**Proyecto Bedu (Bases de Datos)**

**Modulo 1**

**“El impacto sanitario de la calidad del aire”**

La calidad del aire es un problema crítico que afecta la salud pública a nivel global. La exposición a contaminantes del aire, como el dióxido de nitrógeno (NO2), el ozono (O3), y las partículas en suspensión (PM2.5 y PM10), está vinculada a una variedad de problemas de salud, incluyendo enfermedades respiratorias, cardiovasculares y cáncer. En muchas ciudades y regiones, la calidad del aire no se monitorea adecuadamente, lo que impide a las autoridades y a la población tomar medidas informadas para proteger su salud. Además, la falta de datos accesibles y en tiempo real sobre la calidad del aire dificulta la implementación de políticas y estrategias efectivas para reducir la contaminación.

**¿Cuál es la necesidad o problema para resolver? 🚀**

La necesidad para resolver es la falta de monitoreo adecuado y accesible de la calidad del aire, que impacta negativamente la salud pública, provocando enfermedades respiratorias, cardiovasculares y cáncer.

**¿Cuál es la propuesta para resolver esa necesidad o problema? 💡**

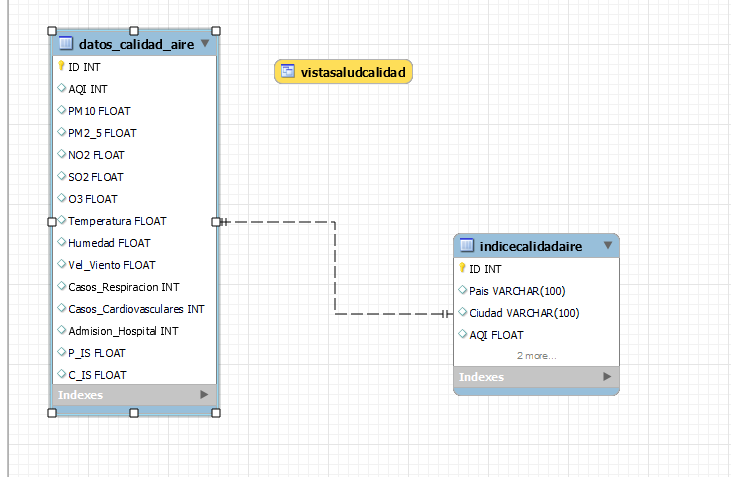
El análisis de bases de datos en MySQL con tablas relacionales y datos sobre la calidad del aire permite identificar tendencias y patrones, detectar anomalías y realizar análisis de correlación entre contaminantes y variables ambientales. Este análisis facilita la segmentación y comparación de datos, generando informes y visualizaciones útiles para la toma de decisiones. Además, proporciona una base sólida para exportar datos a herramientas de modelado predictivo, ayudando a prever futuros niveles de contaminación y evaluar el impacto de intervenciones, contribuyendo así a la mejora de la salud pública y la calidad del aire.

**¿Cómo lo vas a hacer? 📋**

* Elección de fuente de datos

Se busco información de un Dataset a analizar en Kaggle para visualizar la tendencia de este problema publico de salud y hacer una metodología de análisis.

* Definición de la estructura de la base de datos
  1. Datos de Calidad de Aire con parámetros de concentración, estadísticos y parámetros físicos
  2. Datos de pacientes, ubicaciones y ID involucrados
* Diagrama de estructura de Bases de datos en MySQL



* Carga de datos

La carga de datos (Datos sobre el impacto sanitario de la calidad del aire) se descargó de Kaggle y se cargó en MySQL.

**Métricas de calidad del aire**

AQI: Índice de Calidad del Aire, una medida de la contaminación del aire actual o prevista.

PM10: Concentración de partículas de menos de 10 micrómetros de diámetro (μg/m³).

PM2.5: Concentración de partículas de menos de 2,5 micrómetros de diámetro (μg/m³).

NO2: Concentración de dióxido de nitrógeno (ppb).

SO2: Concentración de dióxido de azufre (ppb).

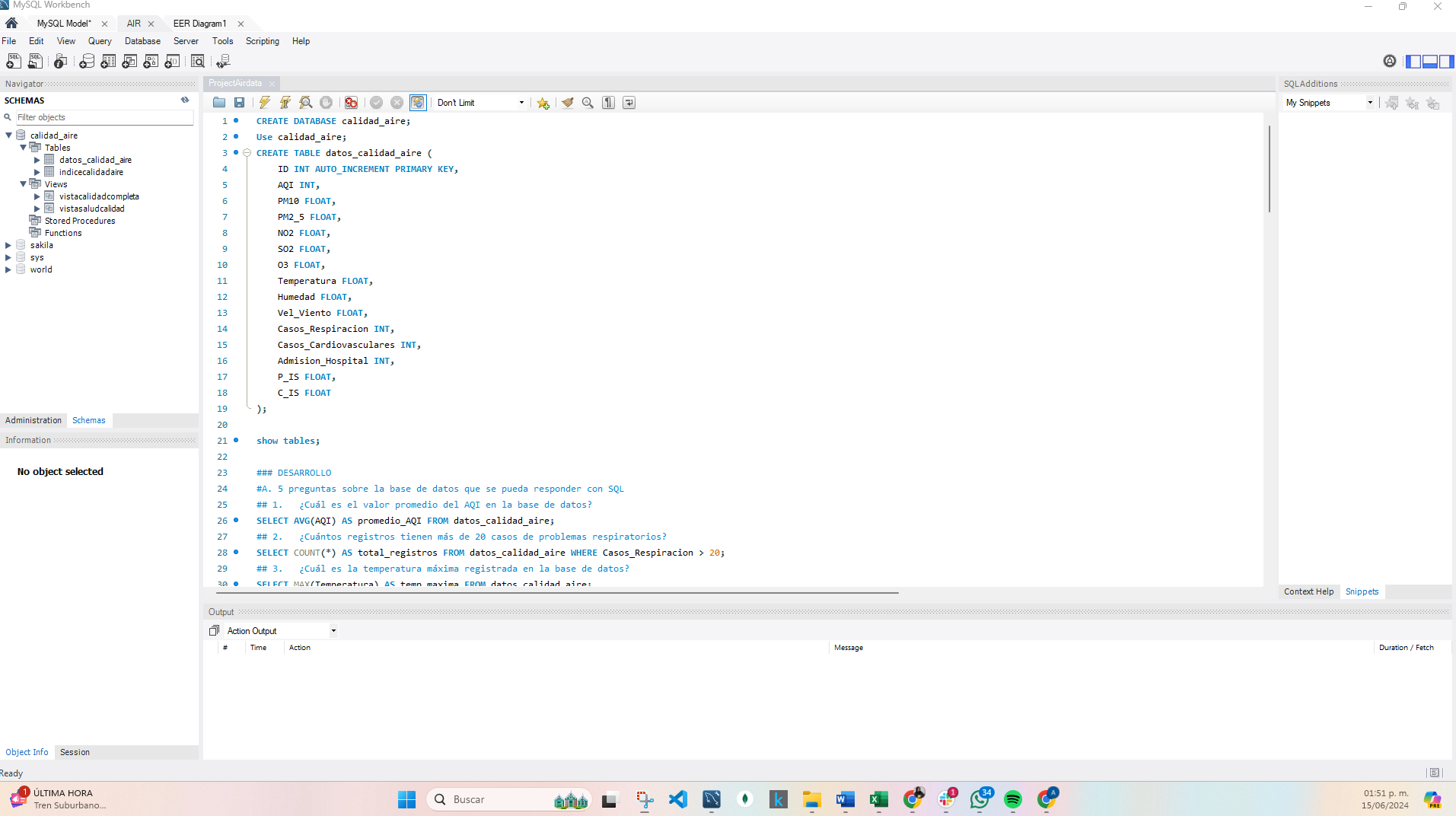
O3: Concentración de ozono (ppb).

**Referencias**

* Rabie El Kharoua. (2024). 🌍 Air Quality and Health Impact Dataset🌍 [Data set]. Kaggle. <https://doi.org/10.34740/KAGGLE/DSV/8675842>

**Desarrollo**

**(Primeras consultas)**



**A. 5 preguntas sobre la base de datos que se pueda responder con SQL**

1. **¿Cuál es el valor promedio del AQI en la base de datos?**

SELECT AVG(AQI) AS promedio\_AQI FROM datos\_calidad\_aire;

1. **¿Cuántos registros tienen más de 20 casos de problemas respiratorios?**

SELECT COUNT(\*) AS total\_registros FROM datos\_calidad\_aire WHERE Casos\_Respiracion > 20;

1. **¿Cuál es la temperatura máxima registrada en la base de datos?**

SELECT MAX(Temperatura) AS temp\_maxima FROM datos\_calidad\_aire;

1. **¿Cuántos registros hay en total en la base de datos?**

SELECT COUNT(\*) AS total\_registros FROM datos\_calidad\_aire;

1. **¿Cuál es el valor mínimo de PM2?5 en la base de datos?**

SELECT MIN(PM2\_5) AS minimo\_PM2\_5 FROM datos\_calidad\_aire;

**B. 5 preguntas con reconocimiento de patrones, funciones de agrupamiento, agrupamiento y subconsultas.**

1. **¿Cuál es el valor promedio de AQI por cada nivel de admisión hospitalaria?**

SELECT Admision\_Hospital, AVG(AQI) AS promedio\_AQI

FROM datos\_calidad\_aire

GROUP BY Admision\_Hospital;

1. **¿Cuántos registros tienen una temperatura mayor al promedio de todas las temperaturas?**

SELECT COUNT(\*) AS total\_registros

FROM datos\_calidad\_aire

WHERE Temperatura > (SELECT AVG(Temperatura) FROM datos\_calidad\_aire);

1. **¿Cuál es la relación entre el promedio de PM10 y PM2?5 por nivel de casos de problemas cardiovasculares?**

SELECT Casos\_Cardiovasculares, AVG(PM10) AS promedio\_PM10, AVG(PM2\_5) AS promedio\_PM2\_5

FROM datos\_calidad\_aire

GROUP BY Casos\_Cardiovasculares;

1. **¿Cuál es la velocidad del viento promedio para los días con más de 5 admisiones hospitalarias?**

SELECT AVG(Vel\_Viento) AS velocidad\_promedio

FROM datos\_calidad\_aire

WHERE Admision\_Hospital > 5;

1. **¿Cuántos casos de problemas respiratorios y cardiovasculares hay por cada nivel de calidad del aire (AQI)?**

SELECT AQI, SUM(Casos\_Respiracion) AS total\_respiratorios, SUM(Casos\_Cardiovasculares) AS total\_cardiovasculares

FROM datos\_calidad\_aire

GROUP BY AQI;

**C. 5 preguntas complejas de uso de llaves primarias. Tipo de relaciones, relación de tablas mediante joins y creación de vistas.**

1. **Crear una vista que combine datos de calidad del aire con información de índice de calidad del aire**

CREATE VIEW VistaCalidadCompleta AS

SELECT

ICA.ID,

ICA.Pais,

ICA.Ciudad,

ICA.AQI AS AQI\_Indice,

ICA.Categoria\_AQI,

DCA.AQI AS AQI\_Detalle,

DCA.PM10,

DCA.PM2\_5,

DCA.NO2,

DCA.SO2,

DCA.O3,

DCA.Temperatura,

DCA.Humedad,

DCA.Vel\_Viento,

DCA.Casos\_Respiracion,

DCA.Casos\_Cardiovasculares,

DCA.Admision\_Hospital

FROM

IndiceCalidadAire ICA

JOIN

datos\_calidad\_aire DCA ON ICA.ID = DCA.ID;

1. **Obtener el promedio de los contaminantes por país**

SELECT

Pais,

AVG(PM10) AS Promedio\_PM10,

AVG(PM2\_5) AS Promedio\_PM2\_5,

AVG(NO2) AS Promedio\_NO2,

AVG(SO2) AS Promedio\_SO2,

AVG(O3) AS Promedio\_O3

FROM

VistaCalidadCompleta

GROUP BY

Pais;

1. **Contar el número de ciudades en cada categoría de AQI**

SELECT

Categoria\_AQI,

COUNT(DISTINCT Ciudad) AS Numero\_Ciudades

FROM

VistaCalidadCompleta

GROUP BY

Categoria\_AQI;

1. **Crear una vista para analizar la relación entre calidad del aire y casos de salud**

CREATE VIEW VistaSaludCalidad AS

SELECT

Pais,

Ciudad,

AQI\_Indice,

Categoria\_AQI,

Casos\_Respiracion,

Casos\_Cardiovasculares,

Admision\_Hospital

FROM

VistaCalidadCompleta;

1. **¿Cuál es la ciudad con el peor índice de calidad del aire (AQI\_Indice) y el mayor número de admisiones hospitalarias?**

SELECT

Ciudad,

MAX(AQI\_Indice) AS Peor\_AQI,

MAX(Admision\_Hospital) AS Max\_Admisiones

FROM

VistaSaludCalidad

GROUP BY

Ciudad

ORDER BY

Peor\_AQI DESC, Max\_Admisiones DESC

LIMIT 1;

**Conclusión**

El análisis de los datos de calidad del aire mediante consultas SQL revela insights cruciales sobre la relación entre la contaminación y la salud pública. Al calcular promedios, máximos y mínimos de contaminantes como AQI, PM2.5 y PM10, se puede evaluar la calidad general del aire y su impacto potencial en la salud. Las funciones de agrupamiento y subconsultas permiten identificar patrones específicos y correlacionar altos niveles de contaminación con aumentos en casos de problemas respiratorios y cardiovasculares, así como con admisiones hospitalarias.

Además, el uso de vistas y JOIN para combinar datos de calidad del aire con información de ubicación y pacientes proporciona un análisis más detallado y contextual. Esto facilita la identificación de áreas geográficas más afectadas y permite a las autoridades locales tomar decisiones informadas para mitigar la contaminación y mejorar la salud pública. En resumen, el análisis detallado de estos datos permite una mejor comprensión de la situación actual y la implementación de estrategias efectivas para reducir la contaminación y proteger la salud de la población.