

Proyecto Final / Emisiones

Introducción

Este proyecto tiene como objetivo el análisis de los datos de emisiones de contaminación obtenidos a lo largo de los años 2020 a 2023 para diversas estaciones de medición. Aprenderás a trabajar con bases de datos, realizar transformaciones de datos, y obtener información clave a través de consultas SQL. El proyecto permite explorar cómo evolucionan los niveles de contaminación a lo largo del tiempo y las diferencias entre las estaciones de medición.

Los datos de contaminación están organizados en tablas anuales (2020, 2021, 2022 y 2023), y se requiere combinar estos datos en una única tabla de resumen, facilitando el análisis posterior de los mismos. Este ejercicio final del módulo está diseñado para que los estudiantes apliquen sus conocimientos sobre bases de datos, manejo de funciones SQL avanzadas, funciones de ventana, y consultas agrupadas para obtener información significativa.

Proyecto Final / Emisiones

Diccionario de Variables:

- **PROVINCIA:** Nombre de la provincia en la que se encuentra la estación de medición.
- **MUNICIPIO:** Nombre del municipio en el que se encuentra la estación de medición.
- **ESTACION:** Nombre o código de la estación de medición de calidad del aire o datos ambientales.
- **MAGNITUD:** Tipo de magnitud o parámetro medido (ej. concentración de un contaminante).
- **PUNTO_MUESTREO:** Ubicación específica dentro de la estación donde se realiza la medición (puede ser un punto geográfico o una zona de la estación).
- **ANO:** Año en el que se tomó la medición.
- **MES:** Mes en el que se tomó la medición (de 1 a 12).
- **D01 a D31:** Columnas que representan los valores diarios de cada día del mes para un determinado parámetro medido (puede ser cualquier tipo de medición de calidad del aire, como concentración de contaminantes, temperatura, humedad, etc.). Cada columna corresponde a un día específico del mes (D01 para el día 1, D02 para el día 2, ..., D31 para el día 31).
- **V01 a V31:** Similar a las columnas D01 a D31, estas columnas pueden representar los valores de alguna variable adicional o diferente asociada a los valores diarios de cada día del mes. Podría ser una variable derivada o un segundo conjunto de mediciones relacionadas con los datos de calidad del aire u otras mediciones ambientales.

Proyecto Final / Emisiones

Descripción de las Preguntas y Objetivos

A continuación, se presentan las preguntas a resolver en este proyecto. Estas preguntas te guiarán a través de diversos ejercicios de análisis de datos, transformaciones y cálculos descriptivos en SQL.

1. Combinación de datos de diferentes años

Tareas:

- A. Crea la base de datos y la tabla total.
- B. Inserta los datos de las tablas de los años 2020, 2021, 2022 y 2023 en la tabla total.
- C. Verifica que todos los datos se han insertado correctamente, mostrando el número total de registros y los años disponibles en la tabla.

2. Creación de la columna valor_dia

Tareas:

- A. Añade la columna valor_dia a la tabla total.
- B. Rellena esta columna con los valores de dia d01

3. Añadir una columna fecha

Tareas:

- A. Crea la columna fecha en la tabla total con formato YYYY-MM-DD
- B. Actualiza la columna fecha con los valores correspondientes a cada mes y año

Proyecto Final / Emisiones

4. Consultar estaciones y contaminantes disponibles

Tarea:

- A. Muestra las estaciones y los contaminantes únicos que existen en la tabla total

5. Filtrar datos por estación y contaminante

Tarea:

- A. compara los valores diarios de contaminación entre las estaciones "Ramon y Cajal" y "Escuelas Aguirre" durante el año 2020, calcula la diferencia entre los valores de ambas estaciones para cada fecha y devuelve las columnas: fecha, valor de "Ramon y Cajal", valor de "Escuelas Aguirre" y la diferencia de valores, filtrando solo los registros donde la magnitud sea 1 y el rango de fechas sea del 1 de enero de 2020 al 31 de diciembre de 2020, ordenando los resultados por fecha.

6. Resumen descriptivo por contaminante (Magnitud)

Tarea:

- A. Muestra el valor mínimo, máximo, promedio y la desviación estándar de los valores de contaminación para cada contaminante.

Proyecto Final / Emisiones

7. Resumen descriptivo por estación

Tarea:

- A. Muestra el valor mínimo, máximo, promedio y la desviación estándar de los valores de contaminación para cada estación

8. Calcular medias mensuales de contaminación

Tarea:

- A. ¿Cómo podemos calcular el promedio mensual de valores por estación y año, transformando los números de los meses en palabras y agrupándolos además en trimestres?

9. Medias mensuales por estación con nombre largo

Tarea:

- A. Muestra la media mensual de contaminación para estaciones con nombres largos (más de 10 caracteres), agrupado por estación, contaminante y mes.

10. Niveles de contaminación acumulados por estación y contaminante

Tarea:

- A. calcula la media anual de las emisiones (valor_dia) por estación y magnitud para cada año, y asigna un ranking de emisiones dentro de cada año y magnitud, donde el primer lugar corresponde al valor más alto. Ordena los resultados por año, magnitud y ranking de emisiones.

Proyecto Final / Emisiones

11. Promedio acumulado de emisiones por estación

Tarea:

- A. ¿Cómo podemos calcular el promedio acumulado anual y el promedio acumulado total de emisiones por estación a lo largo de los años?

12. Días con datos de contaminación por estación y mes

Tarea:

- A. Muestra el número de días con datos de contaminación registrados por estación y nombre del mes

13. Días transcurridos desde la última medición por estación

Tarea:

- A. ¿Cómo podemos identificar la última fecha de registro de datos para cada estación y calcular el número de días transcurridos desde esa fecha hasta hoy?

14. Variación de contaminación entre días anteriores y posteriores

Tarea:

- A. Muestra cómo varió la contaminación de cada estación y contaminante en comparación con el día anterior y el día siguiente, y también muestra el primer y último valor registrado de cada estación y contaminante durante el año.

Proyecto Final / Emisiones

OPCIONAL

15. Usando la tabla total, realiza lo siguiente en un solo query:

- Calcula el promedio anual de emisiones (media_anual) para cada estación (estación) y magnitud (magnitud), solo para la magnitud 1. Considera solo los años 2020 y 2021.
- Genera un ranking por año y magnitud, ordenando las estaciones por su media_anual de emisiones (de mayor a menor).
- Determina el mes con la mayor media mensual de emisiones para cada estación en los años 2020 y 2021.
- Combina todos estos cálculos en un único query que incluya: ano, magnitud, estacion, media_anual, ranking_emisiones, el mes con la mayor media mensual de emisiones (mes_max) y el valor promedio mensual de ese mes (max_media_mensual).
- Ordena los resultados por año, magnitud y ranking_emisiones.