# 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**

1. **Fuente:** National Centers for Environmental Information (NOAA).
2. **Periodo:** 1880–2023 (para este proyecto se toma desde 1958 para coincidir con los registros de CO₂).
3. **Cobertura:** Global (tierra y océano).
4. **Variable principal:**
   * Anomaly: anomalía de temperatura global expresada en grados Celsius (°C), calculada respecto al promedio del siglo XX (1901–2000).
   * Year: Año de la medición.
5. **Frecuencia:** mensual, con agregación anual para análisis.
6. **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.
  + Variables adicionaes: decima date, deseasonalized, ndays, sdev, unc.
* **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 3.11**, utilizando las siguientes librerías especializadas:

* Pandas 2.0+: Manipulación y transformación de DataFrames
* NumPy: Operaciones numéricas y manejo de arrays
* Matplotlib: Configuración de visualizaciones
* SciPy: Análisis estadístico y detección de patrones
* Datetime: Manejo y conversión de fechas

El desarrollo se realizó en **Jupyter Notebooks** para mantener un flujo reproducible y documentado. Se configuró un entorno virtual aislado para garantizar la consistencia de las dependencias.

### Ajustes realizados

#### Procesamiento de dataset de CO₂

1. Creación de índice temporal: se crea un índice datetime para que permita le análisis de series temporales en operaciones como resampling, Rolling windows y detección de tendencias estacionales.
2. Selección de variables relevantes: se conserva solo las variables necesarias (year, month y average) eliminando columnas de control técnico que no aportan valor al modelado predictivo.
3. Agregación anual: esto para homologar la frecuencia temporal con el dataset de temperatura (anual), se calculó el promedio anual de CO₂ y permite el análisis de tendencias a largo plazo.

#### Unificación de datasets

Merge por variable temporal común



Se utilizó inner join para conservar únicamente los años con datos disponibles en ambos datasets, garantizando la integridad del análisis. El período resultante (1958 – 2023) proporciona 68 observaciones, suficientes para análisis estadístico robusto.

#### Creación de variables derivadas

1. Variable década: permite análisis de tendencias por períodos históricos y comparación entre eras climáticas (pre-inductrial, industrialización, aceleración reciente).
2. Ordenamiento cronológico: esencial para cálculos de series temporales. Las operaciones diff() u Rolling() requieren orden temporal correcto.
3. Tasas de cambio (primera derivada): las tasas de cambio anuales son mejores indicadores de aceleración climática que los valores absolutos. Permiten identificar períodos de cambio rápido versus estabilidad.
4. Aceleración (segunda derivada): la aceleración detecta cambios en la velocidad del cambio climático, crucial para entender si el problema se esta intensificando exponencialmente.
5. Tendencias móviles suavizadas: en 10 años filtran la volatilidad interanual y revelan patrones de largo plazo. Son fundamentales para modelado predictivo porque reducen el ruido estadístico.

### Observaciones relevantes y calidad de datos

#### Integridad temporal

* Periodo de análisis: 1958 – 2023 (68 años).
* Cobertura: sin valores faltantes en el período unificado.
* Resolución: anual, apropiada para análisis de cambio climático de largo plazo.

#### Características de las variables

* Anomalías de temperatura:
  + Rango: -0.090°C a +1.210°C
  + Media: +0.402°C (Calentamiento neto confirmado).
  + Distribución: sesgo hacia valores positivos (calentamiento reciente).
* Concentración de CO₂:
  + Rango: 315.237 a 428.608 ppm.
  + Incremento total: 113.371ppm (un aumento del 36% en 67 años).
  + Distribución: crecimiento prácticamente monotónico.
* Correlación inicial:
  + Correlación CO₂ - Temperatura: 0.9565
  + Interpretación: relación extremadamente fuerte, justifica el enfoque de modelado predictivo basado en CO₂.

#### Limitaciones identificadas

* Representatividad espacial: CO₂ medido en un solo punto (Mauna Loa), aunque es aceptado como referencia global.
* Resolución temporal: Datos anuales no capturan variabilidad estacional o eventos extremos.
* Variables omitidas: No incluye otros gases de efecto invernadero, aerosoles o factores solares.

#### Calidad general

El dataset resultante presenta alta calidad para análisis de tendencias climáticas:

* Sin valores atípicos extremos que comprometan el análisis.
* Continuidad temporal adecuada para modelado de series.
* Variables derivadas matemáticamente correctas.
* Correlaciones coherentes con la literatura científica.

#### Dataset final

* Archivo generado: df\_comb\_final.csv
* Dimensiones: 68 observaciones × 11 variables
* Variables disponibles: Year, Anomaly, co2\_avg, decade, temp\_change, co2\_change, temp\_acceleration, co2\_acceleration, temp\_trend, co2\_trend

## Fase 2: Análisis Exploratorio de Datos (EDA)

Esta fase se enfoca en extraer insights significativos que permitan validar hipótesis sobre el cambio climático y fundamentar las decisiones metodológicas para el posterior modelado predictivo.

El EDA se estructuró en múltiples dimensiones de análisis: estadísticas descriptivas, evolución temporal, análisis por regímenes climáticos, correlaciones temporales, y detección de patrones de cambio y aceleración. Cada análisis proporciona una perspectiva complementaria que contribuye a una comprensión integral del fenómeno estudiado.

### Herramientas utilizadas

La preparación de los datos se llevó a cabo en **Python 3.11**, utilizando las siguientes librerías especializadas:

* Pandas para manipulación y agregación de datos por períodos.
* NumPy para cálculos estadísticos y operaciones matriciales.
* Matplotlib y Seaborn para generación de visualizaciones.
* SciPy para análisis estadístico avanzado (regresiones, correlaciones, detección de picos).

### Estadísticas descriptivas básicas

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variable | Observaciones | Media | Desv. Estándar | Mínimo | Máximo |
| Año | 68 | 1991.5 | 19.8 | 1958 | 2025 |
| Anomalía Temperatura | 68 | +0.402°C | 0.331°C | -0.090°C | +1.210°C |
| CO₂ Atmosférico | 68 | 360.593 ppm | 33.165 ppm | 315.237 ppm | 428.608 ppm |

Interpretación de estadísticas clave:

* Anomalía de temperatura positiva: La media de +0.402°C confirma un calentamiento neto del planeta durante el período analizado, representando una desviación significativa respecto al promedio histórico del siglo XX.
* Rango de variación: El incremento desde -0.090°C hasta +1.210°C evidencia una aceleración del calentamiento, particularmente notable en la magnitud del valor máximo.
* Concentración de CO₂: El aumento de 113.371 ppm (36% de incremento en 67 años) demuestra una tendencia exponencial que supera significativamente las variaciones naturales históricas.

### Análisis por décadas

#### Evolución de las variables climáticas por períodos decadales

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Década | Anomalía Promedio (°C) | CO₂ Promedio (ppm) | Tasa Cambio Temp. (°C/año) | Tasa Cambio CO₂ (ppm/año) |
| 1950 | 0.0700 | 315.6 | 0.060 | 0.74 |
| 1960 | 0.0190 | 320.3 | 0.000 | 0.86 |
| 1970 | 0.0860 | 330.9 | 0.010 | 1.22 |
| 1980 | 0.2500 | 345.7 | 0.002 | 1.64 |
| 1990 | 0.4450 | 360.6 | 0.019 | 1.53 |
| 2000 | 0.5510 | 378.8 | 0.027 | 1.91 |
| 2010 | 0.7680 | 408.4 | 0.024 | 2.48 |
| 2020 | 0.9983 | 420.6 | 0.010 | 2.83 |

#### Hallazgos por períodos

* Aceleración del calentamiento:
  + 1950s-1960s: Temperaturas cercanas al promedio histórico (0.07°C, 0.02°C)
  + 2020s: Calentamiento de casi 1°C, representando un incremento de 14x respecto a los años 1950
* Crecimiento exponencial del CO₂:
  + Tasa inicial (1950s): 0.74 ppm/año
  + Tasa actual (2020s): 2.83 ppm/año
  + Factor de aceleración: 3.8x en 70 años
* Cambio de régimen climático: Los datos evidencian una transición clara desde condiciones relativamente estables (1950s-1970s) hacia una aceleración sostenida del cambio climático (1990s-2020s).

### Análisis de correlaciones temporales

#### Correlaciones con rezagos temporales

El análisis de correlaciones con diferentes rezagos temporales permite evaluar si existen desfases entre las variaciones de CO₂ y temperatura:

|  |  |
| --- | --- |
| Rezago (años) | Correlación CO₂-Temperatura |
| 0 | 0.9565 |
| 1 | 0.9542 |
| 2 | 0.9560 |
| 3 | 0.9554 |
| 4 | 0.9555 |
| 5 | 0.9537 |

#### Interpretación de correlaciones temporales

* Sincronización inmediata: La correlación más alta se observa en lag 0 (0.9565), indicando que CO₂ y temperatura cambian prácticamente de forma simultánea en escalas anuales.
* Ausencia de rezagos significativos: Las diferencias mínimas entre correlaciones (0.9537-0.9565) sugieren que no hay efectos de rezago temporal pronunciados en el período estudiado.
* Implicaciones para el modelado: La alta correlación sin rezagos justifica el uso de CO₂ contemporáneo como predictor principal de temperatura en modelos de regresión.

### Análisis de regímenes climáticos

#### Caracterización de períodos históricos

|  |  |  |  |
| --- | --- | --- | --- |
| Período | Pendiente Temp. (°C/año) | Pendiente CO₂ (ppm/año) | Correlación |
| Pre-industrial (1958-1970) | -0.0035 | +0.83 | -0.1585 |
| Industrialización (1970-1990) | +0.0159 | +1.48 | 0.6738 |
| Aceleración (1990-2010) | +0.0135 | +1.83 | 0.6346 |
| Crisis climática (2010-2023) | +0.0262 | +2.44 | 0.8852 |

#### Insights por regímenes

* Régimen más crítico: El período 2010-2023 presenta la mayor pendiente de temperatura (0.0262°C/año), confirmando que la década reciente representa la fase de mayor aceleración del cambio climático registrada.
* Período pre-industrial atípico: El período 1958-1970 muestra una pendiente negativa de temperatura (-0.0035°C/año) y correlación negativa (-0.1585), lo que puede reflejar variabilidad natural o efectos de aerosoles industriales que temporalmente enfriaron el planeta.
* Establecimiento de la correlación: A partir de 1970, se establece una correlación positiva consistente entre CO₂ y temperatura, alcanzando su máximo en el período más reciente (0.8852).
* Progresión exponencial del CO₂: La aceleración es clara: de 0.83 ppm/año (1958-1970) a 2.44 ppm/año (2010-2023), representando un incremento de casi 3x en la tasa de emisión.

### Análisis avanzado: Cambios, aceleración y volatilidad

#### Métricas de aceleración climática

* Factor de aceleración del cambio de temperatura:
  + Cambio reciente (2010+): 0.0187°C/año
  + Cambio histórico (≤1970): 0.0076°C/año
  + Factor de aceleración: 2.5x
* Puntos de inflexión identificados:
  + Últimos picos de temperatura: [2010, 2015, 2024]
  + Patrón: Incremento en la frecuencia de años récord de temperatura

#### Análisis de volatilidad

Las visualizaciones de tasas de cambio anuales revelan:

* Períodos de alta volatilidad: 1970s-1980s con fluctuaciones significativas.
* Estabilización reciente: 2000s-2020s muestran cambios más consistentes y unidireccionales.
* Aceleración sostenida: La tendencia móvil de 10 años muestra una pendiente creciente consistente desde 2000, tras un período de volatilidad en los años 1990.

### Visualizaciones clave y hallazgos

#### Distribuciones estadísticas

Temperatura:

* Distribución sesgada hacia valores positivos (calentamiento).
* Concentración de datos recientes en el extremo superior del rango.

CO₂:

* Distribución bimodal que refleja la aceleración post-1980
* Crecimiento prácticamente monotónico sin reversiones

#### Análisis por décadas (boxplots)

Los boxplots revelan una progresión clara:

* Expansión de rangos: Cada década muestra mayor variabilidad que la anterior.
* Desplazamiento de medianas: Incremento sistemático de valores centrales.
* Outliers mínimos: Alta consistencia en las tendencias observadas.

#### Matriz de correlaciones

Correlaciones principales:

* CO₂ vs Temperatura: 0.957 (relación extremadamente fuerte).
* Año vs variables climáticas: >0.94 (confirma tendencias temporales claras).
* Década vs variables: >0.98 (validación de agrupación temporal).

### Conclusiones del EDA

#### Validación de hipótesis principales

* Relación CO₂-Temperatura confirmada: Correlación de 0.9565 valida la hipótesis de asociación fuerte.
* Aceleración del cambio climático: Factor de 2.5x en tasas de cambio de temperatura.
* Sincronización temporal: Ausencia de rezagos significativos en escalas anuales.
* Progresión no lineal: Evidencia de aceleración exponencial en ambas variables.

#### Implicaciones para el modelado predictivo

Justificación metodológica:

* La alta correlación (0.957) justifica modelos de regresión CO₂→Temperatura.
* La ausencia de rezagos permite usar valores contemporáneos de CO₂.
* Las tendencias claras facilitan la aplicación de modelos de series temporales (ARIMA, Prophet).

#### Variables predictoras óptimas:

* CO₂ atmosférico como predictor principal.
* Variables de tendencia temporal como predictores secundarios.
* Consideración de regímenes climáticos para modelado segmentado.

### Preparación para Fase 3

El EDA proporciona la base sólida necesaria para el desarrollo de modelos predictivos robustos:

* Dataset limpio y caracterizado (68 observaciones, 11 variables)
* Relaciones cuantificadas y validadas estadísticamente
* Identificación de patrones temporales apropiados para proyecciones futuras
* Fundamentación científica para la selección de algoritmos de modelado

Dataset final optimizado: df\_comb\_final.csv con todas las variables derivadas y validadas, listo para la implementación de modelos de predicción climática a 5 años.

**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