



Análisis Exploratorio y Limpieza de Datos de Contaminación Atmosférica en Lima

CONTEXTO

El *dataset* que se presenta contiene datos de contaminación del aire en Lima. Los datos incluyen mediciones horarias de varios contaminantes atmosféricos como PM10, PM2.5, SO2, NO2, O3, y CO, en diferentes estaciones de monitoreo. El objetivo de este ejercicio es realizar un análisis de la calidad del aire y determinar patrones de contaminación en las distintas estaciones, así como observar si existen tendencias significativas entre las concentraciones de contaminantes

OBJETIVOS

- Comprender la estructura de los datos y su significado
- Limpiar los datos para eliminar inconsistencias, valores faltantes o atípicos
- Visualizar los datos usando la librería *seaborn* para identificar patrones o relaciones entre variables

EXPLORACIÓN DE LOS DATOS

- Dataset proporcionado: [datos_horarios_contaminacion_lima.csv](#)
- Los datos contienen las siguientes columnas:
 - **CODIGO ESTACION:** Código único de la estación de monitoreo
 - **ESTACION:** Nombre de la estación de monitoreo
 - **ANO, MES, DIA, HORA:** Fecha y hora de la medición
 - **PM10, PM2.5, SO2, NO2, O3, CO:** Concentraciones de contaminantes en el aire (valores en partes por millón o microgramos por metro cúbico, dependiendo del contaminante)

TAREAS PARA REALIZAR

1. Carga del dataset
2. Revisión del dataset
3. Descripción general (descripción estadística)
4. Limpieza de los datos
 - 4.1. Valores faltantes
 - 4.1.1. Identificar qué columnas tienen valores nulos (NaN) y la cantidad de valores faltantes.
 - 4.1.2. Decidir cómo manejar los valores faltantes. Pueden imputar valores con la media o la mediana o simplemente eliminar las filas/columnas con demasiados datos faltantes
5. Análisis de relación entre variables
 - 5.1. Correlación entre contaminantes
 - 5.1.1. Calcular la matriz de correlación entre los diferentes contaminantes para ver si hay relaciones fuertes entre ellos (por ejemplo, entre PM10 y PM2.5).
 - 5.1.2. Visualizar la correlación usando un *heatmap* de seaborn
6. Visualización de la contaminación
 - 6.1. Gráfico de la distribución de contaminantes
 - 6.1.1. Usar *seaborn* para visualizar la distribución de PM10 o PM2.5 a lo largo del tiempo (puedes usar la variable HORA para representar la hora del día)
 - 6.1.2. Crear un gráfico de líneas que muestre la evolución de uno o más contaminantes en una estación específica
7. Conclusiones
 - 7.1. En esta parte, deberás redactar tus conclusiones basadas en el análisis exploratorio y las visualizaciones realizadas. Algunas preguntas que pueden ayudar son:
 - 7.1.1. ¿Hay algún patrón claro en las concentraciones de contaminantes a lo largo del día?
 - 7.1.2. ¿Qué tipo de tratamiento adicional podría aplicarse a los datos para mejorar su calidad o análisis?