



Universidad de la Habana
Facultad de Matemática y Computación

PROYECTO DE FIN DE SEMESTRE

Autores:

- Mariam Parra
- David Michel García
- Kevin Corbo
- Luis Manuel Notario

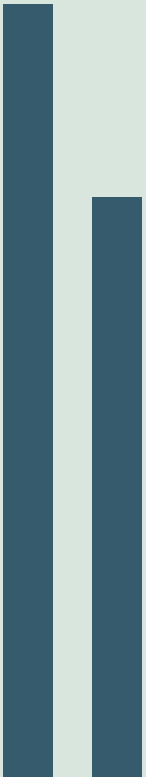
Ciencia de Datos

Asignatura:

Análisis Exploratorio
de Datos



ÍNDICE

1. Introducción
 2. Objetivos del estudio
 3. Metodología aplicada
 4. Resultados obtenidos
 5. Discusión de hallazgos
 6. Conclusiones y futuras líneas de investigación
- 

INTRODUCCIÓN

El Análisis Exploratorio de Datos (AED) es una etapa esencial en la investigación basada en datos, ya que permite detectar patrones, relaciones y posibles anomalías antes de aplicar modelos estadísticos o predictivos. Este estudio analiza el conjunto de datos airquality, que contiene observaciones meteorológicas y de contaminación atmosférica entre mayo y septiembre.

Mediante técnicas de visualización y estadística descriptiva, se explora la relación entre la temperatura, el ozono y otros factores ambientales. Esto permite no solo comprender la variabilidad de la calidad del aire, sino también generar hipótesis sobre su evolución a lo largo del tiempo. Los resultados pueden servir como referencia para estudios actuales sobre contaminación y cambio climático.



OBJETIVOS

General



Realizar un análisis exploratorio de datos (AED) sobre las variables meteorológicas y de calidad del aire del dataset airquality, identificando patrones, correlaciones y tendencias que permitan una mejor comprensión de los factores ambientales que influyen en la contaminación atmosférica.

Específicos



- Examinar la distribución y variabilidad de las variables atmosféricas mediante estadísticas descriptivas y visualizaciones gráficas.
- Identificar relaciones significativas entre temperatura, radiación solar, viento y niveles de ozono utilizando análisis de correlación.
- Aplicar técnicas de limpieza y transformación de datos para mejorar la calidad del análisis exploratorio.
- Evaluar la relación entre la temperatura y el ozono mediante un modelo de regresión lineal, analizando su ajuste y validez estadística.
- Interpretar los resultados obtenidos en el contexto ambiental y de contaminación, formulando hipótesis sobre los efectos de factores climáticos en la calidad del aire.

METODOLOGÍA

Carga del dataset

- Se utilizó el conjunto de datos airquality disponible en R Studio.
- Se revisó su estructura, cantidad de observaciones y posibles valores faltantes.

Limpieza de datos

- Se identificaron valores ausentes en Ozone y Solar.R.
- Se eliminaron las filas con valores faltantes para garantizar la calidad del análisis.

Exploración y visualización de datos

- Se calcularon medidas estadísticas y se generaron gráficos para analizar patrones.
- Se evaluaron correlaciones entre variables.

Análisis de correlación y modelado estadístico

- Se estudiaron correlaciones entre variables meteorológicas y calidad del aire.
- Se aplicaron modelos estadísticos para evaluar patrones y dependencias.

RESULTADOS OBTENIDOS



Distribución de las variables

- La temperatura mostró un aumento progresivo a lo largo del verano, alcanzando valores máximos en julio.
- La radiación solar presentó una distribución bimodal, reflejando la existencia de dos regímenes climáticos diferenciados.



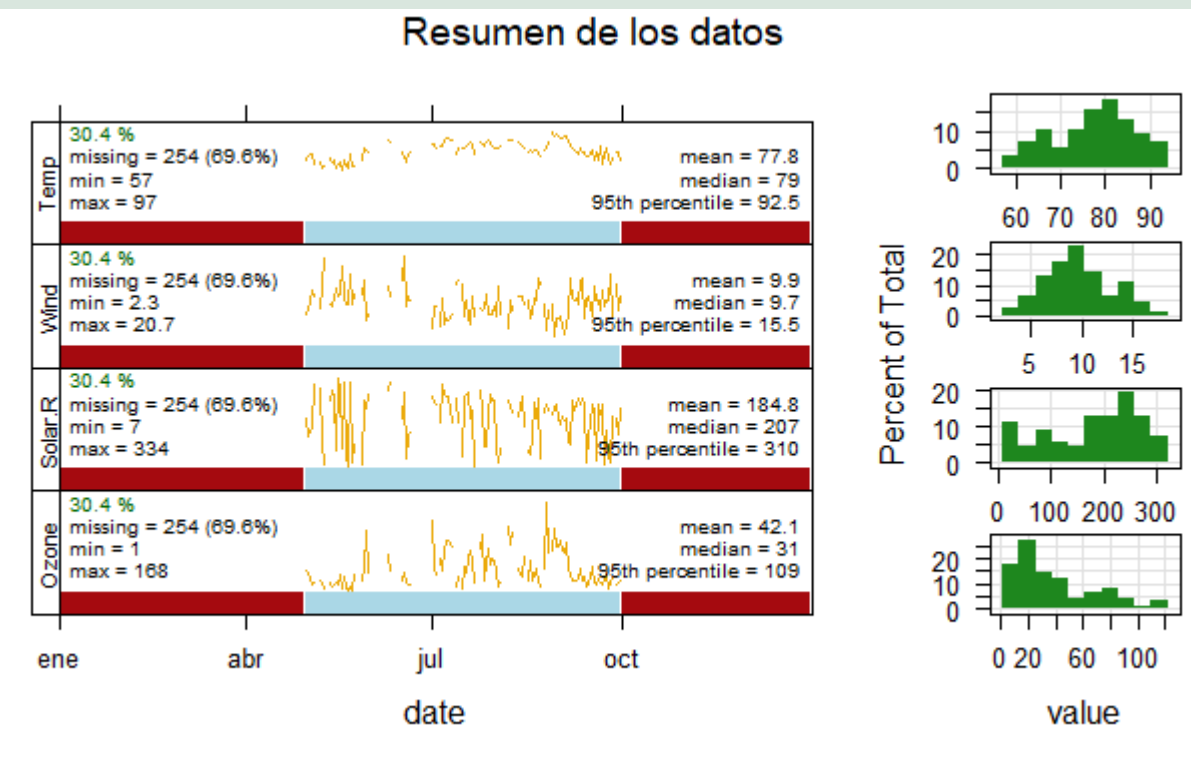
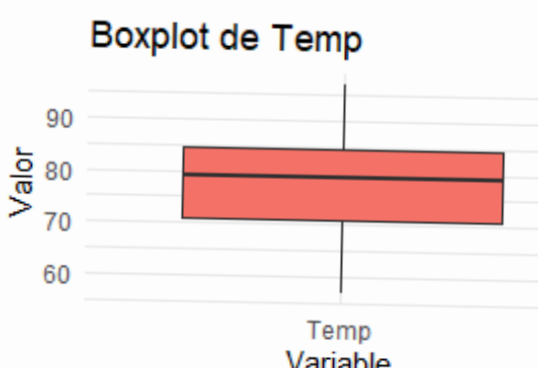
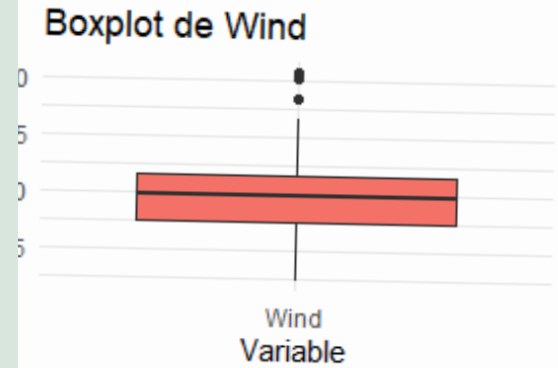
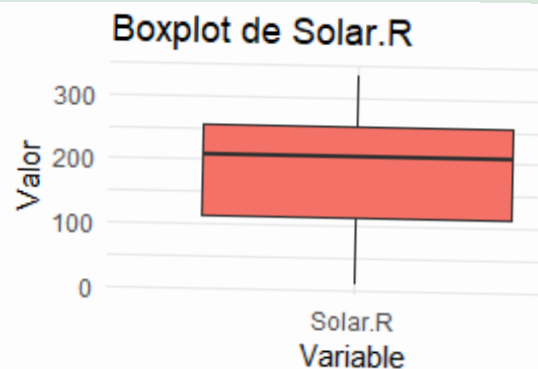
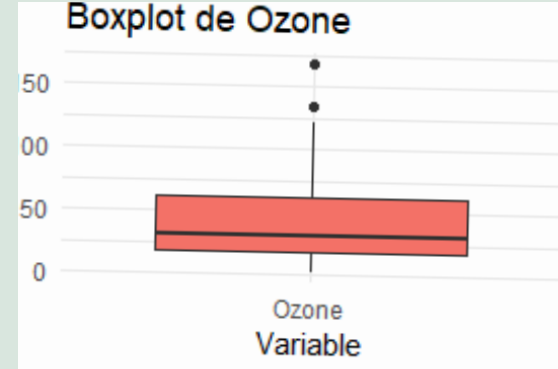
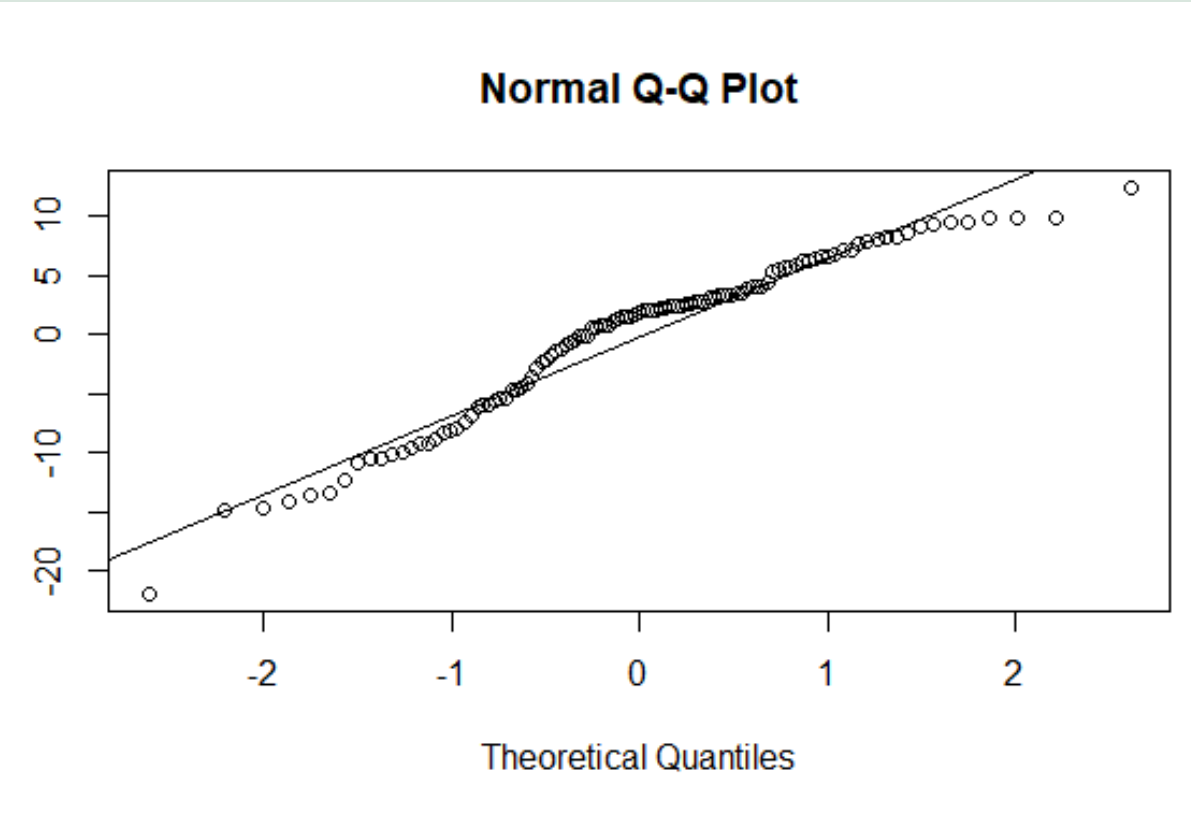
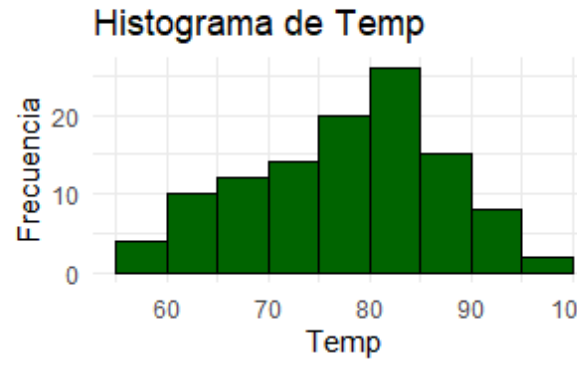
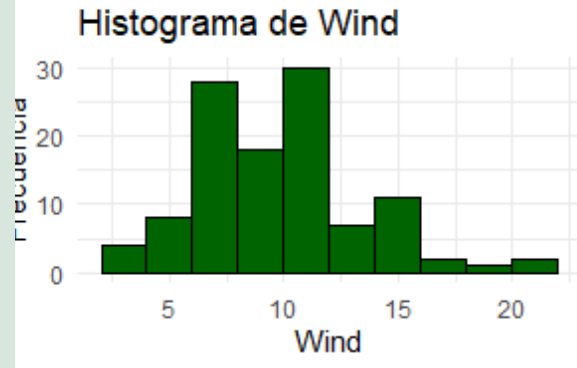
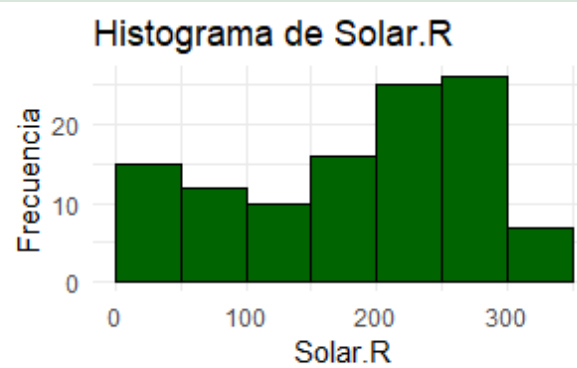
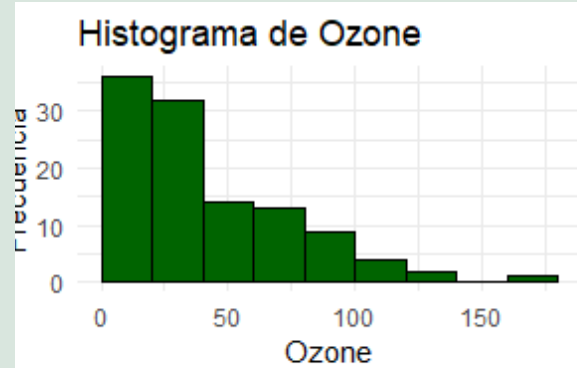
Relación entre temperatura y ozono

- Se encontró una correlación significativa entre ambas variables ($R^2 = 0.48$), indicando que temperaturas más altas pueden estar asociadas con mayores niveles de ozono.



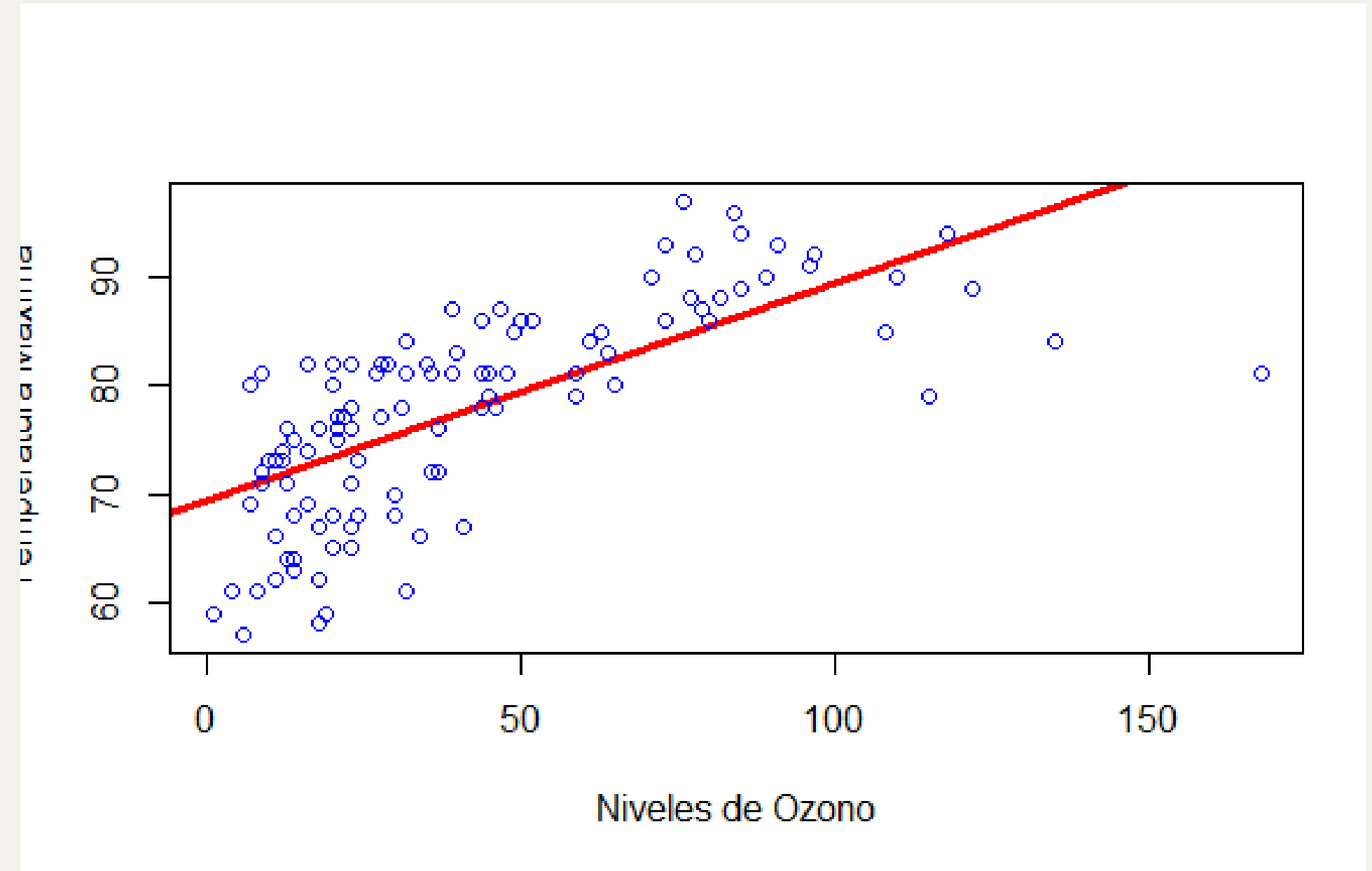
Influencia de factores meteorológicos

- El viento mostró variaciones extremas que pueden influir en la dispersión de contaminantes.
- Se identificó una posible conexión entre radiación solar y niveles de ozono, sugiriendo un efecto ambiental en la calidad del aire.



DISCUSIÓN DE HALLAZGOS

Los resultados indican que la temperatura tiene una relación significativa con los niveles de ozono, lo que sugiere que condiciones más cálidas pueden favorecer su acumulación en la atmósfera. La bimodalidad en la radiación solar y el ozono sugiere la existencia de regímenes ambientales diferenciados, posiblemente influenciados por la variabilidad estacional. Además, la influencia del viento en la dispersión de contaminantes refuerza la importancia de los factores meteorológicos en la calidad del aire. Estos hallazgos son consistentes con estudios sobre contaminación atmosférica y pueden ser clave para evaluar el impacto del cambio climático en la distribución de ozono.



CONCLUSIONES

El análisis exploratorio de datos permitió identificar patrones atmosféricos clave y su relación con la calidad del aire. La correlación entre temperatura y ozono sugiere que condiciones más cálidas pueden favorecer mayores concentraciones de contaminantes. La presencia de bimodalidad en la radiación solar y el ozono revela regímenes ambientales diferenciados, lo que refuerza la importancia de estudiar estos efectos en distintos contextos climáticos.

Los hallazgos destacan la relevancia del AED para interpretar fenómenos ambientales y generar hipótesis sobre la evolución de la contaminación. Este enfoque no solo facilita el entendimiento de datos históricos, sino que también sirve como base para comparar tendencias actuales y evaluar el impacto del cambio climático en la calidad del aire.

FUTURAS LÍNEAS DE INVESTIGACIÓN

Comparación de patrones históricos y actuales

- Analizar cómo han cambiado los niveles de ozono y otros contaminantes desde 1973 hasta la actualidad.
- Evaluar el impacto de factores antropogénicos en la calidad del aire con datos recientes.

Influencia del cambio climático en la contaminación

- Investigar cómo el aumento de temperatura global afecta la acumulación de ozono.
- Examinar la variabilidad de la radiación solar y su relación con eventos extremos de contaminación.

Aplicación de técnicas avanzadas de análisis de datos

- Implementar modelos de aprendizaje automático para detectar patrones complejos.
- Explorar nuevas metodologías para mejorar la predicción de niveles de contaminación atmosférica.



Universidad de la Habana
Facultad de Matemáticas y Computación

MUCHAS GRACIAS

Mayo 2025