

JOSE DAVID CARDONA MAZO - Avance 1

Proyecto Integrado III - Analítica de Datos (Sharon Karin Camacho) -  
PREICA2502B020071

UNIVERSIDAD DIGITAL DE ANTIOQUIA - 2025

Primer Avance	
Tema	Ambiente y Desarrollo Sostenible - Calidad Del Aire En Colombia (Promedio Anual)
Link Dataset	<a href="https://www.datos.gov.co/resource/kek-d-7v7h.json?">https://www.datos.gov.co/resource/kek-d-7v7h.json?</a>
Columnas	28733
Fuente	<a href="#">Portal de Datos Abiertos de Colombia</a>
Objetivo de análisis:	Evaluar la influencia de las variables meteorológicas (temperatura, precipitación, viento, radiación solar, humedad) en la concentración y variabilidad de los contaminantes atmosféricos (PM10, PM2.5, NO2, SO2, CO, O3, entre otros) en las estaciones de monitoreo de calidad del aire de Colombia, para identificar patrones y predecir posibles excedencias de los límites establecidos por las normativas ambientales.
Pregunta de Investigación:	¿Cómo afectan las variables meteorológicas (temperatura, precipitación, viento, radiación solar, humedad) a la concentración y variabilidad de los contaminantes atmosféricos (PM10, PM2.5, NO2, SO2, CO, O3, entre otros) en las estaciones de monitoreo de calidad del aire en Colombia?
Hipótesis	Las condiciones meteorológicas (temperatura, precipitación, viento, radiación solar y humedad) tienen un impacto significativo en las concentraciones y la variabilidad de los contaminantes atmosféricos medidos, donde cambios en las condiciones climáticas, como temperaturas extremas o baja humedad, se correlacionan con niveles más altos de contaminación.
Métricas de Éxito:	Correlación entre las condiciones meteorológicas y los contaminantes ( $R^2$ ).
	Número de días de excedencia de los límites de calidad del aire.
	Desviación estándar de los contaminantes y su relación con las condiciones meteorológicas.
	Precisión del modelo predictivo (evaluado con RMSE o $R^2$ ).
Qué datos tengo para lograr el objetivo	Sí, estos datos son útiles para lograr el objetivo del análisis, Las variables incluyen tanto contaminantes como factores meteorológicos, lo que nos permite analizar cómo afectan las condiciones ambientales a la calidad del aire

## DICCIONARIO DE DATOS

Columna	Descripción	Tipo de Dato	Ejemplo
<b>id_estacion</b>	Identificador único asignado a la estación de monitoreo.	Numérico	20383
<b>autoridad_ambiental</b>	Autoridad ambiental encargada de la estación.	Texto	CORANTIOQUIA
<b>estacion</b>	Nombre de la estación de monitoreo.	Texto	ALTAVISTA
<b>latitud</b>	Coordenada geográfica que indica la latitud de la estación.	Numérico	6.222.584
<b>longitud</b>	Coordenada geográfica que indica la longitud de la estación.	Numérico	-75.628.207
<b>variable</b>	Contaminante o variable meteorológica medida en la estación (PM10, PM2.5, temperatura, etc.).	Texto	PM10
<b>unidades</b>	Unidad de medida para la variable.	Texto	ug/m3
<b>tiempo_de_exposicion_horas</b>	Horas de exposición durante el monitoreo.	Numérico	1
<b>a_o</b>	Año en el que se realizaron las mediciones.	Numérico	2011
<b>promedio</b>	Promedio anual de la medición de la variable (contaminante o meteorológica).	Numérico	75.5
<b>suma</b>	Total acumulado anual de la medición de la variable.	Numérico	329454.0
<b>no_de_datos</b>	Número total de datos válidos recopilados en ese año de medición.	Numérico	4363
<b>representatividad_temporal</b>	Porcentaje del tiempo que tiene datos válidos registrados.	Categórico	50%
<b>excedencias_limite_actual</b>	Número de veces que el contaminante ha superado el límite permitido en ese año.	Numérico	0
<b>porcentaje_excedencias_limite</b>	Porcentaje de veces que se superó el límite permitido durante el año.	Numérico	0.0%
<b>mediana</b>	Mediana de la medición del contaminante durante el año.	Numérico	64.0
<b>percentil_98</b>	Percentil 98 de la medición del contaminante durante el año (valor por debajo del cual está el 98% de los datos).	Numérico	208.0
<b>maximo</b>	Valor máximo de la medición durante el año.	Numérico	524.0
<b>fechas_horas_del_maximo</b>	Fecha y hora del valor máximo registrado durante el año.	Fecha/Tiempo	27/10/2011 8:00:00 p. m.
<b>minimo</b>	Valor mínimo de la medición durante el año.	Numérico	7.0

<b>fechas_horas_del_minimo</b>	Fecha y hora del valor mínimo registrado durante el año.	Fecha/Tiempo	27/10/2011 8:00:00 p. m.
<b>dias_de_excedencias</b>	Número de días en que se superó el límite permitido.	Numérico	0
<b>codigo_del_departamento</b>	Código DANE del departamento donde se encuentra la estación.	Numérico	5
<b>nombre_del_departamento</b>	Nombre del departamento.	Texto	ANTIOQUIA
<b>codigo_del_municipio</b>	Código DANE del municipio donde se encuentra la estación.	Numérico	5001
<b>nombre_del_municipio</b>	Nombre del municipio.	Texto	MEDELLÍN
<b>tipo_de_estacion</b>	Tipo de estación de monitoreo (fija o móvil).	Texto	Fija
<b>ubicacion</b>	Información geoespacial de la estación (coordenadas en formato JSON).	Geoespacial	{'type': 'Point', 'coordinates': [-75.628207, 6.222584]}