# PROYECTO FINAL ANÁLISIS DE DATOS – EXPLORADOR

Richard Vélez  
Andrés Márquez  
Rosibel Panesso  
Juan Sebastián Fajardo

**Titulo:**

**Análisis de las tendencias de emisiones de CO2 por sector en Latinoamérica antes y después de implementar políticas de mitigación entre el 2000 y 2023**

## Identificación del Problema:

## Este proyecto aborda la necesidad crítica de analizar las tendencias de emisiones de CO2 en Latinoamérica entre 2000 y 2023. Se enfoca en analizar el impacto de las políticas de mitigación del cambio climático, identificando a los países con mayores emisiones y aquellos que han logrado reducciones significativas, con el fin de proporcionar información valiosa para la formulación de políticas climáticas más efectivas.

**Objetivo General:**

Analizar las tendencias de emisiones de dióxido de carbono (CO2) por sector en Latinoamérica entre los años 2000 y 2023, con el fin de identificar cambios significativos posteriores a la implementación de políticas de mitigación del cambio climático en la región.

**Objetivos Específicos:**

* Recolectar datos históricos de emisiones por país y sector desde fuentes confiables
* Limpiar los conjuntos de datos para obtener una mejor calidad de las fuentes, manejando valores faltantes, inconsistencias y outliers.
* Analizar las tendencias de emisiones de CO2 por sector (petróleo, gas, carbón, etc.) para cada país de Latinoamérica entre 2000 y 2023.
* Comparar los niveles de emisiones de CO2 por sector entre los diferentes países de Latinoamérica a lo largo del periodo de estudio.
* Analizar los datos después de las políticas de mitigación implementadas en la región, comparando las tendencias de emisiones antes y después de su adopción.
* Elaborar visualizaciones claras y concisas que permitan identificar patrones y tendencias en las emisiones.
* Elaborar un informe final que resuma los hallazgos clave, las conclusiones del análisis y las recomendaciones para futuras políticas de mitigación en Latinoamérica.

**Bases de datos:**

**-** <https://ourworldindata.org/co2-and-greenhouse-gas-emissions>

-(Completar con las otras)

## Alcance / Actividades principales del proyecto:

|  |  |
| --- | --- |
| **Actividad** | **Descripción** |
| Recolección de datos | Recolección de datos históricos de emisiones por país y sector desde fuentes confiables (por ejemplo, la Agencia Internacional de Energía, el Banco Mundial, etc.). |
| Limpieza y procesamiento de datos | Revisión, depuración y estandarización de los conjuntos de datos para asegurar consistencia y calidad. |
| Análisis exploratorio | Evaluación de tendencias generales, anomalías y patrones por país y por sector. |
| Estadísticas descriptivas | Cálculo de medidas como medias, medianas, tasas de cambio y rankings para determinar qué países emiten más y cuáles han logrado reducciones significativas. |
| Análisis de cambios | Comparación de datos antes y después de la implementación de medidas de mitigación (por ejemplo, protocolos internacionales, reformas energéticas). |
| Visualización de datos | Creación de gráficos, que muestren claramente las tendencias y diferencias por regiones y sectores. |
| Informe final | Elaboración de un informe con hallazgos, conclusiones y recomendaciones para políticas de mitigación más efectivas. |

## Entregables:

|  |  |
| --- | --- |
| **Entregable** | **Descripción / Detalles** |
| Conjunto de datos procesados | Datos limpios y organizados listos para análisis, con metadatos documentados. |
| Informe de análisis | Documento con los hallazgos clave, metodología, análisis comparativo y resultados de estadísticas descriptivas. |

## Resumen del cronograma / Principales hitos:

|  |  |  |
| --- | --- | --- |
| Hito | Fecha estimada de finalización | Descripción / Detalles |
| Recolección de datos |  | Obtención de todos los datos necesarios. |
| Procesamiento de datos |  | Datos organizados, limpios y listos para análisis. |
| Análisis exploratorio y descriptivo |  | Evaluación de tendencias, distribución y variaciones entre países. |
| Visualizaciones finales |  | Gráficos listos para entrega. |
| Informe final |  | Documento final con recomendaciones. |

## Fecha estimada de finalización:

09 de Mayo de 2025

**EXPOSICIÓN:**

**(EN AMARILLO YA ESTÁ LISTO)**

**(EN AZUL INCOMPLETO EN MI SCRIPT)**

**(EN ROSADO FALTA EN MI SCRIPT)**

Titulo del proyecto

Objetivo General

Objetivos específicos

Identificación del problema

Limpieza de datos

* **Estado Actual**: El script de Python ya incluye pasos de limpieza:
  + Eliminación de columnas innecesarias ('Annual CO₂ emissions.1') en anual\_co\_emissions\_region.
  + Eliminación de filas con valores nulos en las columnas clave ('Annual CO₂ emissions', 'Year', y otras).
  + Conversión del tipo de dato de la columna 'Year' a entero.
  + Filtrado de los DataFrames para incluir solo datos desde el año 2000 en adelante.
* **¿Qué falta?**
  + **Documentar**: Es crucial que documentes *cada* paso de limpieza en tu informe. Explica por qué eliminaste ciertas columnas o filas, cómo manejaste los valores nulos y por qué realizaste cada conversión de tipo de datos.
  + **Valores atípicos (Outliers)**: Considera si necesitas identificar y manejar valores atípicos, especialmente en las columnas de emisiones. Puedes usar métodos como el rango intercuartílico (IQR) o la desviación estándar para detectarlos. Sin embargo, decide si eliminarlos, transformarlos o dejarlos como están, justificando tu decisión.
  + **Consistencia de datos**: Verifica la consistencia de los nombres de los países/regiones en los diferentes DataFrames. Si hay variaciones (por ejemplo, "Estados Unidos" vs. "USA"), unifícalos para facilitar el análisis.

Análisis exploratorio

 **Estado Actual**: El script ya incluye algo de análisis exploratorio:

* anual\_co.info(), bysector\_.info(), y greenhouse\_.info() para obtener información general de los DataFrames.
* anual\_co.shape, anual\_co.columns, bysector\_.shape, bysector\_.columns para ver las dimensiones y columnas de los DataFrames.
* anual\_co['Entity'].unique(), bysector\_['Entity'].unique() para obtener los valores únicos de países/regiones.
* anual\_co.head(), bysector\_.head(), y greenhouse\_.head() para ver las primeras filas de los DataFrames.
* Impresión de los años mínimos y máximos en los DataFrames.

 **¿Qué falta?**

* **Visualizaciones**: Crea gráficos para explorar los datos. Algunas sugerencias:
  + Histogramas o diagramas de caja para visualizar la distribución de las emisiones.
  + Gráficos de líneas para ver las tendencias de las emisiones a lo largo del tiempo por país/región y sector.
  + Mapas de calor para visualizar la correlación entre diferentes sectores de emisión.
* **Resúmenes estadísticos**: Utiliza describe() para obtener estadísticas descriptivas de las columnas numéricas (media, mediana, desviación estándar, cuartiles, etc.).
* **Análisis de frecuencia**: Cuenta la frecuencia de los países/regiones para identificar los más comunes.

Análisis descriptivo

 **Estado Actual**: Ya hemos comenzado con esto al calcular estadísticas básicas y explorar los datos.

 **¿Qué falta?**

* **Profundizar en las estadísticas**:
  + Calcula las emisiones totales por país/región y sector.
  + Determina las tasas de crecimiento anual de las emisiones.
  + Identifica los países con las emisiones más altas y más bajas.
  + Calcula la mediana y la moda de las emisiones para comparar con la media.
* **Visualizaciones específicas**:
  + Gráficos de barras para comparar las emisiones totales entre países/regiones.
  + Gráficos de líneas para mostrar las tendencias de las emisiones a lo largo del tiempo, con anotaciones de las políticas de mitigación clave.
* **Análisis de series de tiempo**: Si es posible, descompón las series de tiempo de las emisiones en tendencia, estacionalidad y residuos para comprender mejor los patrones.

Análisis estadístico

* **Estado Actual**: No hay análisis estadístico inferencial en el script
* **¿Qué falta?**
  + **Definir qué análisis estadístico es relevante para tus objetivos**:
    - **Pruebas de hipótesis**: Si quieres determinar si hay una diferencia *significativa* en las emisiones antes y después de las políticas de mitigación, podrías usar una prueba t de muestras pareadas o una prueba de Wilcoxon (si los datos no son normales).
    - **Correlación**: Calcula la correlación entre las emisiones de diferentes sectores para ver si están relacionadas.
    - **Regresión**: Si tienes datos suficientes, podrías intentar construir un modelo de regresión para predecir las emisiones futuras basadas en las tendencias pasadas.
  + **Implementar las pruebas en Python**: Usa las funciones de la biblioteca scipy.stats para realizar las pruebas estadísticas.
  + **Interpretar los resultados**: Explica qué significan los resultados de las pruebas en el contexto de tu análisis. Indica si las diferencias observadas son estadísticamente significativas y qué implicaciones tienen.

Imputación de datos si se require

* **Estado Actual**: Ya hemos eliminado algunos valores nulos, pero no hemos realizado imputación.
* **¿Qué falta?**
  + **Evaluar la necesidad de imputación**: Después de la limpieza inicial, determina si quedan muchos valores nulos que podrían afectar tu análisis. Si es así, considera la imputación.
  + **Elegir métodos de imputación**:
    - Para series de tiempo, la imputación con el último valor observado (forward fill) o el siguiente valor observado (backward fill) podría ser apropiada.
    - Para otros datos, la imputación con la media, la mediana o el valor más frecuente podría ser adecuada, dependiendo de la distribución de los datos.
    - Métodos más avanzados como la imputación por regresión o la imputación múltiple también son opciones, pero pueden ser más complejos.
  + **Implementar la imputación en Python**: Usa las funciones de pandas como fillna() para imputar los valores nulos.
  + **Justificar tu elección**: Explica por qué elegiste un método de imputación particular y cómo podría afectar los resultados de tu análisis.

Conclusiones y recomendaciones

 **Estado Actual**: Esta sección aún no está desarrollada.

 **¿Qué falta?**

* **Resumir los hallazgos clave**: Resume los principales resultados de tu análisis descriptivo y estadístico.
* **Responder a los objetivos del proyecto**: Indica claramente cómo tus hallazgos responden a los objetivos específicos y al objetivo general del proyecto.
* **Discutir las implicaciones**: Explica qué significan tus hallazgos en términos de las tendencias de emisiones de CO2 en Latinoamérica y la efectividad de las políticas de mitigación.
* **Recomendaciones**: Basado en tus hallazgos, proporciona recomendaciones para futuras políticas de mitigación, investigaciones o acciones.
* **Limitaciones del estudio**: Reconoce cualquier limitación en tus datos o metodología que podría haber afectado tus resultados.
* **Futuras investigaciones**: Sugiere posibles direcciones para futuras investigaciones relacionadas con tu proyecto.

**ESTRUCTURA SUGERIDA DEL INFORME**

Para organizar tu entrega, te sugiero la siguiente estructura de informe:

1. **Título del Proyecto**
2. **Introducción**
   * Contexto del problema de las emisiones de CO2 en Latinoamérica.
   * Importancia del análisis de las políticas de mitigación.
   * Propósito del proyecto (de tu documento DOCX).
3. **Objetivo General**
4. **Objetivos Específicos**
5. **Identificación del Problema**
6. **Datos y Metodología**
   * Descripción de las fuentes de datos.
   * Descripción de las variables utilizadas.
   * Detalles de la metodología de análisis (limpieza de datos, análisis exploratorio, análisis descriptivo, análisis estadístico, imputación).
7. **Resultados**
   * Análisis Exploratorio (con visualizaciones).
   * Análisis Descriptivo (con visualizaciones).
   * Análisis Estadístico (si aplica, con interpretación).
8. **Discusión**
   * Interpretación de los resultados en el contexto del problema.
   * Comparación con otros estudios (si es posible).
9. **Conclusiones**
   * Resumen de los principales hallazgos.
   * Respuesta a los objetivos del proyecto.
10. **Recomendaciones**
    * Recomendaciones basadas en los hallazgos.
11. **Limitaciones del Estudio**
12. **Futuras Investigaciones**
13. **Referencias**
14. **Apéndices**
    * Código de Python (bien comentado).
    * Tablas adicionales o visualizaciones.

**Consejos Adicionales**

* **Visualizaciones**: Las visualizaciones son clave para comunicar tus hallazgos. Asegúrate de que sean claras, informativas y estén bien etiquetadas.
* **Documentación del Código**: Comenta tu código de Python para que sea fácil de entender. Explica el propósito de cada paso y las decisiones que tomaste.
* **Enfoque en Latinoamérica**: Dado que el enfoque de tu proyecto es Latinoamérica, asegúrate de que tu análisis y tus conclusiones se centren en esta región.
* **Políticas de Mitigación**: Investiga las principales políticas de mitigación implementadas en Latinoamérica durante el período 2000-2023. Esto te ayudará a contextualizar tu análisis y a hacer recomendaciones más informadas.
* **Revisión**: Dedica tiempo a revisar y editar tu informe antes de entregarlo. Verifica la gramática, la ortografía, la claridad y la coherencia.