

# Enunciado del ejercicio

**Título:** Evolución mensual de la calidad del aire en Santiago (2018–2024) con enfoque de desigualdad ambiental

**Stack:** Google Colab · PostgreSQL (Neon) · Python (pandas, SQLAlchemy, matplotlib) · SQL

## 1) Contexto y propósito

La calidad del aire en Santiago afecta de forma desigual a sus comunas. El objetivo de este ejercicio es **medir, visualizar y explicar** esa desigualdad entre 2018 y 2024, generando evidencia útil para sensibilización ciudadana y toma de decisiones.

## 2) Objetivo general

Construir un flujo reproducible (Colab → Neon) que:

1. cargue y modele datos de PM2.5,
2. calcule métricas mensuales por comuna,
3. compare periodos **pre/covid/post**,
4. comunique hallazgos con visualizaciones y texto.

## 3) Preguntas/KPIs a responder

- **KPI 1:** Promedio **mensual** de PM2.5 por **comuna** (2018–2024).
- **KPI 2:** % de **días críticos** por comuna y mes (define umbral, p. ej., promedio diario  $\text{PM}_{2.5} > 15 \mu\text{g}/\text{m}^3$ ).
- **KPI 3:** Comparación por bloques temporales: **pre** (2019), **covid** (2020–2021), **post** (2022–2024).

## 4) Datos requeridos (mínimo viable)

- **Mediciones:** PM2.5 por estación en comunas de Santiago ( $\geq 3$  estaciones).
- **Periodo:** 2018-01-01 a 2024-12-31.
- **Granularidad original:** horaria o diaria (agruparás luego).
- **Fuentes típicas:** SINCA (CSV por estación).  
(Opcional) clima básico (temperatura/viento) si lo tienes.

## 5) Esquema de tablas en PostgreSQL (Neon)

Crea este esquema (o equivalente):

```
CREATE TABLE IF NOT EXISTS stations (  
  station_id TEXT PRIMARY KEY,  
  nombre TEXT NOT NULL,  
  comuna TEXT NOT NULL,  
  lat REAL,  
  lon REAL  
);  
  
CREATE TABLE IF NOT EXISTS measurements (  
  station_id TEXT REFERENCES stations(station_id),  
  ts TIMESTAMP NOT NULL,  
  pm25 REAL,  
  pm10 REAL,  
  temp REAL,  
  wind_ms REAL  
);
```

Reglas de calidad mínimas al cargar:

- **ts** válido; • **pm25** en [0, 1000]; • sin duplicados exactos (**station\_id,ts**).

## 6) Tareas (paso a paso)

1. **Preparación (Colab):** instalar dependencias (`psycopg2-binary`, `sqlalchemy`) y conectar a Neon con `?sslmode=require`.
2. **Modelado:** crear tablas `stations` y `measurements`.
3. **ETL:** leer CSV(s), normalizar columnas (fechas, nombres, unidades), validar nulos/outliers y cargar a `measurements`.
4. **Consultas SQL (aprendizaje guiado):**
  - **Q1:** Promedio **mensual** de PM2.5 por **comuna**.
  - **Q2:** % de **días críticos** por comuna y mes (umbral elegido).
  - **Q3:** Promedio de PM2.5 por **bloque temporal** (`pre`, `covid`, `post`) y comuna.
5. **Validaciones en pandas:** comprobar columnas, tipos, rangos y presencia de datos por periodo.
6. **Visualizaciones:**
  - Serie mensual por comuna (líneas).
  - Barras comparando bloques temporales por comuna.
7. **Comunicación:** escribir 3–5 hallazgos (desigualdad entre comunas, efecto COVID, estacionalidad) y 2–3 limitaciones (cobertura, nulos, calidad).

## 7) Formato de entrega (repo)

```
/calidad-aire-santiago
├── notebooks/analisis_calidad_aire.ipynb
├── sql/schema.sql
├── sql/consultas_base.sql    # skeleton con TODOs (ver abajo)
├── sql/consultas_solucion.sql # solución comentada (para revisión)
├── data/ (opcional, si es liviano o con enlaces)
├── imgs/ (gráficos exportados)
├── requirements.txt (opcional)
└──
```

Badge Colab (README)

[![Abrir en Colab](https://colab.research.google.com/assets/colab-badge.svg)]  
(https://colab.research.google.com/github/USUARIO/REPO/blob/main/notebooks/analisis\_calidad\_aire.ipynb)

## 8) Skeleton SQL con TODOs (para **sql/consultas\_base.sql**)

```
-- Q1) Promedio mensual de PM2.5 por comuna
.
.
.
.

-- Q2) % de días críticos (> umbral) por comuna y mes
.
.
.

-- Q3) Comparación 2019 vs 2020–2021 vs 2022–2024
.
.
.
```