# Tendencias Climáticas y Predicción Global de Anomalías de Temperatura

Por: Eslanny Ramírez

2025

## Introducción

El cambio climático se ha consolidado como uno de los desafíos más urgentes del siglo XXI. La evidencia científica muestra un aumento sostenido de la temperatura global y una concentración creciente de gases de efecto invernadero, particularmente dióxido de carbono (CO₂).

Este proyecto busca analizar de manera rigurosa las tendencias históricas de la temperatura y del CO₂ atmosférico, utilizando técnicas de análisis de datos y ciencia de datos. La finalidad es comprender cómo estas variables han evolucionado en el tiempo, identificar patrones y correlaciones, y realizar modelos predictivos que permitan proyectar escenarios futuros.

El trabajo se desarrolla en un entorno profesional con Python, Jupyter Notebooks y Power BI, y se documenta como parte de un portafolio orientado a la comunicación científica y al análisis aplicado.

## Objetivo

Analizar y predecir el comportamiento de la temperatura global en relación con la concentración de CO₂, mediante el uso de técnicas de análisis exploratorio, modelado predictivo y visualización de datos, con el fin de aportar evidencia clara y accesible sobre las tendencias del cambio climático.

## Alcances del proyecto

1. **Exploración de datos históricos (EDA):**
   * Identificación de tendencias por década.
   * Análisis de estacionalidad y anomalías.
   * Visualización de distribuciones, correlaciones y outliers.
2. **Modelado predictivo:**
   * Desarrollo de modelos de series temporales (ARIMA, Prophet).
   * Regresiones multivariables para evaluar la influencia del CO₂ en la temperatura.
   * Evaluación de desempeño con métricas como RMSE y MAPE.
3. **Visualización ejecutiva en Power BI:**
   * Dashboard interactivo con tendencias históricas, proyecciones y comparaciones.
   * Filtros por década, región y variables.
4. **Documentación del proyecto:** 
   * Elaboración de un informe técnico con las fases realizadas, hallazgos y resultados.
   * Preparación de un README para repositorio en GitHub.

## Metodología / Fases del Proyecto

El proyecto se estructura en diferentes fases que representan el flujo completo de un trabajo en ciencia de datos, desde la preparación del entorno hasta la documentación y publicación de resultados. Cada fase se documenta de forma independiente y se relaciona con las demás para mantener una trazabilidad clara.

**FASE 1: Preparación del entorno**

* Configuración de entorno virtual en Python.
* Instalación de librerías necesarias para análisis, visualización y modelado (pandas, seaborn, matplotlib, scikit-learn, statsmodels, prophet, entre otras).
* Estructuración de carpetas para organización del proyecto (data, notebooks, scripts, outputs, powerbi).

**FASE 2: Recolección y exploración de datos (EDA)**

* Obtención de datasets oficiales de fuentes confiables:
  + NOAA Global Land and Ocean Temperature Anomalies (1880–2023).
  + NOAA Mauna Loa CO₂ Monthly Data (1958–2023).
* Limpieza de datos: estandarización de columnas, manejo de valores faltantes y creación de variables derivadas.
* Análisis exploratorio (EDA): tendencias por década, estacionalidad, visualizaciones (líneas, boxplots, heatmaps), correlaciones y detección de outliers.

**FASE 3: Preparación y modelado predictivo**

* Creación de variables adicionales (feature engineering):
  + Promedios por década.
  + Tendencias móviles (rolling mean).
  + Variables de estacionalidad.
* Modelado:
  + Modelos de series temporales (ARIMA, Prophet).
  + Regresiones multivariables (ej. CO₂ → temperatura).
* Evaluación de desempeño de los modelos mediante métricas como RMSE y MAPE.

**FASE 4: Visualización en Power BI**

* Exportación de datasets limpios y resultados de predicciones a formato CSV.
* Construcción de un dashboard interactivo que integre:
  + Evolución histórica de la temperatura global.
  + Niveles de CO₂.
  + Proyecciones futuras.
  + Filtros por década y variables.

**FASE 5: Documentación y comunicación**

* Documentación detallada de cada fase en un informe técnico.
* Preparación de un archivo README para repositorio en GitHub.
* Publicación del caso de estudio en LinkedIn como parte del portafolio profesional, incluyendo visualizaciones clave y reflexiones sobre los resultados.

## Descripción de los Datasets

**1. NOAA Global Land and Ocean Temperature Anomalies**

* **Fuente:** National Centers for Environmental Information (NOAA).
* **Periodo:** 1880–2023 (para este proyecto se toma desde 1958 para coincidir con los registros de CO₂).
* **Cobertura:** Global (tierra y océano).
* **Variable principal:**
  + Anomaly: anomalía de temperatura global expresada en grados Celsius (°C), calculada respecto al promedio del siglo XX (1901–2000).
* **Frecuencia:** mensual, con agregación anual para análisis.
* **Observaciones:** el uso de anomalías permite comparar la evolución en el tiempo eliminando las diferencias de clima base entre regiones.

**2. NOAA Mauna Loa CO₂ Monthly Data**

* **Fuente:** NOAA Earth System Research Laboratory (Mauna Loa Observatory).
* **Periodo:** 1958–2023.
* **Cobertura:** Local (Hawái), pero aceptado como **referencia global de CO₂**.
* **Variables principales:**
  + Average: concentración mensual promedio de CO₂ en la atmósfera (ppm).
  + Year y Month: permiten organizar los datos temporales.
* **Frecuencia:** mensual, con agregación anual para análisis.
* **Observaciones:** incluye valores interpolados para meses sin medición directa.

## Fase 1: Preparación de los Datos

Antes de iniciar el análisis exploratorio, se realizó una etapa de preparación de los datos con el fin de asegurar su consistencia, calidad y usabilidad. Esta fase incluyó la verificación de formatos, el tratamiento de valores inválidos y la creación de variables adicionales que facilitaran el estudio de tendencias a largo plazo.

### Herramientas utilizadas

La preparación de los datos se llevó a cabo en **Python**, utilizando principalmente la librería **Pandas** para manipulación y limpieza de tablas, así como **NumPy** para operaciones numéricas. El trabajo se desarrolló dentro de un entorno virtual configurado en **Jupyter Notebooks**, lo que permitió documentar cada paso y mantener un flujo reproducible.

### Ajustes realizados

* **Conversión de fechas:**  
  En el dataset de CO₂, se construyó una columna date en formato datetime, combinando year y month. Esto permitió establecer un índice temporal y habilitar operaciones como resampling y agrupaciones por año o década.
* **Estandarización de columnas:**  
  Se renombraron y seleccionaron las variables de interés:
  + Year y Anomaly para el dataset de temperatura.
  + year, month y average (rebautizada como co2\_avg) para el dataset de CO₂.
* **Unificación de datasets:**  
  Se creó un DataFrame combinado (df\_comb) utilizando la variable Year como clave. De este modo, fue posible alinear las anomalías de temperatura global con los promedios anuales de CO₂ de Mauna Loa.
* **Creación de variable década:**  
  Para facilitar el análisis de tendencias a largo plazo, se generó la columna decade, agrupando los registros en intervalos de 10 años. Esto permitió analizar la evolución conjunta de temperatura y CO₂ con una visión más amplia.

**Observaciones relevantes**

* El dataset de **temperatura** contiene anomalías calculadas respecto al promedio del siglo XX (1901–2000), por lo que no refleja temperaturas absolutas, sino desviaciones en °C.
* El dataset de **CO₂** proviene de una única estación (Mauna Loa, Hawái), pero se considera una referencia aceptada a nivel global por la estabilidad y continuidad de sus mediciones desde 1958.
* No se detectaron problemas graves de calidad de datos que comprometieran el análisis; las series temporales resultaron estables y consistentes tras la limpieza.

**Estructura recomendada de la documentación**

1. **Portada**
   * Título del proyecto
   * Autora / fecha
   * (Opcional: institución o logo si quieres algo más formal)
2. **Introducción**
   * Contexto del cambio climático
   * Justificación del análisis (por qué es relevante estudiar temperatura y CO₂)
3. **Objetivo**
   * Lo que busca el proyecto (ya lo tienes ✅)
4. **Alcances del proyecto**
   * Lo que cubre y lo que no cubre (ya lo tienes ✅)
5. **Metodología / Fases del Proyecto**  
   Aquí puedes listar las fases que definimos:
   * Preparación del entorno
   * Recolección de datos
   * Limpieza y exploración (EDA)
   * Modelado predictivo
   * Visualización en Power BI
   * Documentación final  
     *(cada fase se detalla más adelante con su desarrollo y resultados)*
6. **Descripción de los Datasets**
   * Fuente de cada dataset (NOAA, etc.)
   * Variables principales
   * Periodo cubierto
   * Observaciones sobre calidad de datos
7. **Preparación de los Datos**
   * Herramientas utilizadas (Python, Pandas, etc.)
   * Ajustes realizados (formato de fechas, índices, eliminación de valores inválidos, creación de columnas como decade)
   * Observaciones relevantes (ejemplo: -99.99 en CO₂, anomalías expresadas respecto al promedio 1901–2000)
8. **Exploración y Análisis (EDA)**
   * Introducción y herramientas usadas (Matplotlib, Seaborn)
   * Gráficos de evolución temporal
   * Comparación conjunta
   * Tendencias por década
   * Boxplots y outliers
   * Correlaciones (heatmap, coeficientes)
   * Hallazgos principales
9. **Modelado Predictivo**
   * Algoritmos utilizados
   * Preparación de variables
   * Resultados y métricas
10. **Visualización en Power BI**

* Dashboard interactivo
* KPIs clave
* Filtros y visualizaciones

1. **Conclusiones y Aprendizajes**

* Síntesis de hallazgos
* Implicancias del análisis
* Posibles mejoras

1. **Referencias**

* Datasets (NOAA, etc.)
* Bibliografía

✅  **FASE 1: Preparación del entorno profesional**

**1.1 Instalación y entorno local**

✅ Requisitos:

* **Python 3.11+**
* **Visual Studio Code** o **Jupyter Lab (preferido en entorno local)**
* **Power BI Desktop**
* **Git** (para control de versiones y subir a GitHub)

📦 Crear un entorno virtual:

bash

CopiarEditar

python -m venv clima\_env

Activar el entorno:

* Windows: clima\_env\Scripts\activate
* Mac/Linux: source clima\_env/bin/activate

Instalar paquetes base:

bash

CopiarEditar

pip install pandas numpy matplotlib seaborn scikit-learn statsmodels prophet jupyter openpyxl

**📁 Estructura de carpetas**

bash

CopiarEditar

clima\_proyecto/

│

├── data/ # Datos originales

├── notebooks/ # Jupyter notebooks por fases

├── scripts/ # Scripts Python reutilizables

├── outputs/ # Tablas y gráficos exportados

├── powerbi/ # Archivo .pbix final

├── README.md # Descripción del proyecto

└── clima\_env/ # Entorno virtual

**📊 FASE 2: Recolección y exploración de datos**

**2.1 Dataset principal (recomendado)**

* **NOAAGlobalTemp**: temperatura global mensual 1880–2023.
* Complemento: CO₂, nivel del mar, lluvia, etc. (Kaggle – Climate Change Dataset).

📌 Lo descargo y te entrego limpio si deseas, o lo limpiamos juntos.

**2.2 Análisis exploratorio (EDA)**

* Tendencias por década
* Análisis de estacionalidad y anomalías
* Visualización: líneas, mapas, boxplots
* Detección de outliers

**🤖 FASE 3: Preparación y modelado predictivo**

**3.1 Feature Engineering**

* Crear variables como:
  + Temperatura promedio por década
  + Incremento anual
  + Estacionalidad mensual
  + Rolling mean (media móvil)

**3.2 Modelos a aplicar**

* **ARIMA** (predicción temporal)
* **Prophet** (predicción temporal + estacionalidad)
* **Regresión multivariable** (ej: CO₂ → temperatura)

📈 Evaluación:

* MAPE
* RMSE
* Visualización de predicción vs real

**📉 FASE 4: Visualización en Power BI**

**Exportar resultados:**

* CSV con predicciones
* Dataset limpio y estructurado

**Dashboard:**

* Mapa del mundo con anomalías térmicas
* Gráficos de tendencias históricas
* Proyecciones climáticas
* Filtros por región, década, variable

**📄 FASE 5: Documentación y publicación profesional**

**5.1 README para GitHub**

* Objetivo, pasos, herramientas
* Imágenes del dashboard
* Cómo ejecutar localmente

**5.2 Publicación LinkedIn**

* Storytelling: contexto del cambio climático
* Gráficos clave
* Reflexión como analista
* Enlace a GitHub y Power BI