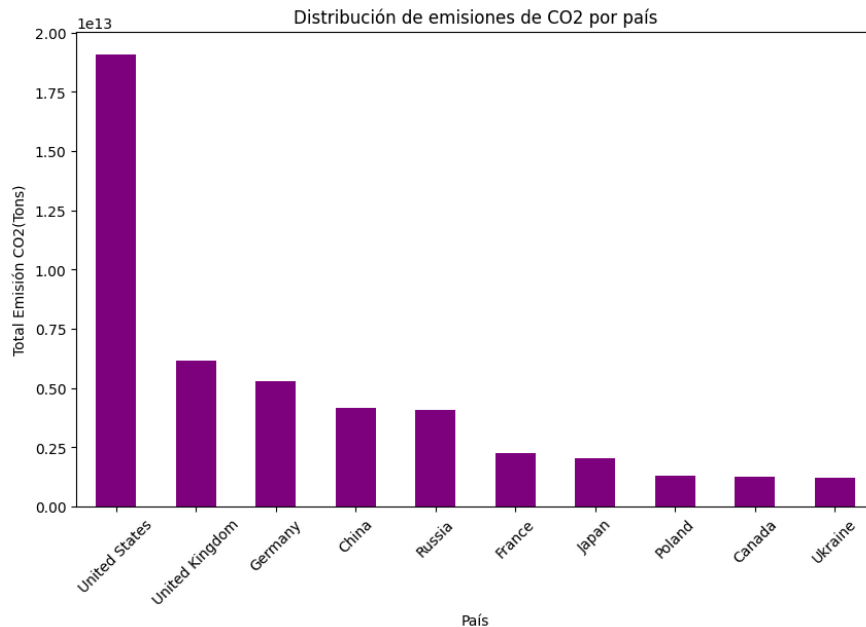
Se realizó un análisis descriptivo inicial para resumir las características principales de las variables numéricas del conjunto de datos, se calcularon estadísticas básicas como la media, mínimo, máximo y desviación estándar para estas variables

Para la visualización de datos se utilizaron distintos gráficos estadísticos como por ejemplo el gráfico de barras para ver la distribución de emisiones de CO2 de los 10 países con mayor emisión, también un gráfico de líneas para ver las tendencias globales de emisiones de CO2 y de igual forma para ver las tendencias de emisiones de CO2 de los 10 países con mayor emisión a lo largo del tiempo. También se utilizó un gráfico de dispersión para observar la relación entre la población y emisión de CO2 de los 10 países con mayor emisión durante el año 2020. Por último, se utilizó un mapa de calor para examinar las correlaciones entre la población, emisiones de CO2 y Área de los países.

## Resultados

Dentro de las visualizaciones obtuvimos los siguientes gráficos

### Gráfico de barras para la distribución de de emisiones de CO2 por país(los 10 países con mayor emisión)



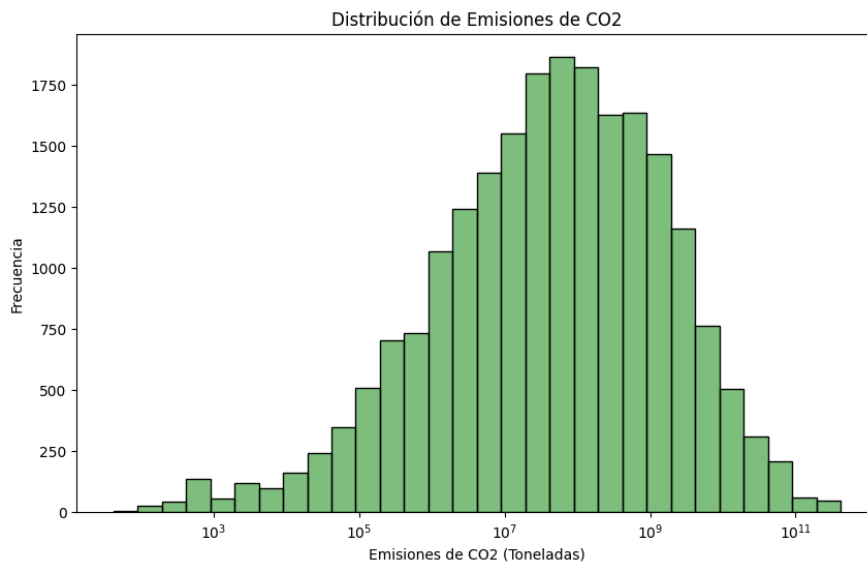
-Estados Unidos es con gran diferencia el mayor emisor de CO2 entre todos los países, con emisiones totales que superan significativamente a la de los otros países.

-Reino Unido, Alemania, China y Rusia siguen a Estados Unidos, pero con cifras notablemente más bajas. Aproximadamente sus emisiones son un tercio o menos en comparación con la de Estados Unidos.

-Francia y Japón presentan niveles de emisión muy similares aunque un poco más bajos que los de China y Rusia.

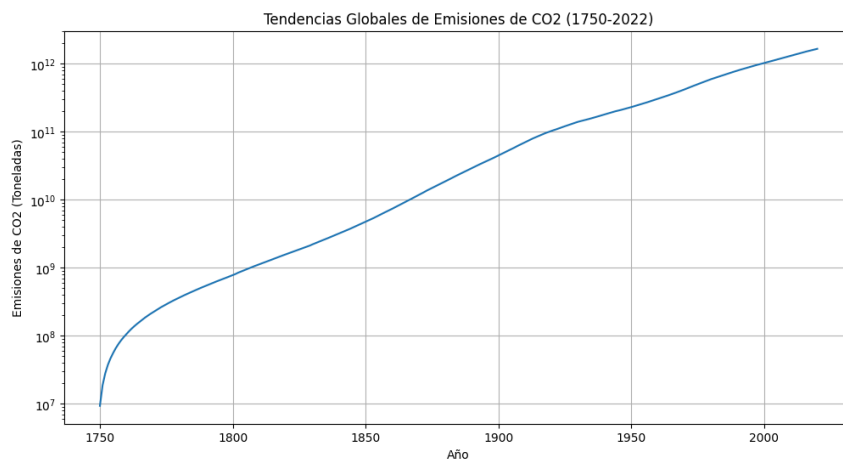
-Polonia, Canadá y Ucrania tienen las emisiones más bajas dentro de este grupo de países, sin embargo igualmente contribuyen significativamente en comparación con los otros países que no aparecen en el gráfico.

## Histograma de emisiones de CO2



- El histograma tiene una forma aproximadamente simétrica, lo que muestra que la mayoría de los valores de emisiones de CO2 se encuentran en el rango de  $10^7$  y  $10^8$  toneladas.
- La mayor frecuencia de datos está en torno a emisiones de CO2 en el orden de  $10^8$  toneladas, es decir que muchos registros de emisiones se encuentran cercanos a ese valor.
- A partir de  $10^9$  toneladas, hay una menor cantidad de registros, aunque sí hay registros con emisiones muy altas como  $10^{10}$  y  $10^{11}$  toneladas, son menos comunes.
- De igual forma se observan pocos registros de emisiones muy pequeñas como  $10^4$  y  $10^3$ , lo cual se podría entender como aquellos países pequeños o los datos de los primeros años cuando las emisiones eran más bajas.

## Gráfico de líneas para ver las tendencias globales de emisiones de CO2 a lo largo del tiempo



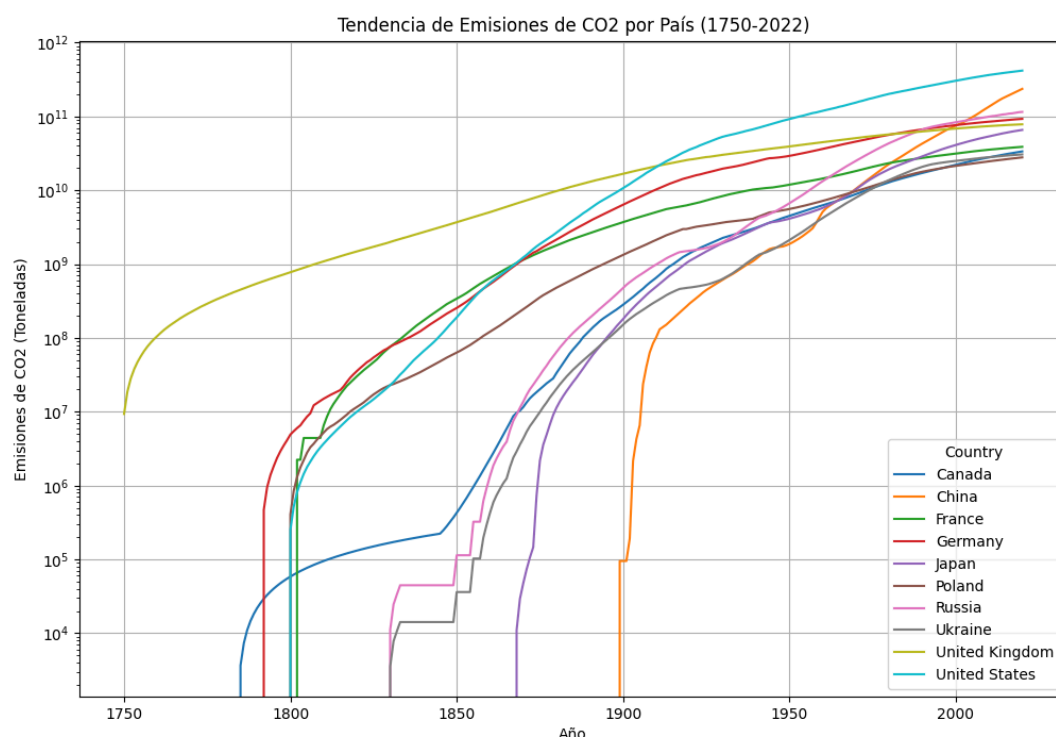


-A partir del año 1850 aproximadamente, ha habido un crecimiento drástico en las emisiones globales de CO<sub>2</sub>. Un punto importante a destacar, es la revolución industrial debido a que durante esta época, se incrementó el uso de combustibles fósiles.

-Anterior a ese año(1850), la emisiones eran algo bajas, lo cual se traduce en una menor actividad industrial y la limitada quema de combustibles fósiles a nivel global.

-Alrededor del año 1950, es decir luego de la segunda guerra mundial, se refleja un crecimiento mayor. Esto se puede deber al crecimiento económico global, la expansión industrial así como el aumento del consumo energético.

### Gráfico de líneas para ver las tendencia de Emisiones de CO<sub>2</sub> por País(10 países con mayor emisión) a lo largo del tiempo



-La mayoría de los países presentan una tendencia creciente de emisiones de CO<sub>2</sub> a lo largo del tiempo.

-China y Estados Unidos presentan un crecimiento rápido y continuo en las emisiones de CO<sub>2</sub> a partir de mediados del siglo XX.

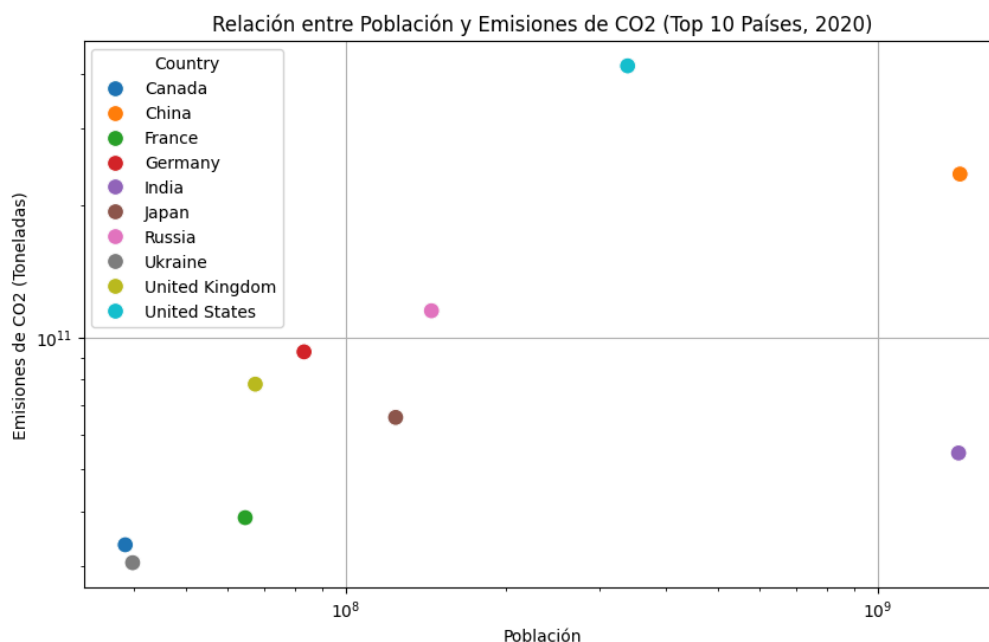
-Estados Unidos lidera las emisiones acumuladas en casi todo el periodo, sin embargo China ha presentado un crecimiento demasiado acelerado en las últimas décadas y es posible que en algún momento alcance a USA.

-Reino Unido, muestra un rápido aumento en las emisiones de CO<sub>2</sub> en el siglo XIX, tomando en cuenta que es pionero en la Revolución Industrial, sin embargo durante el siglo XX su crecimiento se estabiliza.

-Francia también muestra un patrón de crecimiento temprano, aunque no tan pronunciado, siendo su curva más suave.

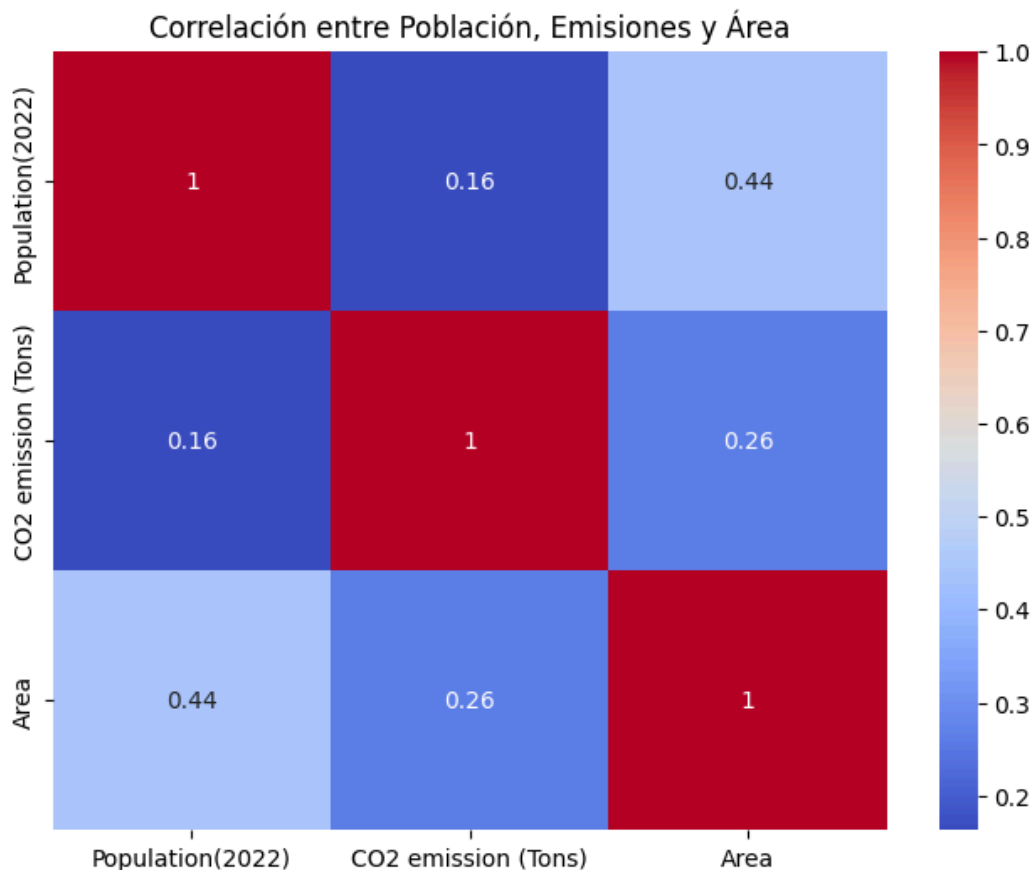
- Alemania, Japón y Rusia presentan un crecimiento moderado, aumentando sus emisiones constantemente en el siglo XX.
- Canadá y Ucrania presentan emisiones acumuladas mucho más bajas que los países considerados grandes emisores, aunque muestran un aumento constante en sus emisiones a lo largo del tiempo.

### Gráfico de dispersión de emisiones vs población para el año 2020



- Se eligieron los datos del año 2020 debido a que es el último año que aparece registrado en el dataset, por lo tanto se quiere saber cómo está la relación entre la población y las emisiones de CO2 recientemente.
- Como se observa en el gráfico, Estados Unidos es el mayor emisor de CO2 para este año, sin embargo si comparamos la población con la de China e India, la de USA es menor.
- Las emisiones de China son comparables con las de Estados Unidos, pero con una población mucho mayor.
- En términos de emisiones per cápita, las de China no son tan altas como la de Estados Unidos, a pesar de que China tiene la mayor cantidad de población y emisiones muy elevadas. Sin embargo, Estados Unidos si tiene emisiones muy altas en relación con su población.
- Aunque India tiene una gran población, sus emisiones totales son realmente muy inferiores a las de USA y China, lo que indica emisiones per cápita mucho más bajas.
- Francia, Reino Unido y Canadá presentan emisiones y poblaciones mucho menores en comparación con los gigantes de China y USA.

## Mapa de calor para ver la correlación entre las variables numéricas relevantes



-La correlación de 0.16 es muy baja e indica que un país con mayor población, no necesariamente tendrá mayores emisiones de CO2.

-Por otra parte, la correlación de 0.44 es moderada lo cual indica que los países más grandes a nivel territorial, tienden a tener un mayor número de población.

-Por último, la correlación de 0.26 es débil e indica que solo porque un país sea más grande va a emitir más CO2.

## Discusión

Luego de este extenso análisis, se puede decir que desde mediados del siglo XX las emisiones de CO2 han venido aumentando de manera significativa a lo largo del tiempo, sobre todo en países mayormente industrializados como China, USA, Rusia los cuales han experimentado un crecimiento constante en sus emisiones, producto de la misma industrialización.

También se puede decir que el tamaño de la población, no necesariamente es un factor que va a indicar una mayor emisión de CO2, como es el caso de China e India, a pesar de ser los países más poblados no necesariamente son los que emiten mayor emisión, el nivel de industrialización y el uso de combustible fósiles sí podrían ser factores determinantes. Países como China y USA, deberían considerar

el impacto global de sus políticas ambientales, debido a que son los mayores contribuyentes al cambio climático.

## **Conclusiones**

-Los resultados que se obtuvieron indican que la emisión de CO2 globalmente ha venido en crecimiento, sobre todo a partir de eventos como la Revolución Industrial o la Segunda Guerra Mundial, porque estos acontecimientos dieron pauta al progreso industrial y económico lo cual trae consigo un incremento de las emisiones. Los países más industrializados como USA y China son los mayores contaminantes y emisores de CO2 a nivel mundial, es por eso que se debería regular mediante la imposición de multas a las grandes corporaciones por no acatar la leyes ambientales de dichos países. Sería muy importante identificar los factores claves que influyen en la emisión a nivel global, sin embargo este conjunto de datos no nos permite realizar un análisis tan profundo debido a que no presenta tantas características que permitan identificar esos factores claves.

## **Referencias**

- <https://www.kaggle.com/datasets/moazzimalibhatti/co2-emission-by-countries-year-wise-17502022>
- <https://unfccc.int/es/news/las-emisiones-mundiales-de-co2-repuntaron-en-2021-hasta-su-nivel-mas-alto-de-la-historia>