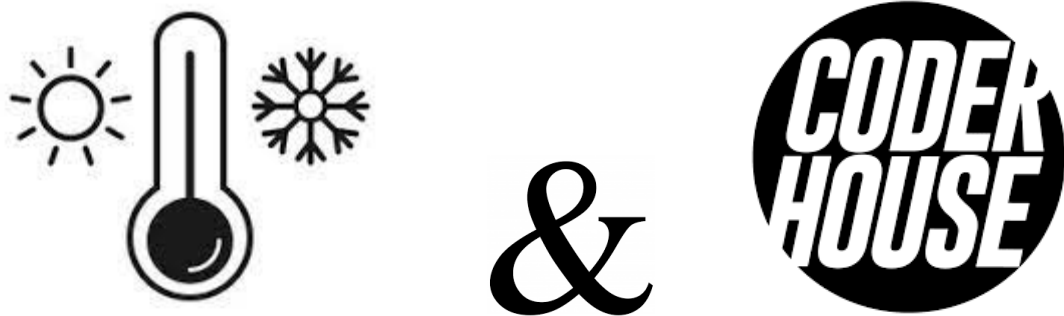


# PROYECTO

## “Temperatura y calidad del aire”



Institución: CODERHOUSE

Autora: Méndez, Julieta Milagros

Data Analytics 2024

## Índice

1. Introducción .....	2
1.1 Descripción de la temática.....	2
1.2 Objetivo .....	2
1.3 Hipótesis.....	2
1.4 Alcance y nivel de aplicación .....	3
2. Información sobre el Database .....	3
2.1 Diagrama Entidad-Relación .....	3
2.2 Tablas .....	4
Las tablas del modelo son las siguientes:.....	4
3. Carga y transformación en Power BI.....	6
3.1 Transformaciones realizadas.....	6
3.2 Medidas calculadas .....	7
4. Visualización de datos en Power BI .....	8
4.1 Solapa "ÍNDICE" .....	8
4.2 Solapa "GLOSARIO" .....	9
4.3 Solapa "TEMPERATURA AMERICA LATINA" .....	9
4.4 Solapa "TEMPERATURA BUENOS AIRES" .....	10
4.5 Solapa "AIRE AMERICA LATINA" .....	11
4.6 Solapa "PRECIPITACIÓN Y EVAPOTRANSPIRACIÓN" .....	12
5. Conclusiones .....	13
6. Futuras Líneas .....	15
7. Bibliografía.....	16

# 1. Introducción

## 1.1 Descripción de la temática

Este proyecto se centra en analizar el cambio en la temperatura y la calidad del aire en América Latina, una región caracterizada por su diversidad climática. El cambio climático es tema de agenda a nivel global desde hace años al igual que la contaminación del aire debido, en parte, a la urbanización acelerada del continente.

Este conjunto de datos contiene tres archivos CSV con información relacionada de varios países de América Latina y de la ciudad de Buenos Aires. Los datos incluyen el nombre del país, la ciudad capital, las temperaturas (desde 1991 a 2024) y los datos de calidad del aire (desde 2022 a 2024).

## 1.2 Objetivo

Para este trabajo se plantea el objetivo siguiendo la metodología SMART (\*):

Contribuir al conocimiento sobre el estado ambiental de América Latina analizando las variaciones en la temperatura y los niveles de contaminación del aire desde el año 1991 a 2024. Para esto se obtendrá el dataset desde

[https://www.kaggle.com/datasets/anycaroliny/latin-america-weather-and-air-quality-data?select=LA\\_daily\\_climate.csv](https://www.kaggle.com/datasets/anycaroliny/latin-america-weather-and-air-quality-data?select=LA_daily_climate.csv)

<https://data.buenosaires.gob.ar/dataset/registro-temperatura-ciudad>

y con las herramientas de análisis y limpieza de datos aprendidas se logrará identificar cómo evoluciona su comportamiento al finalizar el curso.

(\*) S: *specific* (contribuir al conocimiento sobre el estado ambiental de América Latina) - M: *measurable* (analizando datos del 1991 al 2024) - A: *achievable* (partiendo de un dataset con los datos necesarios) - R: *realistic* (con herramientas vistas durante el curso) - T: *timebound* (finalizado al culminar el curso)

## 1.3 Hipótesis

Se espera observar un aumento de la temperatura en América Latina a medida que transcurren los años, relacionado con un deterioro en la calidad del aire, particularmente en las zonas urbanas.

Se buscará comprobar que el cambio será más pronunciado en regiones con alta densidad poblacional y crecimiento industrial, lo que sugiere que el calentamiento global y la urbanización están contribuyendo significativamente a la degradación ambiental en la región.

## 1.4 Alcance y nivel de aplicación

El proyecto abarca el análisis de las temperaturas y la calidad del aire en países de América Latina, cubriendo el período de 2020 a 2024. Para Buenos Aires, el análisis de temperatura se extiende desde 1991 hasta 2024.

El enfoque del proyecto está orientado hacia el análisis estratégico, con el fin de evaluar la magnitud del cambio climático y su impacto en la región. El objetivo principal es comprender la evolución de este fenómeno y diseñar planes a largo plazo para mitigar sus efectos negativos.

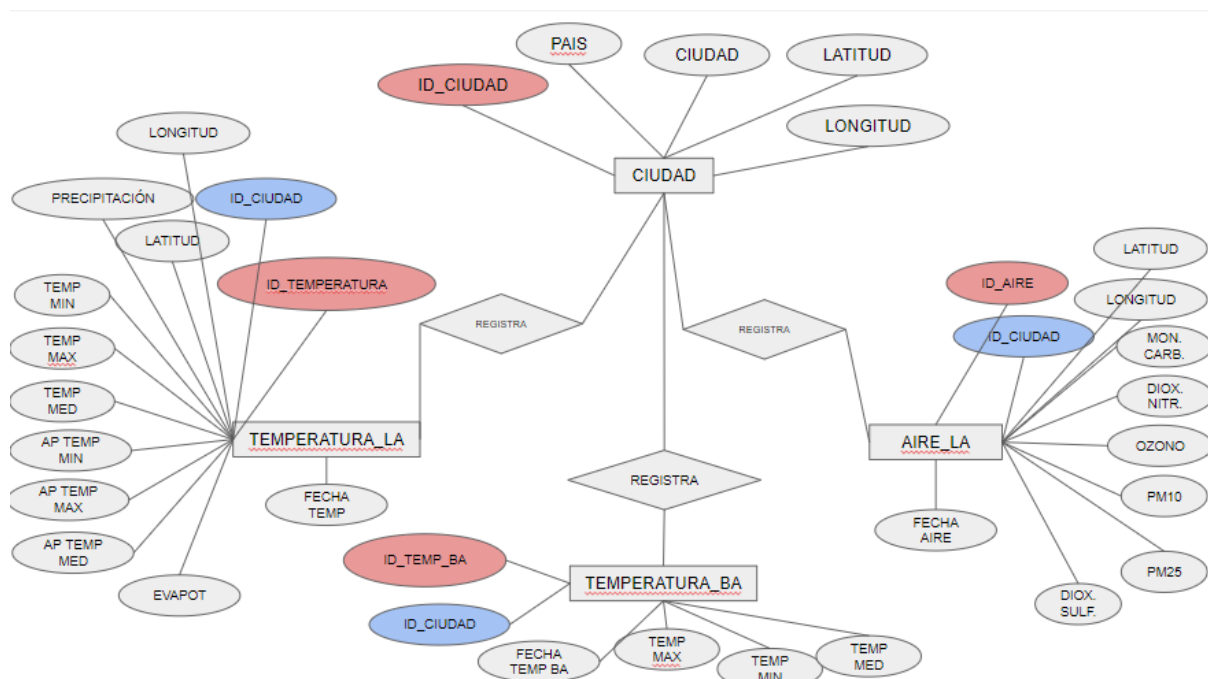
El proyecto está dirigido a los líderes de los países latinoamericanos comprometidos con la lucha contra el cambio climático y que lo han integrado como un tema prioritario en sus agendas futuras.

## 2. Información sobre el Dataset

### 2.1 Diagrama Entidad-Relación

A continuación, se muestra el diagrama Entidad-Relación detallando las entidades: CIUDAD, TEMPERATURA\_LA, TEMPERATURA\_BA y AIRE\_LA, junto con sus atributos y relaciones.

Las claves primarias de cada entidad están resaltadas en rojo y las claves foráneas están resaltadas en celeste.



## 2.2 Tablas

Las tablas del modelo son las siguientes:

- CIUDAD: contiene todas las ciudades a analizar junto con su país, su latitud y su longitud. Los campos son:
  - ID\_CIUADAD: identifica de manera única a cada ciudad de la tabla.  
**Tipo de dato: INT. Es clave primaria.**
  - PAIS: nombre del país de cada ciudad.  
**Tipo de dato: VARCHAR (100).**
  - CIUDAD: nombre de la ciudad.  
**Tipo de dato: VARCHAR (100).**
  - LATITUD: número que corresponde a la latitud de cada ciudad.  
**Tipo de dato: FLOAT.**
  - LONGITUD: número que corresponde a la longitud de cada ciudad.  
**Tipo de dato: FLOAT.**
- TEMPERATURA\_LA: contiene todos los registros de temperatura de cada ciudad de Latinoamérica por fecha, junto con datos sobre las precipitaciones y el índice de evapotranspiración. Los campos son:
  - ID\_TEMPERATURA: identifica de manera única a cada registro de la tabla.  
**Tipo de dato: INT. Es clave primaria.**
  - ID\_CIUADAD: identifica de manera única a cada ciudad de la tabla.  
**Tipo de dato: INT. Es clave foránea.**
  - LATITUD: número que corresponde a la latitud de cada ciudad.  
**Tipo de dato: FLOAT.**
  - LONGITUD: número que corresponde a la longitud de cada ciudad.  
**Tipo de dato: FLOAT.**
  - PRECIPITACIÓN: número que corresponde a las precipitaciones en milímetros por día de cada ciudad.  
**Tipo de dato: FLOAT.**
  - TEMP MIN: número que corresponde a las temperaturas mínimas en grados Celsius de cada ciudad.  
**Tipo de dato: FLOAT.**
  - TEMP MAX: número que corresponde a las temperaturas máximas en grados Celsius de cada ciudad.  
**Tipo de dato: FLOAT.**
  - TEMP MED: número que corresponde a las temperaturas medias en grados Celsius de cada ciudad.  
**Tipo de dato: FLOAT.**
  - AP TEMP MIN: número que corresponde a la sensación térmica de la temperatura mínima en grados Celsius de cada ciudad.  
**Tipo de dato: FLOAT.**
  - AP TEMP MAX: número que corresponde a la sensación térmica de la temperatura máxima en grados Celsius de cada ciudad.  
**Tipo de dato: FLOAT.**

- AP TEMP MED: número que corresponde a la sensación térmica de la temperatura media en grados Celsius de cada ciudad.  
**Tipo de dato: FLOAT.**
- EVAPOT: número que corresponde al índice de evapotranspiración de cada ciudad en milímetros por día.  
**Tipo de dato: FLOAT.**
- FECHA TEMP: fecha en la que fueron tomadas las mediciones.  
**Tipo de dato: DATE.**
- TEMPERATURA\_BA: contiene todos los registros de temperatura de la ciudad de Buenos Aires por fecha. Los campos son:
  - ID\_TEMP\_BA: identifica de manera única a cada registro de la tabla.  
**Tipo de dato: INT. Es clave primaria.**
  - ID\_CIUDAD: identifica de manera única a cada ciudad de la tabla.  
**Tipo de dato: INT. Es clave foránea.**
  - TEMP MIN: número que corresponde a las temperaturas mínimas en grados Celsius de la ciudad.  
**Tipo de dato: FLOAT.**
  - TEMP MAX: número que corresponde a las temperaturas máximas en grados Celsius de la ciudad.  
**Tipo de dato: FLOAT.**
  - TEMP MED: número que corresponde a las temperaturas medias en grados Celsius de la ciudad.  
**Tipo de dato: FLOAT.**
  - FECHA TEMP BA: fecha en la que fueron tomadas las mediciones.  
**Tipo de dato: DATE.**
- AIRE\_LA: contiene todos los registros sobre la calidad del aire de cada ciudad de Latinoamérica por fecha. Los campos son:
  - ID\_AIRE: identifica de manera única a cada registro de la tabla.  
**Tipo de dato: INT. Es clave primaria.**
  - ID\_CIUDAD: identifica de manera única a cada ciudad de la tabla.  
**Tipo de dato: INT. Es clave foránea.**
  - LATITUD: número que corresponde a la latitud de cada ciudad.  
**Tipo de dato: FLOAT.**
  - LONGITUD: número que corresponde a la longitud de cada ciudad.  
**Tipo de dato: FLOAT.**
  - MON. CARB.: número que corresponde al índice de monóxido de carbono de cada ciudad.  
**Tipo de dato: FLOAT.**
  - DIOX. NITR.: número que corresponde al índice de dióxido de nitrógeno de cada ciudad.  
**Tipo de dato: FLOAT.**
  - OZONO: número que corresponde al índice de ozono de cada ciudad.  
**Tipo de dato: FLOAT.**
  - PM10: número que corresponde al índice de PM 10 de cada ciudad.  
**Tipo de dato: FLOAT.**
  - PM25: número que corresponde al índice de PM 2.5 de cada ciudad.

**Tipo de dato: FLOAT.**

- DIOX. SULF.: número que corresponde al índice de dióxido sulfúrico de cada ciudad.

**Tipo de dato: FLOAT.**

- FECHA AIRE: fecha en la que fueron tomadas las mediciones.

**Tipo de dato: DATE.**

Entre las tablas las relaciones son mediante el campo ID\_CIUDAD que es clave primaria en la tabla CIUDAD y es clave foránea en las demás:

- La tabla CIUDAD se relaciona con la tabla TEMPERATURA\_LA mediante el campo ID\_CIUDAD.
- La tabla CIUDAD se relaciona con la tabla TEMPERATURA\_BA mediante el campo ID\_CIUDAD.
- La tabla CIUDAD se relaciona con la tabla AIRE\_LA mediante el campo ID\_CIUDAD.

### 3. Carga y transformación en Power BI

#### 3.1 Transformaciones realizadas

Al cargar el modelo en Power BI se deben normalizar los datos para su correcto análisis. Para las tablas TEMPERATURA\_LA, AIRE\_LA y TEMPERATURA\_BA el primer paso es asegurarse de la consistencia de los datos y de que cada registro sea único, para esto se quitan filas en blanco y registros duplicados, se definen correctamente los tipos de datos de cada columna y se agrega una columna índice que forma parte del paso intermedio para crear los ID.

Finalmente, en estas tres tablas se crean los siguientes ID para identificar unívocamente a cada registro:

- ID\_TEMPERATURA: tabla TEMPERATURA\_LA
- ID\_TEMP\_BA: tabla TEMPERATURA\_BA
- ID\_AIRE: tabla AIRE\_LA

El paso siguiente es crear la tabla CIUDAD que resume todas las ciudades de la base de datos con su latitud, longitud, país correspondiente y un ID\_CIUDAD.

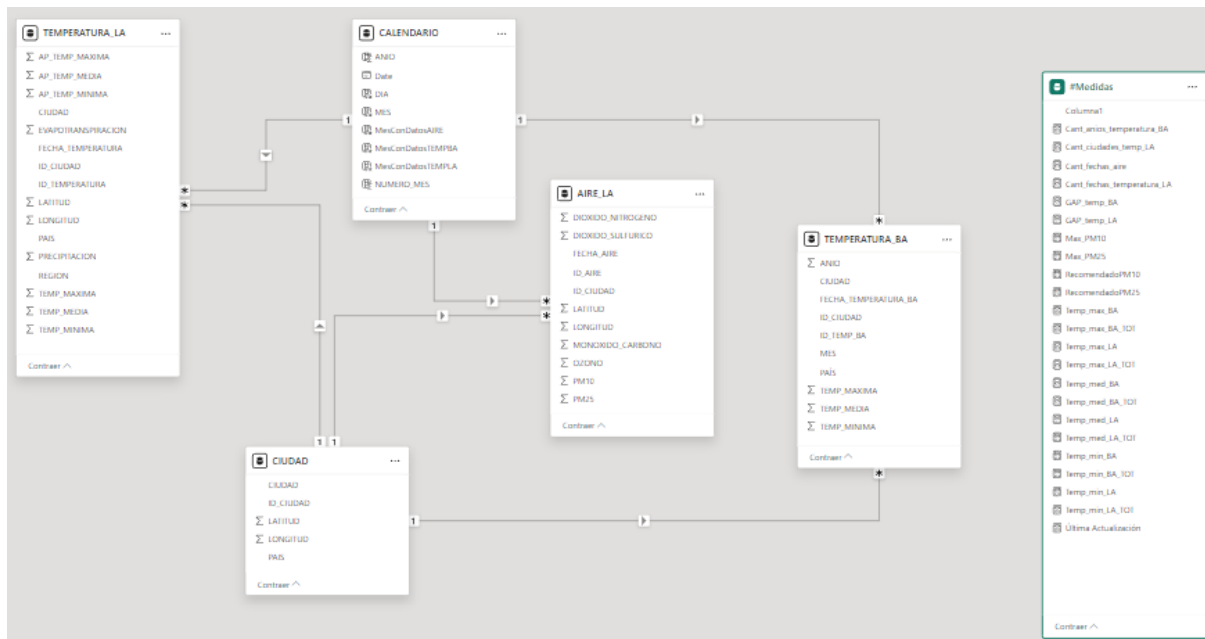
Luego este ID\_CIUDAD es el que se relaciona con las demás tablas: *TEMPERATURA\_LA*, *TEMPERATURA\_BA*, y *AIRE\_LA*.

A continuación, se crea una tabla calendario que contiene la totalidad del rango de fechas de la base de datos y se relaciona con las demás tablas.

También se crea, a modo de apoyo, la tabla REGIONES\_AMERICA que clasifica a los países de América Latina en "América del Sur" y "América Central y el Caribe" con el objetivo de poder segmentar en dos grupos el gran volumen de datos.

Por último, se procede a la crear la tabla #Medidas que contiene las medidas calculadas generadas con sus fórmulas.

Se detalla la vista modelo de Power Bi:



## 3.2 Medidas calculadas

Las medidas calculadas utilizadas en las solapas fueron las siguientes:

- Temp\_max\_LA\_TOT: calcula la temperatura máxima registrada en América Latina independientemente del filtro que se aplique, no se verá afectada.
- Temp\_med\_LA\_TOT: calcula la temperatura media registrada en América Latina independientemente del filtro que se aplique, no se verá afectada.
- Temp\_min\_LA\_TOT: calcula la temperatura mínima registrada en América Latina independientemente del filtro que se aplique, no se verá afectada.
- Temp\_max\_LA: calcula la temperatura máxima registrada en América Latina y se verá afectada por los filtros aplicados.
- Temp\_med\_LA: calcula la temperatura media registrada en América Latina y se verá afectada por los filtros aplicados.
- Temp\_min\_LA: calcula la temperatura mínima registrada en América Latina y se verá afectada por los filtros aplicados.
- Temp\_max\_BA\_TOT: calcula la temperatura máxima registrada en Buenos Aires independientemente del filtro que se aplique, no se verá afectada.
- Temp\_med\_BA\_TOT: calcula la temperatura media registrada en Buenos Aires independientemente del filtro que se aplique, no se verá afectada.
- Temp\_min\_BA\_TOT: calcula la temperatura mínima registrada en Buenos Aires independientemente del filtro que se aplique, no se verá afectada.
- Temp\_max\_BA: calcula la temperatura máxima registrada en Buenos Aires y se verá afectada por los filtros aplicados.



- Temp\_med\_BA: calcula la temperatura media registrada en Buenos Aires y se verá afectada por los filtros aplicados.
- Temp\_min\_BA: calcula la temperatura mínima registrada en Buenos Aires y se verá afectada por los filtros aplicados.
- RecomendadoPM25: es una constante con el valor recomendado por la OMS de partículas en el aire de un diámetro aerodinámico igual o inferior a 2.5  $\mu\text{m}$ .
- RecomendadoPM10: es una constante con el valor recomendado por la OMS de partículas en el aire de un diámetro aerodinámico igual o inferior a 10  $\mu\text{m}$ .
- Gap\_temp\_LA: calcula la diferencia entre la temperatura máxima y mínima registrada en América Latina.
- Max\_PM25: calcula el máximo valor registrado de partículas en el aire de un diámetro aerodinámico igual o inferior a 2.5  $\mu\text{m}$ , en América Latina.
- Max\_PM10: calcula el máximo valor registrado de partículas en el aire de un diámetro aerodinámico igual o inferior a 10  $\mu\text{m}$ , en América Latina.
- Última actualización: muestra fecha y hora de la última actualización del tablero.

## 4. Visualización de datos en Power BI

### 4.1 Solapa "INDICE"

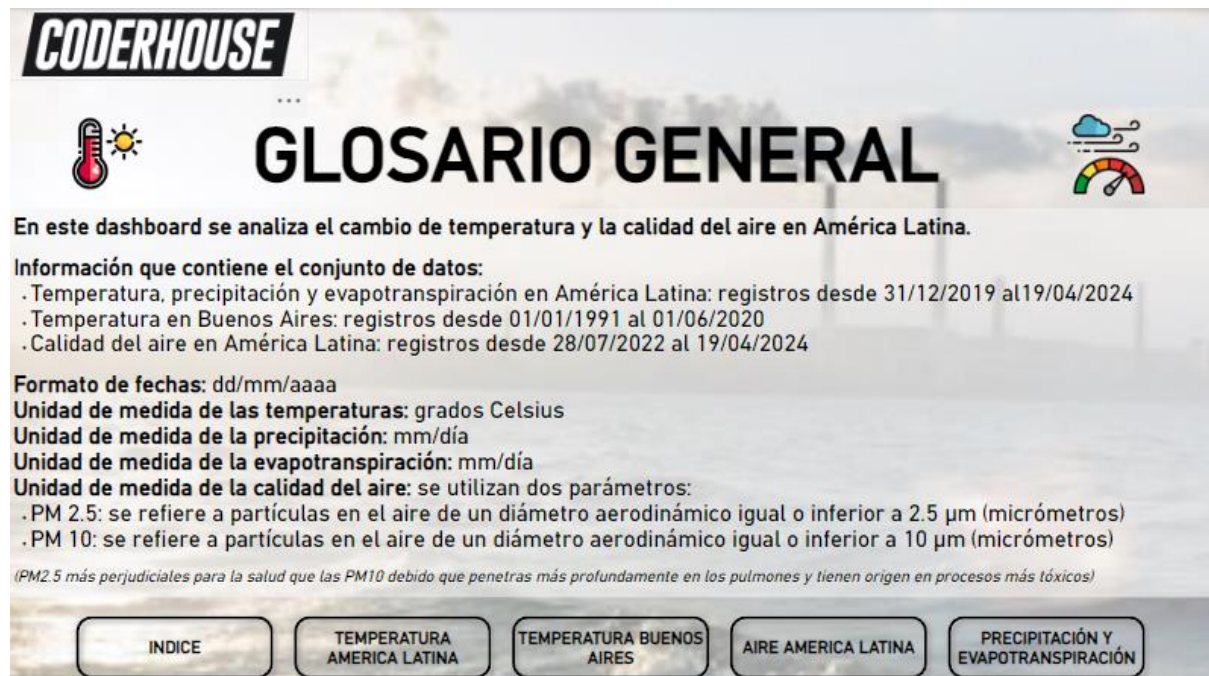
Esta primera solapa contiene el título del proyecto, la hora y fecha de la última actualización que indica rápidamente qué tan reciente se refrescaron los indicadores, logos que hacen referencia a la temática y botones de navegación entre solapas que permiten al usuario navegar fácilmente a través del dashboard.



## 4.2 Solapa "GLOSARIO"

Esta segunda solapa contiene, además de los logos de referencia, un glosario con todos los conceptos utilizados en el dashboard, palabras y significados técnicos sobre la temática elegida que permiten contextualizar el análisis y que sea comprendido por cualquier usuario sin necesidad de ser un experto en el tema.

Además, contiene los botones de navegación entre solapas.



## 4.3 Solapa "TEMPERATURA AMERICA LATINA"

Esta solapa contiene el análisis sobre las temperaturas de América Latina en el período 2019 a 2024.

Contiene el título informativo, el logo, y el botón para regresar al índice.

Se utiliza el patrón Z para el storytelling:

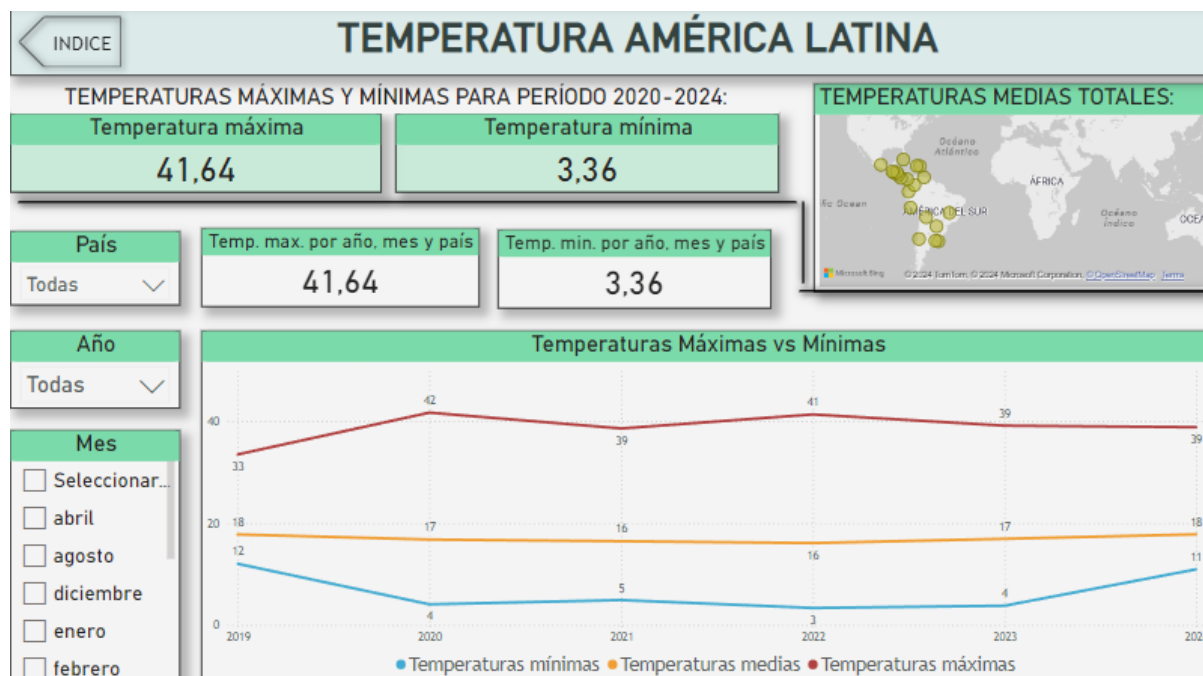
- las tarjetas principales y el mapa que contienen datos sobre el total del período y de los países, están en la parte superior de modo que impactan en mayor medida a primera vista. Dichas tarjetas se refieren a la totalidad de países y del período y no se ven afectadas por los filtros aplicados.
- los filtros de segmentación de datos que permiten filtrar por país, mes y año se encuentran a la izquierda.

Se utiliza un gráfico de líneas ya que se muestra la evolución en el tiempo de la temperatura de los diferentes países del continente.

Tanto el gráfico de líneas como las tarjetas que se encuentran por encima de este, se verán afectados por los filtros aplicados, de manera que el usuario puede interactuar y decidir sobre cuál categoría visualizar los datos. Por ejemplo, sería útil conocer cómo fue la

variación de la temperatura en los países más poblados durante los meses de enero, febrero y marzo dado que se sufre más del calor.

También se implementa un tooltip al gráfico que permite obtener información sobre la diferencia entre la temperatura máxima y mínima según la región del gráfico en donde nos situemos. Esta herramienta da a conocer de manera rápida la amplitud térmica sin necesidad de realizar cálculos mentales.



#### 4.4 Solapa "TEMPERATURA BUENOS AIRES"

Esta solapa contiene el análisis sobre las temperaturas de Buenos Aires en el período 1991 a 2020.

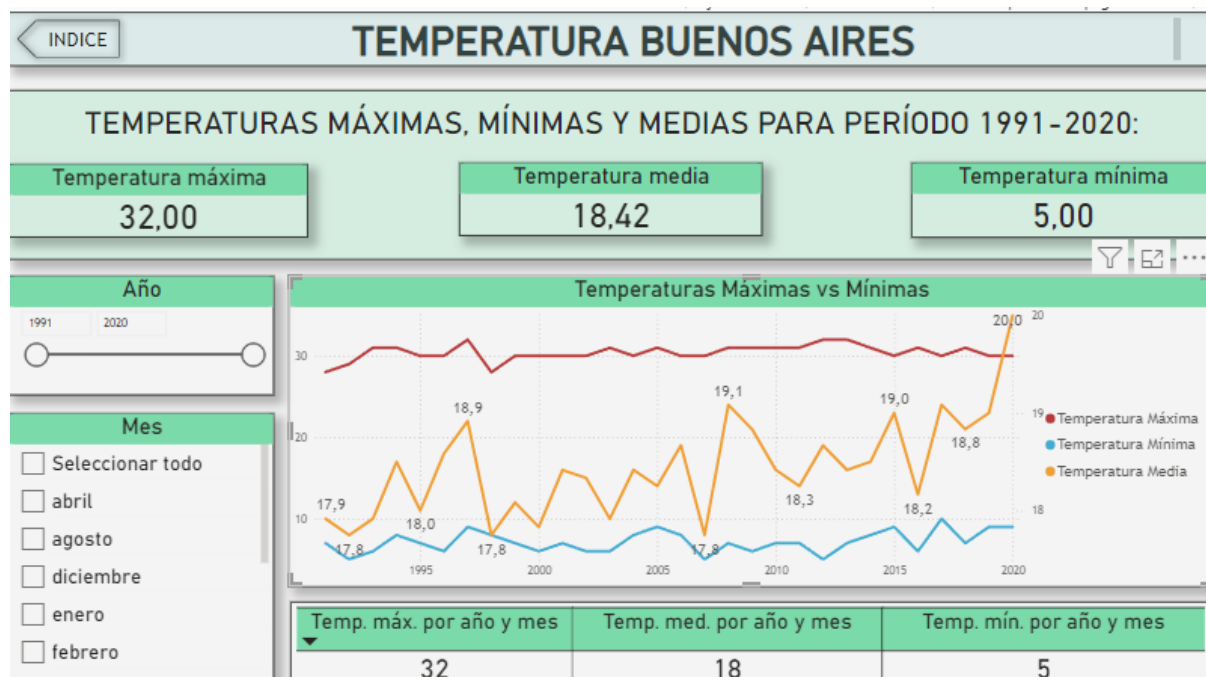
Contiene el título informativo, el logo, y el botón para regresar al índice.

Se utiliza el patrón Z para el storytelling:

- las tarjetas principales que contienen datos sobre el total del período están en la parte superior, no se verán afectadas por los filtros aplicados.
- los filtros de segmentación de datos que permiten filtrar por mes y año se encuentran a la izquierda.

Se utiliza un gráfico de líneas ya que se muestra la evolución en el tiempo de la temperatura de la ciudad.

Tanto el gráfico de líneas como la tabla de temperaturas que se encuentra por debajo de este, se verán afectados por los filtros aplicados, de manera que el usuario puede interactuar con los datos a analizar.



## 4.5 Solapa "AIRE AMERICA LATINA"

Esta solapa contiene el análisis sobre la calidad del aire en América Latina en el período 2022 a 2024

Contiene el título informativo, el logo, y el botón para regresar al índice.

También se utiliza el patrón Z para el storytelling, los filtros de segmentación de datos que permiten filtrar por país, mes y año se encuentran a la izquierda.

Se utilizan gráficos de barras ya que se muestran valores de diferentes categorías, cada categoría representa a un país:

- Los gráficos situados a la izquierda muestran el Top 5 de países con valores máximos más altos de PM 2.5 y PM 10.
- Los gráficos situados a la derecha muestran los valores promedios de estos parámetros para cada país.  
Para estos, se utiliza un tooltip que permite conocer tanto el valor promedio como el máximo de PM 2.5 y PM 10.  
En el caso que estos valores superen los recomendados se pintan de rojo, caso contrario de verde, esto permite conocer rápidamente que tan contaminante es la región sin necesidad de consultar continuamente al valor recomendado.



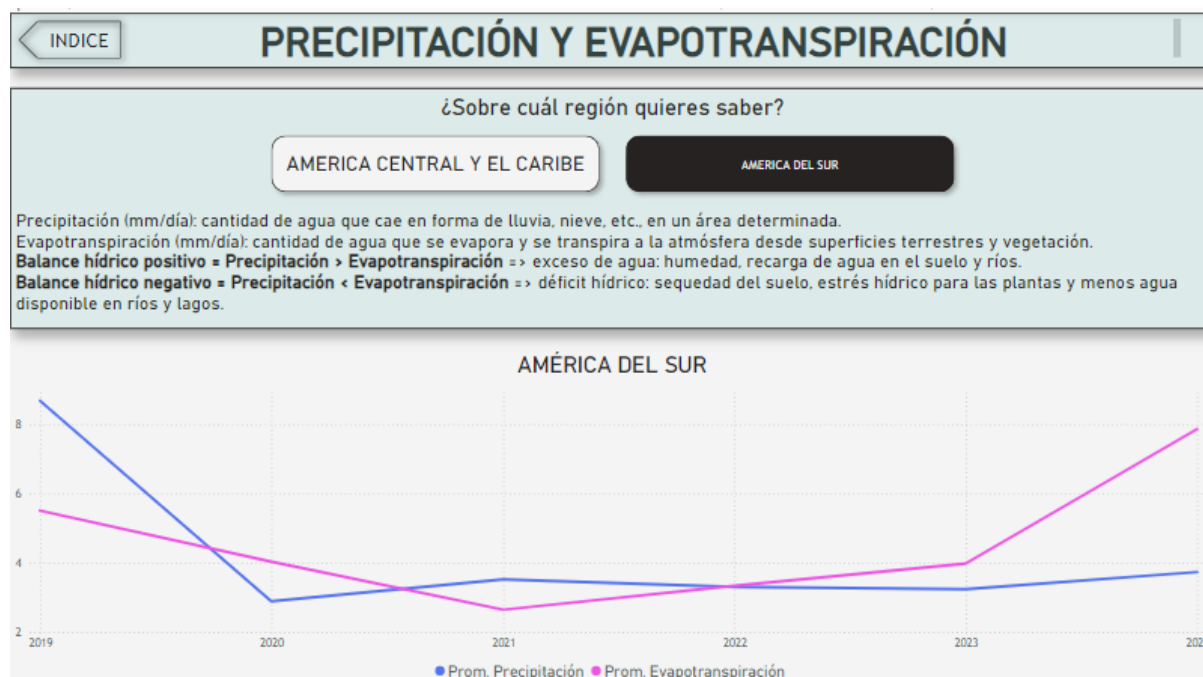
## 4.6 Solapa “PRECIPITACIÓN Y EVAPOTRANSPIRACIÓN”

Esta última solapa introduce al usuario en tema sobre la relación entre los valores de precipitación y evapotranspiración.

Se agrupan los países de Latinoamérica en América del Sur y América Central y el Caribe para segmentar de manera más clara el volumen de datos y volcarla en el gráfico.

Para la visualización se aplica la dinámica de marcadores donde, es el usuario, el que decide qué región desea visualizar.

Un gráfico de líneas es el más adecuado ya que se muestra la evolución de los dos parámetros a lo largo del periodo 2019 a 2024.





## 5. Conclusiones

Repasando la hipótesis, se esperaba observar que el cambio de temperatura sea más pronunciado en regiones con alta densidad poblacional (habitantes por km<sup>2</sup>), lo que sugiere que el calentamiento global contribuye a la degradación ambiental.

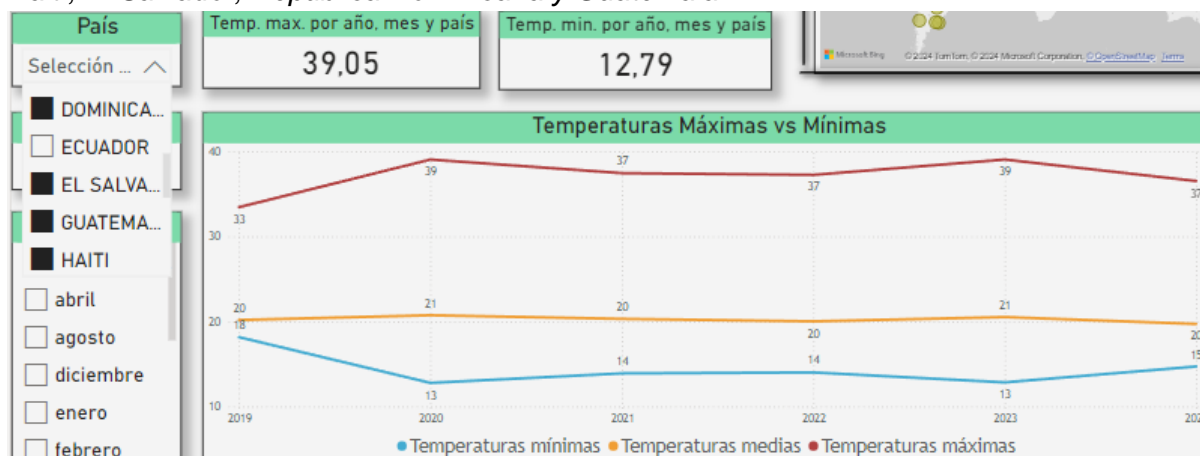
Dentro de América Latina cuatro de los países con mayor densidad poblacional son Haití, El Salvador, República Dominicana y Guatemala, mientras que cuatro de los países con menor densidad son Uruguay, Paraguay, Argentina y Bolivia.

Comparando los siguientes gráficos, en los países con mayor densidad se observa una amplitud térmica menor y mayores valores de la temperatura media.

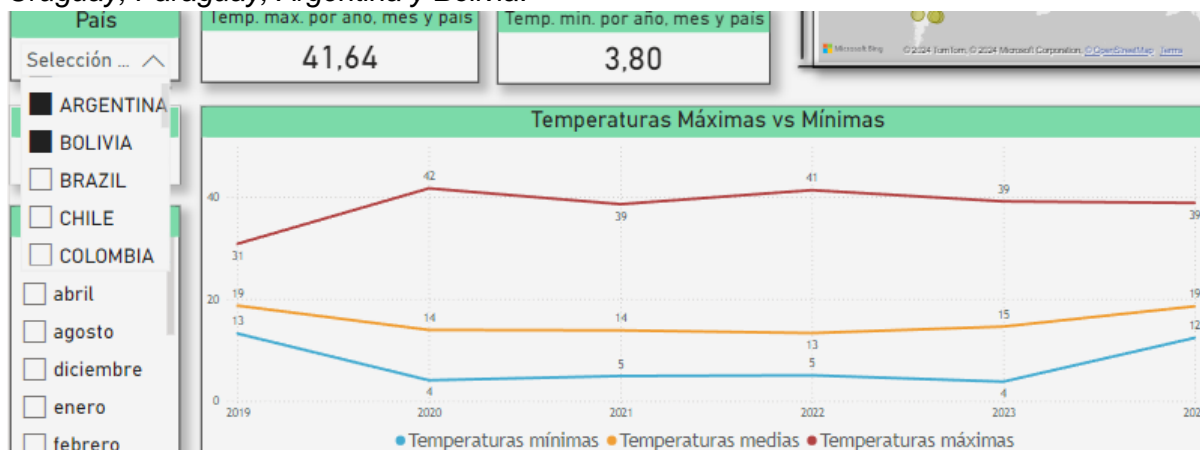
Ambas son manifestaciones del cambio climático, en lugares donde la temperatura media es mayor, el clima tiende a ser más cálido, con noches más templadas y días de calor más intensos.

Es decir que, el calentamiento global, aumenta la temperatura media y, al mismo tiempo, reduce la amplitud térmica diaria.

*Haití, El Salvador, República Dominicana y Guatemala:*



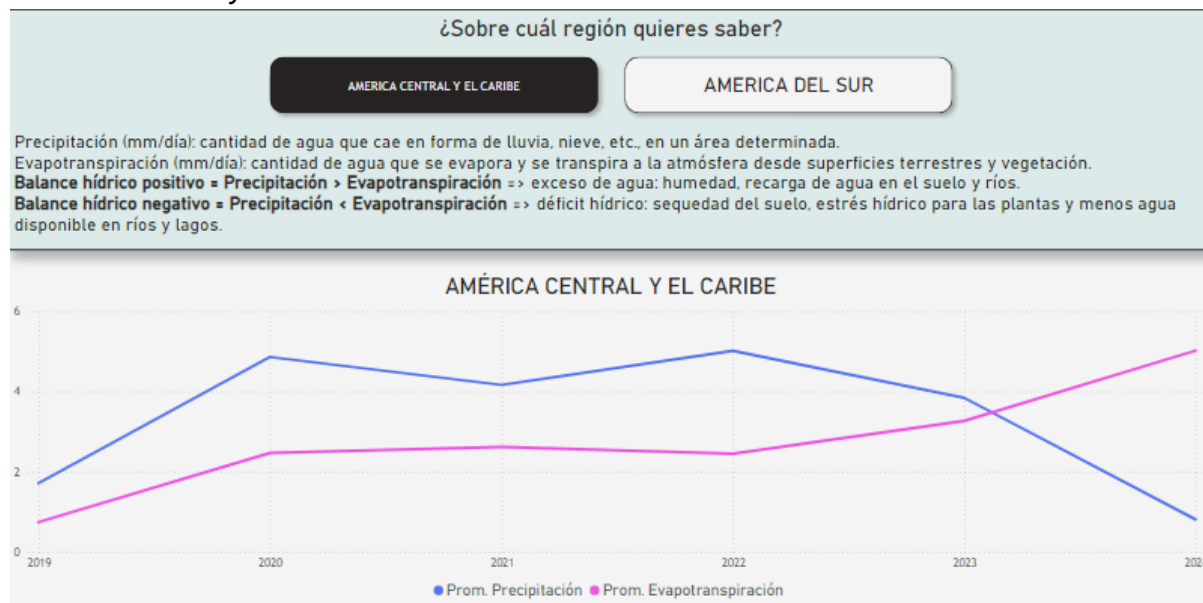
*Uruguay, Paraguay, Argentina y Bolivia:*



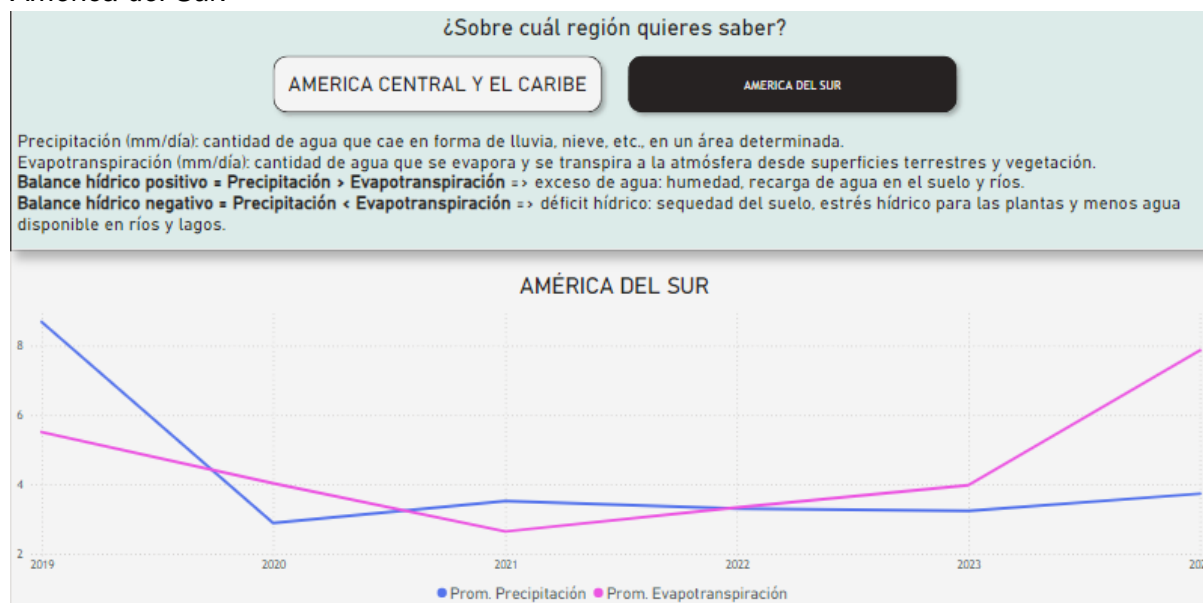
Además, este análisis encuentra relación con la evapotranspiración y la precipitación, una mayor temperatura media genera un aumento de la evapotranspiración y cambia los

patrones de precipitación, esto se visualiza si comparamos los gráficos de América Central y el Caribe (mayor densidad poblacional) vs América del Sur (menor densidad poblacional):

#### América Central y el Caribe:



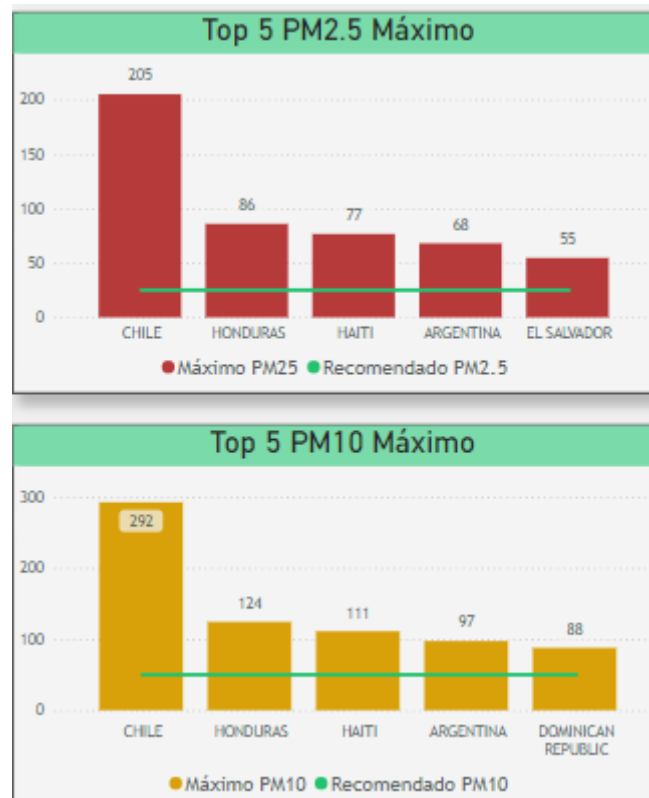
#### América del Sur:



Finalmente, analizando la calidad del aire, en los siguientes gráficos se puede observar el top 5 de países con mayor valor de partículas en el aire PM 2.5 y PM 10.

Tres de estos países son de América Central (Honduras, Haití y El Salvador) lo cual tiene sentido ya que, como vimos anteriormente, son países con mayor densidad poblacional y afectados por el cambio climático.

Los dos países restantes corresponden a América del Sur (Argentina y Chile) lo cual puede corresponderse a que son regiones más desarrolladas que las anteriores y con mayor actividad industrial, factor que afecta a la calidad del aire y la contaminación.



En conclusión, con los datos analizados se podría aceptar la hipótesis planteada, dado que se observa mayor impacto del calentamiento global y de la contaminación en países altamente poblados y con alta actividad industrial.

## 6. Futuras Líneas

Para complementar este proyecto se pueden llevar a cabo diversas iniciativas enfocadas en la evaluación de los datos, la comunicación entre países, y la implementación de planes de mitigación:

- Desarrollo de una plataforma de datos climáticos y ambientales: para centralizar los datos sobre temperatura y calidad del aire. Esto facilitaría el acceso a información en tiempo real y a la toma de decisiones en función de datos históricos y pronósticos.
- Programa de capacitación y cooperación internacional: a líderes políticos y equipos especializados para que sean capaces de implementar estrategias de mitigación. Además, esto ayudaría a promover la colaboración regional y el intercambio de conocimientos.
- Sistemas de alerta temprana y monitoreo de eventos extremos: para prever y mitigar los efectos de eventos climáticos extremos y evitar pérdidas humanas y materiales, además permitiría una mejor planificación de los recursos.
- Planes de adaptación climática por sector: refiriéndonos a sectores claves de la economía, como el sector agropecuario, de energía y salud. Aseguraría que estos sectores puedan seguir operando y creciendo.



- Fomento de inversiones verdes y sostenibles: estimularía el crecimiento económico sostenible y reduciría las emisiones de gases de efecto invernadero, contribuyendo a los objetivos de mitigación a largo plazo.
- Campañas de sensibilización pública: con el fin de crear conciencia pública sobre el cambio climático y su impacto en la región.
- Creación de un fondo regional para la acción climática: poder financiar proyectos de mitigación y adaptación al cambio climático, especialmente en áreas vulnerables y con menos recursos.

Estas iniciativas permitirían a los países latinoamericanos mejorar su capacidad de respuesta al cambio climático y trabajar juntos hacia un futuro más sostenible, basándose en el análisis de datos históricos y en la colaboración entre gobiernos y organizaciones internacionales.

## 7. Bibliografía

[https://www.kaggle.com/datasets/anycaroliny/latin-america-weather-and-air-quality-data?select=LA\\_daily\\_climate.csv](https://www.kaggle.com/datasets/anycaroliny/latin-america-weather-and-air-quality-data?select=LA_daily_climate.csv)

<https://data.buenosaires.gob.ar/dataset/registro-temperatura-ciudad>

*Material brindado por Coderhouse*