

# Proyecto Databricks: Transporte Público en A Coruña

## Arquitectura Medallón (Bronze–Silver–Gold) con PySpark y Delta Lake

Este proyecto implementa un pipeline de datos para analizar el transporte público de A Coruña usando la arquitectura de medallón en Databricks.

### ■ Capa Bronze:

- Ingesta de datos crudos desde CSV (viajes y usuarios).
- Definición de esquemas con PySpark.
- Almacenamiento en Delta Lake.

### ■ Capa Silver:

- Conversión de horas a timestamp.
- Cálculo de duración de viajes.
- Normalización de transbordo (Sí/No -> 1/0).
- Join con usuarios para enriquecer la información.
- Persistencia en Delta.

### ■ Capa Gold:

- Modelo dimensional (Hecho\_Viajes + Dim\_Usuarios, Dim\_Lineas, Dim\_Paradas, Dim\_Tiempo).
- Creación de vistas de métricas clave encima del modelo dimensional:
  - vw\_total\_viajes\_por\_linea
  - vw\_viajes\_por\_tipo\_abono
  - vw\_top\_paradas\_bajada
  - vw\_ranking\_usuarios
  - vw\_hora\_pico

### Beneficio:

- Garantiza datos consistentes y listos para BI (Power BI, Tableau, Streamlit).
- Combina flexibilidad (modelo dimensional) y usabilidad (vistas con métricas).

### Recomendación para GitHub:

- Subir los archivos CSV de entrada.
- Crear carpetas para Bronze, Silver y Gold (con código PySpark).
- Guardar el README con el diagrama de arquitectura y ejemplos de vistas.
- Incluir capturas de notebooks de Databricks mostrando resultados.