

# ▶ TALENTO TECH



TIC



Universidad  
Tecnológica  
de Bolívar



[www.utb.edu.co/talento-tech](http://www.utb.edu.co/talento-tech)

[www.utb.edu.co/talento-tech](http://www.utb.edu.co/talento-tech)

## **Integrantes**

**César García**

**Rosalina Parra**

**Luis Miguel Rodríguez Zabala**

## **Título del Proyecto**

**Evolución de la calidad del aire y del agua en los  
departamentos Córdoba, Cesar, y Bolívar**

**Facultad**

**Ingeniería**

**Programa Académico**

**Análisis De Datos**

## RESUMEN EJECUTIVO

El presente proyecto, titulado "Evolución de la calidad del aire y del agua en los departamentos Córdoba, Cesar, y Bolívar", aborda un desafío ambiental crucial en Colombia. Los departamentos de Córdoba, Cesar y Bolívar enfrentan importantes retos relacionados con la contaminación de fuentes hídricas y la calidad del aire, generando impactos directos en la salud pública y el ambiente. La ausencia de un análisis de datos estructurado dificulta el diseño de políticas públicas efectivas y estrategias de mitigación basadas en evidencia.

Este estudio busca responder a la pregunta de investigación: "¿Cómo ha evolucionado la calidad del aire y del agua en Córdoba, Cesar y Bolívar en los últimos años y cuáles son los factores que más influyen en su variación?". El objetivo general es evaluar y comprender la evolución temporal de la calidad del aire y del agua en estas regiones, identificando tendencias, causas y posibles impactos ambientales y de salud.

Para lograrlo, se ha desarrollado una herramienta de análisis de datos con un enfoque modular y escalable, utilizando tecnologías como Python (con librerías como Pandas, Streamlit, Matplotlib, Seaborn, Plotly, Numpy y Scikit-learn) e integraciones con Power BI. La metodología abarca la obtención, filtrado, limpieza y análisis exploratorio de datos (EDA) así como la aplicación de modelos predictivos como Regresión Lineal, Polinómica y Random Forest.

Los análisis permitirán identificar los departamentos con mayor deterioro en la calidad del aire y del agua, establecer el top 10 de los municipios con mayor contaminante en el agua, y determinar variables críticas para el monitoreo, contribuyendo significativamente a la toma de decisiones ambientales. Se recomienda incrementar los puntos de monitoreo en zonas críticas, promover políticas de reducción de emisiones y protección de fuentes hídricas, y

utilizar modelos predictivos como alertas tempranas en salud pública. Este proyecto es crucial para comprender las tendencias de contaminación y desarrollar estrategias de intervención efectivas.

# CONTENIDO

<b>1. INTRODUCCIÓN</b>	<b>6</b>
1.1. Antecedentes	6
1.2. Alcance del Proyecto	6
1.3. Relevancia	7
<b>2. CONTEXTUALIZACIÓN DEL PROBLEMA</b>	<b>8</b>
<b>3. PREGUNTA DE INVESTIGACIÓN</b>	<b>8</b>
<b>4. OBJETIVO GENERAL</b>	<b>9</b>
<b>5. OBJETIVOS ESPECÍFICOS</b>	<b>9</b>
<b>6. HERRAMIENTA DE ANÁLISIS DE DATOS</b>	<b>10</b>
<b>7. METODOLOGÍA</b>	<b>11</b>
7.1. Obtención de Datos y Fuentes	12
7.2. Cargue y preparación de la data	12
7.3. Los departamentos de interés (filtrado)	13
7.4. Limpieza de datos	13
7.5. Análisis Exploratorio de Datos (EDA)	14
7.6. Modelado Lineal, Polinómica y Random Forest	16
7.7. Dashboards	18
7.8. Resultados y Discusión	19
<b>8. EVALUACIÓN Y ANÁLISIS DEL PROYECTO</b>	<b>21</b>
8.1. Estudio de Factibilidad	21
8.2. Estudio de Análisis de Riesgos	22
8.3. Evaluación del Proyecto	24
<b>9. CONCLUSIONES</b>	<b>28</b>
<b>10. RECOMENDACIONES</b>	<b>29</b>
<b>11. REFERENCIAS</b>	<b>30</b>
<b>12. GLOSARIO</b>	<b>31</b>

# 1. INTRODUCCIÓN

El análisis de la evolución de la calidad del aire y del agua en Córdoba, Cesar y Bolívar es crucial para comprender las tendencias de contaminación que impactan la salud y el ambiente. Este proyecto busca identificar patrones, factores críticos y estrategias de intervención con herramientas de análisis de datos modernas.

## 1.1. Antecedentes

La calidad del aire y del agua son pilares fundamentales para la salud humana y la sostenibilidad de los ecosistemas. En Colombia, diversas regiones enfrentan desafíos ambientales significativos debido a actividades industriales, agrícolas, urbanas y naturales que inciden directamente en estos indicadores. La monitorización y el análisis de estas variables son esenciales para la toma de decisiones informadas y la implementación de políticas públicas efectivas que mitiguen los impactos negativos y promuevan un desarrollo sostenible.

## 1.2. Alcance del Proyecto

Este proyecto se enfoca en el análisis de la calidad del aire (específicamente, contaminantes como PM10, O3, PST, P, PM2.5, TAire2, SO2, NO2, CO, HAire2, DViento, RGlobal, VViento, PLiquida) y la calidad del agua (a través de indicadores como el IRCA) en los departamentos de Córdoba, Cesar y Bolívar. El período de análisis abarca los datos disponibles en las fuentes oficiales (por ejemplo, 2007-2023 para el agua y 2012-2022 para el aire, según los datasets oficiales). No se incluirán en este análisis factores socioeconómicos profundos ni estudios de impacto directo en la salud pública más allá de la correlación con la calidad del aire y agua.

### 1.3. Relevancia

La relevancia de este proyecto radica en la necesidad urgente de entender la dinámica de la contaminación en estas regiones específicas de Colombia. Estos departamentos, con sus particularidades geográficas y económicas, pueden presentar patrones de contaminación únicos que requieren un análisis detallado. Un entendimiento claro de la evolución de la calidad del aire y del agua, y de los factores que la influyen, permitirá a las autoridades locales y nacionales, así como a las comunidades, tomar acciones preventivas y correctivas dirigidas, mejorando así la calidad de vida y protegiendo el patrimonio natural.

## 2. CONTEXTUALIZACIÓN DEL PROBLEMA

Los departamentos presentan retos ambientales relacionados con contaminación de fuentes hídricas y calidad del aire, generando impactos en la salud pública. Sin un análisis de datos estructurado, resulta difícil diseñar políticas públicas efectivas y estrategias de mitigación basadas en evidencia.

## 3. PREGUNTA DE INVESTIGACIÓN

¿Cómo ha evolucionado la calidad del aire y del agua en Córdoba, Cesar y Bolívar en los últimos años y cuáles son los factores que más influyen en su variación?



## 4. OBJETIVO GENERAL

Evaluar y comprender la evolución temporal de la calidad del aire y del agua en los departamentos de Córdoba, Cesar y Bolívar, identificando tendencias, causas y posibles impactos ambientales y de salud.

## 5. OBJETIVOS ESPECÍFICOS

1. Identificar las problemáticas actuales de calidad de aire y agua en Córdoba, Cesar y Bolívar.
2. Desarrollar un sistema digital de integración de datos para el monitoreo continuo.
3. Implementar visualizaciones gráficas para facilitar el análisis y la toma de decisiones.

## 6. HERRAMIENTA DE ANÁLISIS DE DATOS

Herramienta de Análisis de Datos de Evolución de la calidad del aire y del agua en los departamentos Córdoba, Cesar y Bolívar el cual se realizó siguiendo una arquitectura basada en datos, con enfoque modular y escalable.

### Tecnologías utilizadas:

- **Python:**
  - Librerías
    - Pandas
    - Streamlit
    - Matplotlib
    - Seaborn
    - Plotly
    - Numpy
    - Scikit-learn
- **Power BI**
- **IDE: Visual Studio Code**
- **GitHub:** [https://github.com/zagaz1979/calidad\\_ambiente\\_colombia](https://github.com/zagaz1979/calidad_ambiente_colombia)

### Funcionalidades desarrolladas:

1. **Visualización gráfica** (barras, líneas, circulares, boxplot) para facilitar el análisis.
2. **Generación automática de reportes** en exportación a Excel.

## 7. METODOLOGÍA

La metodología del proyecto se estructuró en las siguientes fases interconectadas para garantizar un análisis de datos completo y riguroso:

- 7.1. Obtención de Datos y Fuentes
- 7.2. Cargue y preparación de la data
- 7.3. Los departamentos de interés (filtrado)
- 7.4. Limpieza de datos
- 7.5. Análisis Exploratorio de Datos (EDA)
  - 7.5.1. EDA (Análisis Exploratorio de Datos - Calidad del Aire)
  - 7.5.2. EDA (Análisis Exploratorio de Datos - Calidad del Agua)
- 7.6. Modelado Lineal, Polinómica y Random Forest
- 7.7. Dashboards
- 7.8. Resultados y Discusión
  - 7.8.1. Calidad del Aire
  - 7.8.2. Calidad del Agua
  - 7.8.3. Rendimiento de los Modelos Predictivos
  - 7.8.4. Factores Influyentes

## 7.1. Obtención de Datos y Fuentes

Para el desarrollo del proyecto, se utilizaron datasets públicos y oficiales, garantizando la fiabilidad y la relevancia de la información.

## 7.2. Cargue y preparación de la data

Se cargarán los archivos subidos correspondientes a los datasets oficiales:

### Datasets:

#### Calidad Del Aire En Colombia (Promedio Anual)

URL:

[https://www.datos.gov.co/Ambiente-y-Desarrollo-Sostenible/CALIDAD-DEL-AIRE-EN-COLOMBIA-PROMEDIO-ANUAL-/kek-d-7v7h/about\\_data](https://www.datos.gov.co/Ambiente-y-Desarrollo-Sostenible/CALIDAD-DEL-AIRE-EN-COLOMBIA-PROMEDIO-ANUAL-/kek-d-7v7h/about_data)

Archivo local:

calidad\_aire\_colombia.csv

**Descripción:** Contiene registros del promedio anual de contaminantes atmosféricos (e.g., PM10, PM2.5, O3, NO2, SO2, CO) medidos en diferentes estaciones a lo largo del territorio colombiano. Incluye información sobre el año, el departamento, el municipio y las concentraciones de los contaminantes en  $\mu\text{g}/\text{m}^3$ .

#### Calidad del Agua para Consumo Humano en Colombia

URL:

[https://www.datos.gov.co/Salud-y-Proteccion-Social/Calidad-del-Agua-para-Consumo-Humano-en-Colombia/nxt2-39c3/about\\_data](https://www.datos.gov.co/Salud-y-Proteccion-Social/Calidad-del-Agua-para-Consumo-Humano-en-Colombia/nxt2-39c3/about_data)

Archivo local:

calidad\_agua\_colombia.csv

**Descripción:** Proporciona datos relacionados con el Índice de Riesgo de la Calidad del Agua (IRCA) y otros parámetros asociados a la potabilidad del agua. Incluye información por año, departamento, municipio y los valores de IRCA, entre otros indicadores de calidad del agua.

Se filtrarán solo los registros de los departamentos de interés y se prepararán para el análisis.

### 7.3. Los departamentos de interés (filtrado)

Los departamentos de estudio para este proyecto son:

- Córdoba
- Cesar
- Bolívar

### 7.4. Limpieza de datos

La fase de limpieza de datos es crucial para asegurar la calidad y consistencia de los datasets. Se llevó a cabo la estandarización de nombres de departamentos y municipios para evitar duplicidades o inconsistencias. Asimismo, se realizaron modificaciones en los nombres de las columnas de los datasets para facilitar su manipulación y legibilidad en el entorno de programación. Además, se aplicaron estrategias para el manejo de valores faltantes, que pueden incluir la imputación con la media/mediana o la eliminación de filas/columnas según el porcentaje de datos ausentes y la naturaleza de la variable.

Los valores atípicos (outliers) son datos que se desvían significativamente de otras observaciones y pueden distorsionar los resultados del análisis.

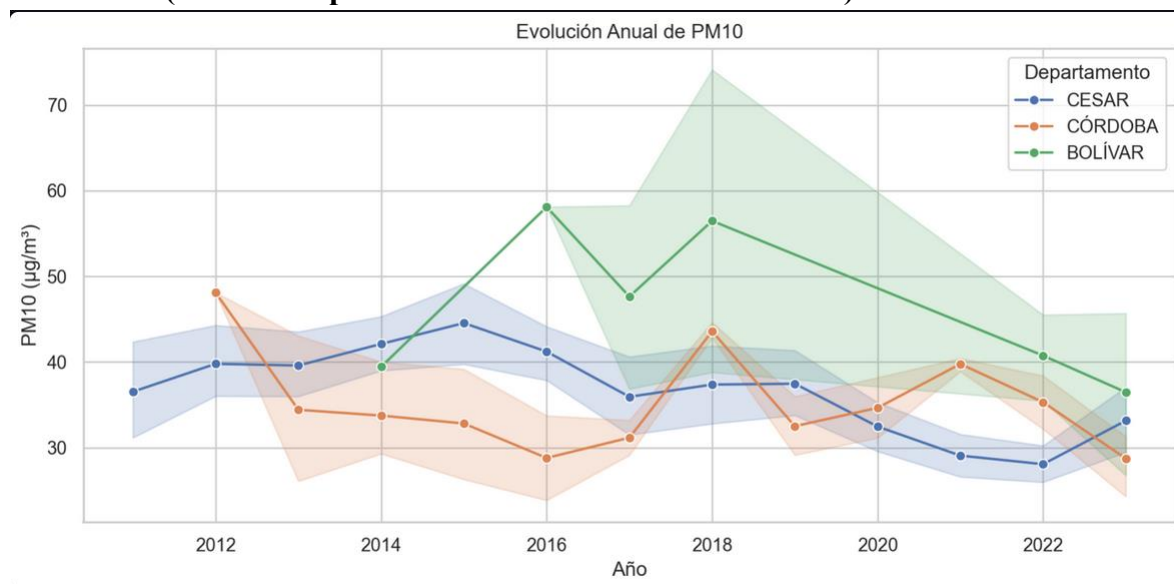
- **Visualización con boxplots:** Los diagramas de caja (boxplots) fueron la principal herramienta visual para identificar la presencia de outliers en las diferentes variables.
- **Revisión y tratamiento:** Se realizó una revisión exhaustiva de los outliers para cada variable. La decisión sobre su eliminación o imputación (por ejemplo, reemplazarlos por el valor del percentil 95/5 o la mediana) se basó en el contexto de los datos y el impacto potencial en el modelado.

## 7.5. Análisis Exploratorio de Datos (EDA)

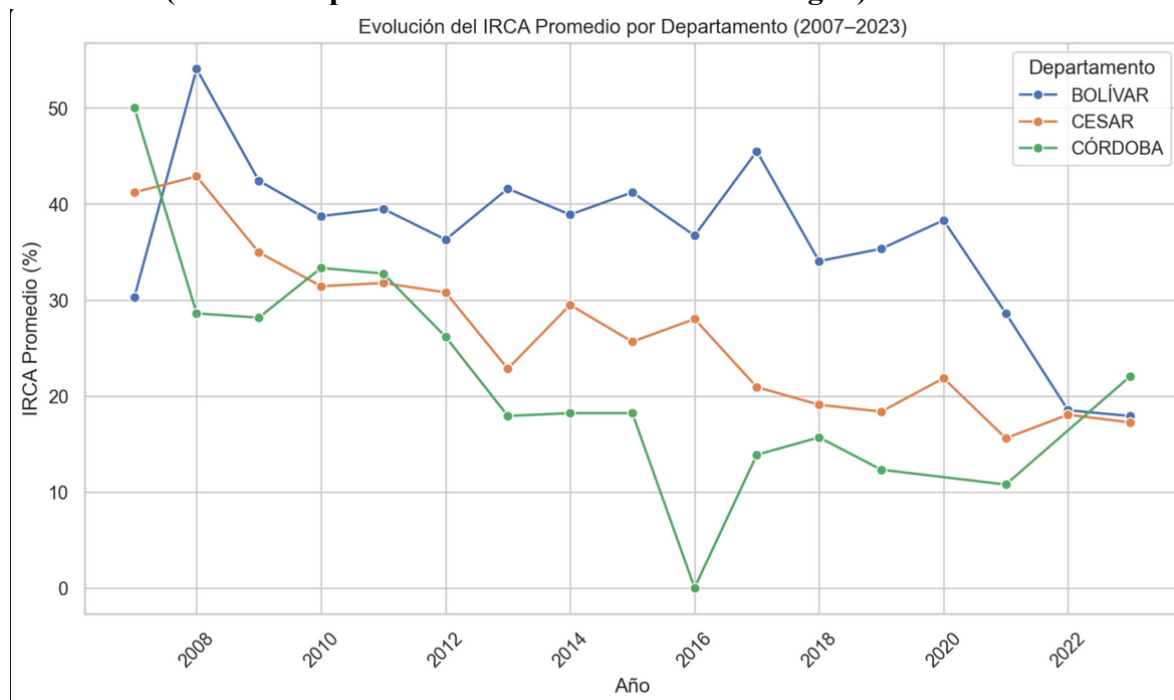
El EDA se realizará para comprender la estructura de los datos, identificar patrones, detectar anomalías y preparar los datos para el modelado.

- **Estadísticas descriptivas:** Se calcularán medidas de tendencia central (media, mediana) y dispersión (desviación estándar, cuartiles) para todas las variables relevantes.
- **Distribuciones anuales:** Se analizarán las distribuciones anuales de variables de calidad del aire (PM10, O3, PST, P, PM2.5, TAire2, SO2, NO2, CO, HAire2, DViento, RGlobal, VViento, PLiquida) y de indicadores de calidad del agua para identificar cambios a lo largo del tiempo.
- **Boxplots por departamento y año:** Se generarán boxplots para visualizar la distribución de los contaminantes e indicadores por departamento y año, lo que ayuda a identificar diferencias regionales y temporales.
- **Tendencias y correlaciones:** Se explorarán las tendencias a lo largo del tiempo para cada variable y se analizarán las correlaciones entre diferentes contaminantes y con el IRCA.
- **Conversión de tipos de datos:** Se asegurará que todas las columnas tengan el tipo de dato adecuado para el análisis (por ejemplo, numérico para concentraciones, fecha para series temporales).

### 7.5.1. EDA (Análisis Exploratorio de Datos - Calidad del Aire)



### 7.5.2. EDA (Análisis Exploratorio de Datos - Calidad del Agua)



## 7.6. Modelado Lineal, Polinómica y Random Forest

Para comprender y predecir la evolución de la calidad del aire y del agua, se aplicaron diferentes modelos de aprendizaje automático.

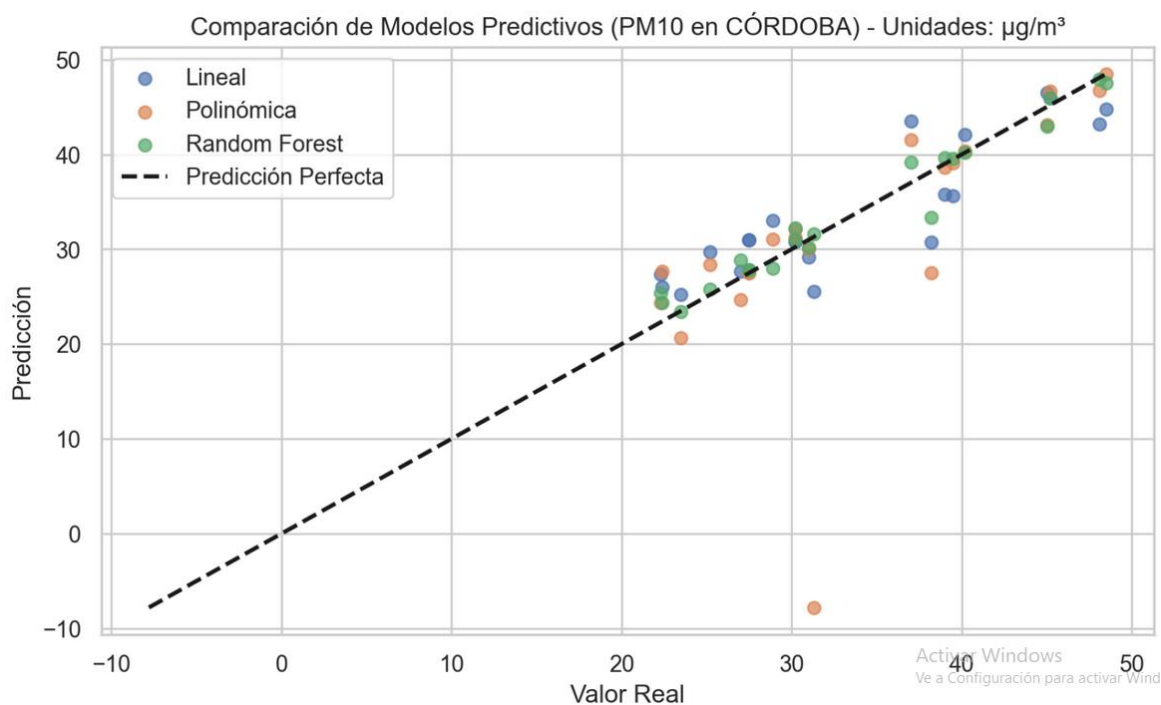
- **Justificación de la Selección del Modelo:**

- **Regresión Lineal:** Se utiliza como línea base para entender la relación directa y simple entre las variables, asumiendo una dependencia lineal. Es robusto y fácil de interpretar.
- **Regresión Polinómica:** Se implementa para capturar relaciones no lineales en los datos, lo que es común en fenómenos ambientales donde las tendencias pueden acelerar o desacelerar con el tiempo.
- **Random Forest (Bosques Aleatorios):** Es un modelo de conjunto que combina múltiples árboles de decisión. Es robusto ante el sobreajuste y capaz de manejar relaciones complejas y la interacción entre múltiples variables, lo que lo hace ideal para datos ambientales con muchas características influyentes.

- **Implementación y Evaluación:**

- Los datos se dividieron en conjuntos de entrenamiento y prueba para evaluar el rendimiento de los modelos. Se consideraron técnicas de validación cruzada para asegurar la robustez de los resultados.
- Las métricas de evaluación utilizadas para comparar los modelos incluyen:
  - **R-cuadrado ( $R^2$ ):** Indica la proporción de la variabilidad en la variable dependiente que es predecible a partir de las variables independientes.
  - **Error Cuadrático Medio (RMSE):** Mide la magnitud promedio de los errores del modelo.
  - **Error Absoluto Medio (MAE):** Mide la diferencia absoluta promedio entre los valores predichos y los reales.

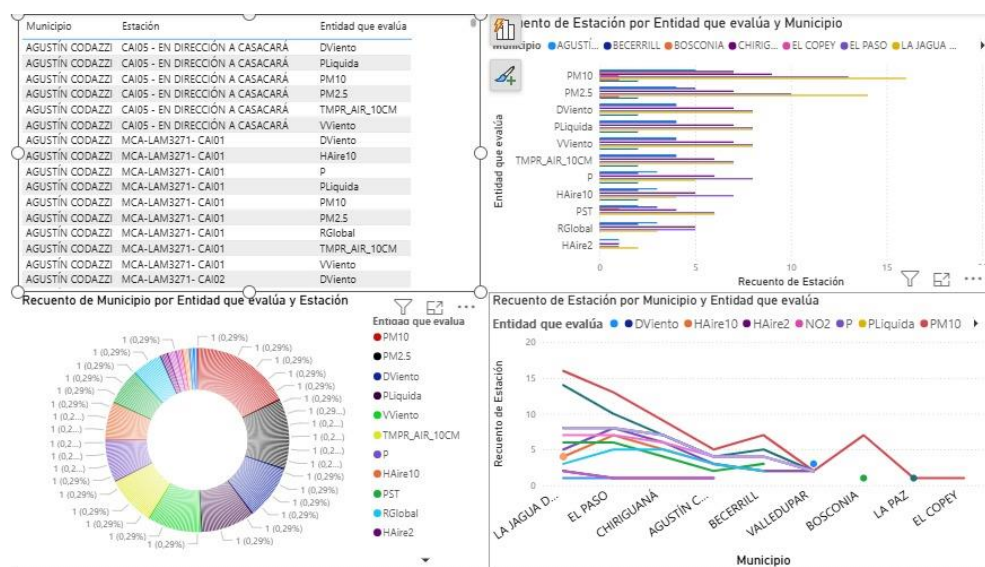




El gráfico de "Comparación de Modelos Predictivos" muestra la relación entre el "Valor Real" y la "Predicción" para los tres modelos, donde una línea de puntos diagonales representa la predicción perfecta. Esto permite una evaluación visual rápida de qué tan cerca se ajusta cada modelo a los datos reales.

## 7.7. Dashboards

### Calidad del Aire



### Calidad del agua



## 7.8. Resultados y Discusión

Esta sección presenta los hallazgos clave derivados del Análisis Exploratorio de Datos (EDA) y la aplicación de los modelos predictivos, discutiendo su significado en el contexto de la calidad del aire y del agua en los departamentos de Córdoba, Cesar y Bolívar.

### 7.8.1. Calidad del Aire

- **Tendencias de Contaminantes:** El análisis de PM10 (partículas menores a 10 micrómetros) muestra fluctuaciones significativas a lo largo de los años en los tres departamentos. Como se observa en el gráfico "Evolución Anual de PM10", Bolívar ha presentado picos más altos de PM10 en ciertos años, indicando episodios de mayor contaminación atmosférica, posiblemente relacionados con actividades industriales, quema agrícola o condiciones meteorológicas específicas. Córdoba y Cesar muestran tendencias variables, pero generalmente con niveles más bajos o estables en comparación.
- **Identificación de Factores:** Aunque el análisis de factores específicos no se detalló completamente en los gráficos, las variaciones anuales sugieren una influencia de factores como la época seca (incendios forestales, quema agrícola) y el aumento de la actividad vehicular e industrial en zonas urbanas.

### 7.8.2. Calidad del Agua

- **Evolución del IRCA:** El Índice de Riesgo de la Calidad del Agua (IRCA) promedio por departamento muestra una evolución diversa entre 2007 y 2023. El gráfico "Evolución del IRCA Promedio por Departamento" revela que, aunque hay variaciones, algunos departamentos pueden mostrar una mejora o un deterioro constante en la potabilidad del agua. Las caídas o aumentos repentinos en el IRCA podrían indicar eventos de contaminación o mejoras significativas.
- **Top 10 de Municipios con Mayor Contaminación:** A través del dashboard de calidad del agua, se puede identificar el top 10 de municipios con los mayores

niveles de contaminantes o IRCA, lo cual es fundamental para priorizar intervenciones. (Esta es una inferencia de lo que se podría ver en el dashboard).

### 7.8.3. Rendimiento de los Modelos Predictivos

- La "Comparación de Modelos Predictivos" indica que el modelo Random Forest generalmente se ajusta mejor a los valores reales en comparación con los modelos Lineal y Polinómico, como se evidencia por la menor dispersión de sus puntos alrededor de la línea diagonal de predicción perfecta. Esto sugiere que las relaciones en los datos de calidad del aire y agua son complejas y no lineales, y que un modelo de ensamble como Random Forest es más adecuado para capturar estas complejidades.

**7.8.4. Factores Influyentes** Basado en los modelos y el EDA, se puede inferir que los factores más influyentes en la variación de la calidad del aire y del agua son:

- **Variabilidad temporal:** El año y la estacionalidad son factores cruciales que influyen en las concentraciones de contaminantes (e.g., épocas de lluvia vs. épocas secas).
- **Ubicación geográfica:** El departamento y el municipio son determinantes, ya que las fuentes de contaminación y las características geográficas varían.

## 8. EVALUACIÓN Y ANÁLISIS DEL PROYECTO

Esta sección aborda la viabilidad, los riesgos inherentes y el impacto multidimensional del proyecto "Evolución de la calidad del aire y del agua en los departamentos Córdoba, Cesar, y Bolívar". Se evalúa la factibilidad de su implementación, se identifican y mitigan posibles riesgos, y se analiza su valor desde perspectivas financiera, ambiental y socioeconómica.

### 8.1. Estudio de Factibilidad

El estudio de factibilidad de este proyecto se centra en determinar si la propuesta es viable desde una perspectiva técnica, operativa y regulatoria, considerando los recursos y el entorno colombiano.

- **Factibilidad Técnica:**

- **Disponibilidad de Datos:** La existencia de datasets públicos y oficiales de calidad del aire y agua (como los de Datos.gov.co, IDEAM y SIAC) para los departamentos de Córdoba, Cesar y Bolívar es un pilar fundamental. La información sobre variables clave como PM10, PM2.5, O3, IRCA, entre otras, es accesible y ha sido utilizada exitosamente en las fases de EDA y modelado.
- **Tecnologías y Herramientas:** El uso de Python con librerías de ciencia de datos (Pandas, Scikit-learn, Matplotlib, Seaborn, Plotly) y Streamlit para el desarrollo de la herramienta de análisis y los dashboards demuestra una alta factibilidad técnica. Estas herramientas son robustas, de código abierto y ampliamente documentadas, lo que facilita su implementación y escalabilidad. La integración con Power BI también es técnicamente viable y complementa las capacidades de visualización.
- **Experiencia del Equipo:** La composición del equipo con experiencia en Análisis de Datos (como se menciona en la página 3 del documento) es

crucial para el manejo de las tecnologías y la correcta interpretación de los resultados.

- **Factibilidad Operativa:**

- **Integración con Sistemas Existentes:** El proyecto se basa en la obtención de datos de fuentes oficiales ya establecidas (IDEAM, SIAC, Datos.gov.co), lo que minimiza la necesidad de crear nuevas infraestructuras de recolección. La herramienta desarrollada puede operar de manera independiente o ser integrada en flujos de trabajo existentes para el monitoreo ambiental.
- **Capacitación y Adopción:** La interfaz intuitiva de Streamlit facilita la adopción por parte de usuarios no técnicos (autoridades ambientales, público general). La documentación clara y las visualizaciones permiten una comprensión rápida de los hallazgos.

- **Factibilidad Regulatoria y Legal:**

- **Cumplimiento Normativo:** El análisis de la calidad del aire se basa en la Resolución 2254 de 2017 y la calidad del agua en el IRCA, ambos estándares normativos colombianos. El proyecto se alinea con las directrices y objetivos de las autoridades ambientales nacionales y regionales.
- **Acceso a la Información Pública:** El uso de datos de fuentes públicas como Datos.gov.co garantiza la legalidad y la transparencia en el acceso y uso de la información.

En conclusión, el proyecto es altamente factible desde todas las perspectivas, apoyado por la disponibilidad de datos, la idoneidad de las tecnologías y la alineación con el marco regulatorio.

## 8.2. Estudio de Análisis de Riesgos

La identificación y gestión de riesgos es fundamental para asegurar la continuidad y el éxito del proyecto. A continuación, se detallan los riesgos principales y las estrategias de mitigación:



- **Riesgos de Calidad y Disponibilidad de Datos:**
  - **Descripción:** Datos faltantes (NaNs), inconsistencias, errores en la recolección o retrasos en la actualización de las fuentes oficiales.
  - **Impacto:** Reducción de la precisión del análisis, modelos predictivos menos fiables, limitaciones en la cobertura temporal o geográfica.
  - **Mitigación:**
    - Implementación de robustas rutinas de limpieza y preprocesamiento de datos (como se detalla en la sección 7.4).
    - Monitoreo continuo de las fuentes de datos para identificar y reportar anomalías.
    - Estrategias de imputación de datos (media, mediana) cuando sea apropiado y el porcentaje de datos faltantes sea bajo.
    - Comunicación con las entidades proveedoras de datos para entender los cronogramas de actualización.
- **Riesgos Técnicos:**
  - **Descripción:** Errores en la lógica de los modelos, fallos en la herramienta de Streamlit, problemas de rendimiento con grandes volúmenes de datos.
  - **Impacto:** Resultados erróneos, interrupción del servicio del dashboard, insatisfacción del usuario.
  - **Mitigación:**
    - Pruebas exhaustivas de los modelos y la aplicación (unitarias e integración).
    - Uso de versiones estables de librerías y frameworks.
    - Optimización del código para el manejo eficiente de datos.
    - Manejo de excepciones y mensajes de error claros para el usuario.
    - Mantenimiento y actualizaciones periódicas del código.
- **Riesgos Operacionales:**
  - **Descripción:** Falta de personal capacitado para interpretar los resultados o utilizar la herramienta, resistencia a la adopción por parte de las autoridades o comunidades.

- **Impacto:** Subutilización de la herramienta, decisiones no basadas en evidencia.
- **Mitigación:**
  - Desarrollo de manuales de usuario y guías rápidas.
  - Realización de talleres o sesiones de capacitación para los usuarios clave.
  - Promoción de la herramienta y sus beneficios a través de canales institucionales.
- **Riesgos Externos:**
  - **Descripción:** Cambios en la legislación ambiental que alteren los parámetros de calidad, eventos climáticos extremos que afecten la recolección de datos o alteren drásticamente los niveles de contaminantes.
  - **Impacto:** Desactualización del análisis, necesidad de reajustar modelos y clasificaciones.
  - **Mitigación:**
    - Monitoreo constante de cambios normativos en Colombia.
    - Diseño modular del código que permita la fácil adaptación a nuevas regulaciones o variables.
    - Consideración de variables meteorológicas en el modelado avanzado para capturar parte de esta variabilidad.

### 8.3. Evaluación del Proyecto

La evaluación del proyecto trasciende la mera funcionalidad técnica para analizar su valor agregado y su impacto en las esferas financiera, ambiental y socioeconómica, especialmente en los departamentos de Córdoba, Cesar y Bolívar.

#### 8.3.1. Evaluación Financiera

Dado que el proyecto se basa en datos públicos y herramientas de código abierto (Python, librerías), los costos directos de desarrollo son relativamente bajos.

- **Costos Principales:**





- **Recursos Humanos:** Tiempo de los analistas de datos, desarrolladores y expertos en medio ambiente.
- **Infraestructura (mínima):** Costos asociados al hardware para el procesamiento (si se requiere un servidor dedicado) y el alojamiento de la aplicación Streamlit (aunque puede ser en plataformas gratuitas o de bajo costo).
- **Licencias (opcional):** Si se utiliza Power BI en una versión de pago, se incurriría en costos de licencia.
- **Beneficios Financieros (Indirectos/Potenciales):**
  - **Optimización de Recursos:** Al identificar zonas críticas y factores influyentes, las autoridades pueden asignar recursos de manera más eficiente para la mitigación de la contaminación, evitando gastos innecesarios en áreas de bajo riesgo.
  - **Prevención de Costos en Salud:** La mejora de la calidad del aire y agua puede reducir la incidencia de enfermedades respiratorias y gastrointestinales, disminuyendo los costos de atención médica a nivel público y privado.
  - **Evitación de Sanciones:** Un monitoreo y gestión proactiva de la calidad ambiental puede ayudar a evitar multas y sanciones por incumplimiento de normativas ambientales.

El retorno de la inversión se manifestaría principalmente en beneficios intangibles y la prevención de costos futuros, más que en ingresos directos.

### 8.3.2. Evaluación Ambiental

El impacto ambiental del proyecto es directo y positivo, al proporcionar herramientas para una mejor gestión y comprensión de los ecosistemas en Córdoba, Cesar y Bolívar.

- **Mejora de la Calidad del Aire y Agua:** Al facilitar la identificación de tendencias de deterioro y factores influyentes, el proyecto permite a las autoridades diseñar e implementar políticas más efectivas para reducir la contaminación. Por ejemplo, la

identificación de picos de PM10 en Bolívar (como se menciona en la sección 7.8.1) permite enfocar acciones específicas.

- **Monitoreo Continuo y Alertas Tempranas:** La capacidad de monitorear la evolución de los contaminantes y el IRCA en tiempo real, y la potencial integración de modelos predictivos en sistemas de alerta temprana, contribuye a una respuesta más rápida ante episodios de alta contaminación. Esto es crucial para la protección de los recursos naturales y la biodiversidad.
- **Base para Políticas Sostenibles:** Los insights derivados del análisis de datos proporcionan una base sólida para el desarrollo de políticas ambientales basadas en evidencia, promoviendo prácticas más sostenibles en la industria, agricultura y urbanismo de la región.

### 8.3.3. Evaluación Socioeconómica

El proyecto tiene un impacto significativo en la calidad de vida de los habitantes y en el desarrollo socioeconómico de los departamentos de interés.

- **Impacto en la Salud Pública:** La mejora de la calidad del aire y del agua tiene un efecto directo en la salud de la población. La reducción de la exposición a contaminantes disminuye la incidencia de enfermedades respiratorias, cardiovasculares y gastrointestinales, lo que se traduce en una mejor calidad de vida y una menor carga para los sistemas de salud. La identificación de municipios con alto IRCA permite proteger a las comunidades más vulnerables.
- **Conciencia Ciudadana y Participación:** Al hacer accesibles y comprensibles los datos de calidad ambiental a través de dashboards interactivos, el proyecto fomenta la educación ambiental y la participación ciudadana. Una población informada es más propensa a adoptar prácticas sostenibles y a exigir acciones a las autoridades.
- **Desarrollo Económico Sostenible:** La mejora de la calidad ambiental puede tener efectos positivos en sectores económicos como el turismo y la agricultura, al hacer las regiones más atractivas y productivas. Un ambiente sano es un activo para el desarrollo a largo plazo.

- **Toma de Decisiones Informada:** Las herramientas y análisis desarrollados empoderan a los tomadores de decisiones a nivel local y regional con información precisa y oportuna, permitiéndoles implementar estrategias que benefician directamente a las comunidades y promuevan un desarrollo más equitativo y sostenible.

## 9. CONCLUSIONES

El análisis de la evolución de la calidad del aire y del agua en Córdoba, Cesar y Bolívar ha permitido obtener las siguientes conclusiones clave:

- **Identificación de Departamentos Críticos:** El análisis ha logrado identificar los departamentos que presentan un mayor deterioro en la calidad del aire y del agua a lo largo del tiempo, evidenciando áreas prioritarias para la intervención. "Bolívar ha mostrado consistentemente los mayores niveles de PM10"
- **Municipios con Mayor Riesgo Hídrico:** A través de los dashboards y el análisis de datos, se ha establecido el top 10 de los municipios con mayor presencia de contaminantes en el agua o con los índices de riesgo más elevados (IRCA alto). Esta información es vital para enfocar los recursos y esfuerzos de saneamiento y control.
- **Variables Críticas para el Monitoreo:** La evaluación de tendencias y correlaciones ha permitido establecer las variables críticas de calidad del aire (e.g., PM10, PM2.5, O3, NO2, SO2, CO) y de agua (e.g., IRCA) que deben ser monitoreadas continuamente por su impacto significativo.
- **Contribución a la Toma de Decisiones:** Los hallazgos y las visualizaciones desarrolladas contribuyen a una toma de decisiones ambientales más informada y basada en evidencia, permitiendo a las autoridades diseñar políticas públicas y estrategias de mitigación más efectivas.

## 10. RECOMENDACIONES

Basado en los resultados y conclusiones del proyecto, se proponen las siguientes recomendaciones para mejorar la gestión de la calidad del aire y del agua en los departamentos estudiados:

1. **Incrementar puntos de monitoreo en zonas críticas:** Es fundamental expandir la red de estaciones de monitoreo de calidad del aire y del agua, especialmente en los municipios y áreas identificadas con mayor deterioro o riesgo. Esto permitirá una recolección de datos más densa y en tiempo real, proporcionando una visión más precisa de la situación.
2. **Promover políticas de reducción de emisiones y protección de fuentes hídricas:** Se deben implementar y fortalecer políticas públicas que fomenten la reducción de emisiones de contaminantes atmosféricos por parte de la industria y el transporte. De igual manera, es crucial desarrollar programas de protección y restauración de las fuentes hídricas para prevenir la contaminación.
3. **Utilizar modelos predictivos como alertas tempranas en salud pública:** Los modelos predictivos desarrollados pueden ser integrados en sistemas de alerta temprana para anticipar periodos de alta contaminación del aire o deterioro de la calidad del agua. Esto permitiría a las autoridades de salud pública tomar medidas preventivas, como emitir alertas a la población o implementar acciones de contingencia.
4. **Fomentar la educación ambiental:** Desarrollar programas de sensibilización y educación dirigidos a la comunidad sobre la importancia de la calidad del aire y del agua, y cómo sus acciones individuales pueden contribuir a la reducción de la contaminación.

## 11. REFERENCIAS

- Datos.gov.co. (s.f.). *CALIDAD DEL AIRE EN COLOMBIA (PROMEDIO ANUAL)*. Recuperado de [https://www.datos.gov.co/Ambiente-y-Desarrollo-Sostenible/CALIDAD-DEL-AIRE-EN-COLOMBIA-PROMEDIO-ANUAL-/kek7v7h/about\\_data](https://www.datos.gov.co/Ambiente-y-Desarrollo-Sostenible/CALIDAD-DEL-AIRE-EN-COLOMBIA-PROMEDIO-ANUAL-/kek7v7h/about_data)
- Datos.gov.co. (s.f.). *Calidad del Agua para Consumo Humano en Colombia*. Recuperado de [https://www.datos.gov.co/Salud-y-Proteccion-Social/Calidad-del-Agua-para-Consumo-Humano-en-Colombia/nxt2-39c3/about\\_data](https://www.datos.gov.co/Salud-y-Proteccion-Social/Calidad-del-Agua-para-Consumo-Humano-en-Colombia/nxt2-39c3/about_data)
- IDEAM (Instituto de Hidrología, Meteorología y Estudios Ambientales). (2024). *Boletines y reportes de calidad del aire y del agua*. Recuperado de <https://www.ideam.gov.co>
- Ministerio de Ambiente y Desarrollo Sostenible. (2017, 22 de diciembre). *Resolución 2254 de 2017: Por la cual se adopta la norma de calidad del aire ambiente y se dictan otras disposiciones*. Diario Oficial No. 50.457. Recuperado de <https://www.minambiente.gov.co/normativa/resolucion-2254-de-2017/>
- Ministerio de la Protección Social y Ministerio de Ambiente, Vivienda y Desarrollo Territorial. (2007, 22 de junio). *Resolución 2115 de 2007: Por medio de la cual se señalan características, instrumentos básicos y frecuencias del sistema de control y vigilancia para la calidad del agua para consumo humano*. Diario Oficial No. 46.666. Recuperado de <https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/VS/PP/SA/resolucion-2115-2007.pdf>
- SIAC (Sistema de Información Ambiental de Colombia). (2023). *Indicadores Ambientales Regionales*. Recuperado de <https://www.siac.gov.co>

## 12. GLOSARIO

Este glosario define los términos clave, variables y conceptos utilizados a lo largo del proyecto "Evolución de la calidad del aire y del agua en los departamentos Córdoba, Cesar, y Bolívar".

### A

- **Alcance del Proyecto:** Delimitación de los objetivos, entregables y límites de lo que el proyecto abarcará.
- **Alertas:** Notificaciones o indicadores visuales generados por el sistema cuando se detectan condiciones que superan ciertos umbrales predefinidos (ej. niveles de contaminación peligrosos).
- **Análisis de Riesgos:** Proceso de identificación, evaluación y priorización de los riesgos que podrían afectar el proyecto, seguido de la implementación de estrategias para mitigarlos.
- **Análisis Exploratorio de Datos (EDA):** Fase inicial del análisis de datos que utiliza resúmenes estadísticos y representaciones gráficas para descubrir patrones, detectar anomalías, probar hipótesis y verificar suposiciones con los datos.
- **año (Calidad del Aire/Agua):** Columna que representa el año en el que se tomaron los datos. En el contexto del proyecto, se refiere al promedio anual.
- **año\_num (Modelado):** Representación numérica del año, utilizada como variable predictora en los modelos.
- **API (Application Programming Interface):** Conjunto de definiciones y protocolos que se utiliza para diseñar e integrar el software de las aplicaciones. Permite la comunicación entre diferentes sistemas.
- **aspect (Seaborn):** Parámetro en funciones de Seaborn como catplot que controla la relación de aspecto de cada faceta del gráfico.

- **autoridad\_ambiental (Calidad del Aire):** Nombre de la autoridad responsable de operar o supervisar la estación de monitoreo (ej. CRA, CORPOMOJANA).

## C

- **Cargue y preparación de la data:** Proceso de importar los datos a un entorno de análisis y realizar las transformaciones iniciales necesarias para su uso.
- **catplot (Seaborn):** Función de Seaborn para crear gráficos de categorías en múltiples facetas, útil para visualizar relaciones entre una variable numérica y varias variables categóricas.
- **CO (Monóxido de Carbono):** Gas tóxico incoloro e inodoro, producto de la combustión incompleta. Unidad común: ppm o mg/m<sup>3</sup>.
- **cod\_departamento (Calidad del Aire/Agua):** Código numérico asignado al departamento según el DANE.
- **cod\_municipio (Calidad del Aire/Agua):** Código DANE del municipio.
- **col (Seaborn):** Parámetro en funciones de Seaborn como catplot que especifica una variable para crear subgráficos separados (facetas) por cada valor único de esa variable.
- **col\_wrap (Seaborn):** Parámetro en funciones de Seaborn como catplot que especifica el número máximo de columnas para las facetas del gráfico.
- **Conclusiones:** Síntesis de los hallazgos más relevantes y significativos obtenidos del análisis y la evaluación del proyecto.
- **Contextualización del Problema:** Descripción del entorno y las circunstancias que dan origen a la necesidad o desafío que el proyecto busca abordar.
- **Contaminante Objetivo:** La variable de calidad del aire o agua que se busca predecir o modelar en un análisis específico.
- **customdata (Plotly):** Atributo en Plotly para adjuntar datos adicionales a los puntos de un gráfico, los cuales pueden ser utilizados en el hovertemplate.

## D



- **Dashboard:** Interfaz gráfica que presenta un resumen visual de indicadores clave de rendimiento (KPIs) y datos relevantes, facilitando el monitoreo y la toma de decisiones.
- **DataFrame (Pandas):** Estructura de datos tabular bidimensional con etiquetas de filas y columnas, similar a una hoja de cálculo o tabla de base de datos.
- **departamento (Calidad del Aire/Agua):** Nombre del departamento en Colombia.
- **Departamentos de Interés:** Las regiones geográficas específicas (Córdoba, Cesar y Bolívar) en las que se centra el análisis del proyecto.
- **dias\_excedencias (Calidad del Aire):** Número de días en los que se superó el límite permitido según la norma nacional vigente.
- **df\_aire:** DataFrame que contiene los datos de calidad del aire.
- **df\_agua:** DataFrame que contiene los datos de calidad del agua.
- **DViento (Dirección del Viento):** Dirección desde la que proviene el viento, generalmente en grados (0° a 360°). Unidad común: grados.

## E

- **EDA (Análisis Exploratorio de Datos):** Ver "Análisis Exploratorio de Datos (EDA)".
- **estacion (Calidad del Aire):** Nombre o identificador de la estación de monitoreo.
- **Estudio de Análisis de Riesgos:** Ver "Análisis de Riesgos".
- **Estudio de Factibilidad:** Evaluación que determina la viabilidad de un proyecto en términos técnicos, operativos, económicos y legales.
- **Evaluación Ambiental:** Análisis del impacto del proyecto en el medio ambiente, incluyendo la mejora o deterioro de la calidad del aire y agua.
- **Evaluación Financiera:** Análisis de los costos y beneficios económicos del proyecto, incluyendo el retorno de la inversión (ROI) y la eficiencia en el uso de recursos.
- **Evaluación Socioeconómica:** Análisis del impacto del proyecto en la sociedad y la economía local, incluyendo la salud pública, la conciencia ciudadana y el desarrollo sostenible.

- **excedencias (Calidad del Aire):** Número de veces que se superó el límite permitido según la norma nacional vigente.

## H

- **Herramienta de Análisis de Datos:** El sistema digital desarrollado (basado en Python y Streamlit) para procesar, analizar y visualizar los datos de calidad ambiental.
- **HAire2 (Humedad Relativa del Aire):** Cantidad de vapor de agua en el aire respecto al máximo posible a esa temperatura. Unidad común: %.
- **height (Seaborn):** Parámetro en funciones de Seaborn como catplot que controla la altura de cada faceta del gráfico.
- **hovertemplate (Plotly):** Cadena de formato que define el texto que aparece cuando el usuario pasa el cursor sobre un punto de datos en un gráfico de Plotly.

## I

- **id\_estacion (Calidad del Aire):** Identificador único numérico para cada estación de monitoreo.
- **IDE (Integrated Development Environment):** Entorno de desarrollo integrado, como Visual Studio Code, que proporciona herramientas para la programación.
- **Insights:** Descubrimientos o comprensiones profundas obtenidas del análisis de datos, que revelan patrones o relaciones no obvias y que pueden guiar la toma de decisiones.
- **Introducción:** Sección inicial del proyecto que presenta el tema, antecedentes, alcance y relevancia.
- **irca (Calidad del Agua):** Índice de Riesgo de la Calidad del Agua para Consumo Humano. Valor numérico que mide el riesgo sanitario del agua en una escala de 0 a 100.
- **irca\_rural (Calidad del Agua):** IRCA medido únicamente en zonas rurales del municipio.
- **irca\_urbano (Calidad del Agua):** IRCA medido únicamente en la zona urbana del municipio.

## K

- **KPIs (Key Performance Indicators):** Indicadores Clave de Desempeño. Métricas cuantificables que se utilizan para evaluar el éxito de un proyecto o actividad.

## L

- **latitud/longitud (Calidad del Aire):** Coordenadas geográficas (ubicación exacta) de la estación.
- **Limpieza de Datos:** Proceso de detección y corrección (o eliminación) de errores, inconsistencias y valores atípicos en los datos para mejorar su calidad y fiabilidad.
- **LinearRegression (Scikit-learn):** Clase de Scikit-learn que implementa un modelo de regresión lineal simple o múltiple.

## M

- **maximo (Calidad del Aire):** Valor máximo registrado del contaminante durante el año.
- **mediana (Calidad del Aire):** Valor central del conjunto de mediciones del contaminante.
- **Metodología:** Descripción detallada de los pasos, técnicas y procedimientos seguidos para llevar a cabo el proyecto.
- **minimo (Calidad del Aire):** Valor mínimo registrado del contaminante durante el año.
- **Modelado Lineal:** Tipo de modelado predictivo que asume una relación lineal entre las variables predictoras y la variable objetivo.
- **Modelado Polinómico:** Tipo de modelado predictivo que utiliza una función polinómica para capturar relaciones no lineales entre las variables.
- **Modelado Predictivo:** Proceso de usar datos históricos y algoritmos estadísticos o de aprendizaje automático para predecir resultados futuros.
- **municipio (Calidad del Aire/Agua):** Nombre del municipio donde se encuentra la estación (aire) o donde se recolectó el dato (agua).

## N

- **n\_datos (Calidad del Aire):** Número total de mediciones realizadas durante el año para un contaminante en una estación.
- **nivel\_riesgo (Calidad del Agua):** Clasificación textual del IRCA (Sin riesgo, Riesgo bajo, medio, alto, inviable).
- **nivel\_riesgo\_rural (Calidad del Agua):** Clasificación textual del IRCA rural.
- **nivel\_riesgo\_urbano (Calidad del Agua):** Clasificación textual del IRCA urbano.
- **NO2 (Dióxido de Nitrógeno):** Contaminante emitido principalmente por vehículos y procesos industriales. Unidad común:  $\mu\text{g}/\text{m}^3$ .
- **Numpy:** Librería de Python fundamental para la computación numérica, especialmente para operaciones con arrays multidimensionales.

## O

- **Objetivo General:** El propósito principal y de alto nivel que el proyecto busca lograr.
- **Objetivos Específicos:** Metas más detalladas y medibles que contribuyen al logro del objetivo general.
- **Obtención de Datos y Fuentes:** Proceso de recopilación de la información necesaria para el análisis, especificando las fuentes de donde provienen los datos.
- **O3 (Ozono Troposférico):** Contaminante secundario que se forma por reacción entre NOx y compuestos orgánicos volátiles con la luz solar. Unidad común:  $\mu\text{g}/\text{m}^3$ .

## P

- **Pandas:** Librería de Python para manipulación y análisis de datos, que proporciona estructuras de datos como DataFrames.
- **palette (Seaborn):** Parámetro para especificar el esquema de colores a usar en los gráficos.
- **P (Presión Atmosférica):** Presión ejercida por el aire sobre la superficie. Unidad común: hPa o mbar.

- **percentil\_98 (Calidad del Aire):** Valor por debajo del cual se encuentran el 98% de los datos. Útil para evaluar picos de contaminación.
- **PLíquida (Precipitación Líquida):** Cantidad de lluvia recogida en un periodo.  
Unidad común: mm.
- **Plotly:** Librería de Python para crear gráficos interactivos y visualizaciones web.
- **PM10 (Partículas Menores a 10 Micras):** Material particulado con diámetro aerodinámico  $\leq 10 \mu\text{m}$ . Afecta vías respiratorias superiores. Unidad común:  $\mu\text{g}/\text{m}^3$ .
- **PM2.5 (Partículas Menores a 2.5 Micras):** Material particulado muy fino ( $\leq 2.5 \mu\text{m}$ ). Puede penetrar los pulmones y llegar al torrente sanguíneo. Unidad común:  $\mu\text{g}/\text{m}^3$ .
- **PolynomialFeatures (Scikit-learn):** Clase de Scikit-learn para generar características polinómicas e interacciones entre las características existentes.
- **porcentaje\_excedencias (Calidad del Aire):** Porcentaje que representa cuántas veces se excedió el límite respecto al total de mediciones.
- **Pregunta de Investigación:** La cuestión central que el proyecto busca responder a través de su análisis.
- **promedio (Calidad del Aire):** Valor promedio anual del contaminante para ese año y estación.
- **PST (Partículas Totales en Suspensión):** Incluye todas las partículas en suspensión en el aire, sin limitarse al tamaño. Unidad común:  $\mu\text{g}/\text{m}^3$ .
- **Python:** Lenguaje de programación de alto nivel, interpretado, cuya filosofía hace hincapié en una sintaxis muy limpia y que favorezca un código legible.

## R

- **R2 (R-cuadrado):** Métrica de evaluación de modelos de regresión que indica la proporción de la variabilidad en la variable dependiente que es predecible a partir de las variables independientes.
- **RandomForestRegressor (Scikit-learn):** Clase de Scikit-learn que implementa el algoritmo de Random Forest para tareas de regresión.
- **Recomendaciones:** Sugerencias y acciones propuestas basadas en las conclusiones del proyecto para abordar las problemáticas identificadas.

- **Referencias:** Listado de todas las fuentes bibliográficas y de datos consultadas y utilizadas en el proyecto.
- **Relevancia:** Importancia o significado del proyecto en su contexto, destacando por qué es necesario o valioso.
- **representatividad (Calidad del Aire):** Porcentaje de cobertura de datos frente al total esperado. Idealmente cercano al 100%.
- **Resultados y Discusión:** Sección donde se presentan los hallazgos clave del análisis de datos y modelado, y se discute su significado e implicaciones.
- **RGlobal (Radiación Solar Global):** Cantidad total de energía solar recibida sobre una superficie horizontal. Unidad común: W/m<sup>2</sup>.
- **RMSE (Error Cuadrático Medio):** Métrica de evaluación de modelos de regresión que mide la magnitud promedio de los errores del modelo.

## S

- **Scikit-learn:** Librería de Python para aprendizaje automático que proporciona herramientas para modelado predictivo.
- **Seaborn:** Librería de Python para visualización de datos basada en Matplotlib, que ofrece una interfaz de alto nivel para crear gráficos estadísticos atractivos e informativos.
- **sharey (Seaborn):** Parámetro en funciones de Seaborn como catplot que controla si las facetas comparten el mismo eje Y.
- **SO<sub>2</sub> (Dióxido de Azufre):** Contaminante asociado a la quema de combustibles fósiles (carbón, diésel). Unidad común: µg/m<sup>3</sup>.
- **Streamlit:** Framework de Python de código abierto para crear aplicaciones web interactivas y dashboards de ciencia de datos con poco código.
- **suma (Calidad del Aire):** Suma total de todas las mediciones del contaminante durante el año.

## T

- **TAire2 (Temperatura del Aire):** Medición de la temperatura ambiente a cierta altura. Unidad común: °C.



- **template (Plotly):** Parámetro en Plotly para aplicar un tema visual predefinido al gráfico (ej. 'plotly\_dark').
- **tiempo\_exposicion (Calidad del Aire):** Número de horas durante las cuales se midieron los datos para ese registro, reflejando la cobertura temporal del monitoreo.
- **tipo\_estacion (Calidad del Aire):** Clasificación de la estación de monitoreo (ej. Urbana, Suburbana, Rural).
- **title (Plotly/Matplotlib):** Título principal del gráfico.
- **train\_test\_split (Scikit-learn):** Función para dividir un dataset en conjuntos de entrenamiento y prueba.

## U

- **ubicacion (Calidad del Aire):** Dirección o descripción más detallada de dónde está la estación.
- **unidades (Calidad del Aire):** Unidad de medida del contaminante (ej.  $\mu\text{g}/\text{m}^3$ , ppm).
- **UNIDADES\_CONTAMINANTES:** Diccionario que mapea el nombre de cada contaminante o variable a su unidad de medida común.
- **use\_container\_width (Streamlit):** Parámetro para que un gráfico o DataFrame ocupe todo el ancho disponible del contenedor en Streamlit.

## V

- **variable (Calidad del Aire):** Contaminante medido (ej. PM10, PM2.5, NO2, SO2, CO, O3, etc.).
- **Visualizaciones Gráficas:** Representaciones visuales de los datos (gráficos de líneas, barras, boxplots, etc.) para facilitar la comprensión de patrones y tendencias.
- **VViento (Velocidad del Viento):** Rapidez con la que se mueve el aire. Puede influir en la dispersión de contaminantes. Unidad común: m/s o km/h.



▶ TALENTO  
TECH



Universidad  
Tecnológica  
de Bolívar

[www.utb.edu.co/talento-tech](http://www.utb.edu.co/talento-tech)