**Portada**

**Nombre del curso:** Curso Básico de Introducción al Big Data

**Título de la tarea:** Proyecto Final

**Nombre del docente:** Prof. Ing. Richard D. Jiménez-R.

**Nombre de los estudiantes:** Orlando Semidei, Néstor Morel, María Luján Ibarra Benegas

**Identificador de grupo:** 9

**Fecha de entrega:** 09 de mayo de 2025

**Análisis de la distribución de vacunas**

# **1. Introducción**

Este proyecto tiene como objetivo analizar la distribución de vacunas aplicadas a la población, utilizando un conjunto de datos público de aproximadamente 1.000.000 registros. El análisis busca responder preguntas como cuáles son los establecimientos con mayor actividad de vacunación, qué tipos de vacunas se aplicaron con mayor frecuencia y cómo varió la vacunación a lo largo del tiempo.

# **2. Arquitectura del Sistema y Modelo de Datos**

Se utilizó PostgreSQL como sistema de base de datos. Se diseñó una tabla principal llamada 'vacunados' con los siguientes campos: id, nombre, apellido, cedula, establecimiento, fecha\_aplicacion, dosis, descripcion\_vacuna, actualizado\_al.

A continuación, se presenta el código SQL para la creación de la tabla:

CREATE TABLE vacunados (  
 id SERIAL PRIMARY KEY,  
 nombre VARCHAR(100),  
 apellido VARCHAR(100),  
 establecimiento VARCHAR(100),  
 fecha\_aplicacion TIMESTAMP,  
 cedula VARCHAR(20) UNIQUE,  
 dosis INTEGER,  
 descripcion\_vacuna VARCHAR(100),  
 actualizado\_al TIMESTAMP  
);

# **3. Proceso ETL (Extracción, Transformación y Carga)**

Se utilizó Pentaho Kettle para el proceso ETL. Se realizó la extracción desde un archivo CSV, seguido de transformaciones como limpieza de columnas, normalización de nombres de vacunas, conversión de fechas y eliminación de duplicados. Finalmente, los datos se cargaron en la tabla 'vacunados'.

# **4. Análisis y KPIs**

A continuación se presentan algunas consultas SQL utilizadas en el análisis:

* Vacunas aplicadas por mes

SELECT DATE\_TRUNC('month', fecha\_aplicacion) AS mes, COUNT(\*) AS total\_vacunados  
FROM vacunados  
GROUP BY mes  
ORDER BY mes;

* Personas vacunadas por establecimiento

SELECT establecimiento, COUNT(\*) AS total  
FROM vacunados  
GROUP BY establecimiento  
ORDER BY total DESC  
LIMIT 10;

* Distribución por tipo de vacuna

SELECT descripcion\_vacuna, COUNT(\*) AS cantidad  
FROM vacunados  
GROUP BY descripcion\_vacuna  
ORDER BY cantidad DESC;

* Promedio de dosis aplicadas por persona

SELECT AVG(dosis) FROM vacunados;

# **5. Visualización**

Se utilizaron herramientas como Power BI, Tableau o Metabase para generar dashboards interactivos. Algunas visualizaciones realizadas incluyen:  
- Gráfico de barras mensual.  
- Gráfico circular por tipo de vacuna.  
- Mapa de calor por establecimiento.

# **6. Conclusiones**

Se observaron patrones claros en los datos de vacunación, destacando algunos establecimientos como los más activos. Se identificaron picos mensuales que reflejan campañas de vacunación intensiva. Las vacunas más aplicadas fueron Pfizer y AstraZeneca. Este análisis podría complementarse con datos geográficos o poblacionales para mejorar la tasa de cobertura estimada. También se podrían desarrollar modelos predictivos para planificar futuras campañas de vacunación.

# **7. Anexos o Código**

En esta sección se pueden incluir enlaces al repositorio de código, scripts ETL, o acceso al dataset original utilizado.