

---

## 1. Introducción

Un buen transporte público es vital para la vida diaria de una ciudad: mueve a la gente, evita atascos y reduce el impacto ambiental. A medida que las ciudades crecen y los desplazamientos se complican, se vuelve imprescindible conocer a fondo cómo funciona este servicio. Para lograrlo, necesitamos datos abiertos, completos y bien ordenados que permitan medir su desempeño de manera clara y comprobable.

En este contexto, el *General Transit Feed Specification* (GTFS) se ha consolidado como el estándar de facto para publicar información sobre rutas, paradas, horarios y agencias. El formato distribuye la información en múltiples archivos de texto con una estructura relacional que facilita su procesamiento mediante herramientas modernas de análisis de datos y la generación de visualizaciones que revelan patrones operativos.

En el formato GTFS la información de transporte se divide en varios archivos, y para convertirla en datos útiles como rutas, horarios o paradas es necesario reunirla y limpiarla. Como advierte Wu *et al.* [Wu et al., 2023], la utilización de estos *feeds* supone un reto tanto para planificadores de transporte como para investigadores, debido a la complejidad de su comprensión, procesamiento y aprovechamiento. Además, la falta de visualizaciones accesibles —señalada por Para *et al.* [Para et al., 2024]— dificulta descubrir patrones clave, como las zonas con mayor densidad de paradas o las rutas más utilizadas, desaprovechando así gran parte del valor de la información de transporte público. No obstante, experiencias como *G2Viz* [Para et al., 2024] y el estudio de Calgary de Prommaharaj *et al.* [Prommaharaj et al., 2020] demuestran que, una vez integrados y depurados, los *feeds* GTFS permiten extraer evidencia objetiva para la toma de decisiones.

Motivada por estos hallazgos, trabajamos con el archivo de datos GTFS que publica la *San Francisco Municipal Transportation Agency* (SFMTA). Primero revisamos qué tablas incluye y cómo se relacionan. Luego limpiamos los datos, corrigiendo o quitando valores faltantes y reuniendo la información que estaba dispersa. Con el conjunto ya ordenado, identificamos las rutas y los tipos de transporte, medimos con qué frecuencia circulan y en qué zonas operan. Finalmente, convertimos estos resultados en gráficas fáciles de entender. Este camino —revisar, limpiar y visualizar— ofrece una imagen clara de la organización y el funcionamiento del sistema de transporte de San Francisco.

## 2. Objeto de Estudio

El objeto de estudio de este trabajo es el **sistema de transporte público operado por la SFMTA en San Francisco**, usando los datos disponibles en formato GTFS. Para el análisis se utiliza el archivo `muni_gtfs-current.zip`, descargado desde el sitio oficial de la agencia.

Este archivo contiene información organizada en tablas sobre:

- Las rutas disponibles (tipos: bus, tranvía, cable car), con sus códigos, colores y descripciones.
- Las paradas del sistema, con coordenadas geográficas, nombres y códigos.
- Los viajes planificados por día, incluyendo su asociación a rutas, calendarios y direcciones.
- Los horarios exactos de paso por cada parada (llegada y salida).
- Las líneas formadas por puntos geográficos que indican cómo se mueve cada ruta sobre el mapa.

El análisis busca detectar qué tipos de transporte son más usados, qué zonas tienen más cobertura, cómo varía el servicio entre semana y fines de semana, y cuáles rutas son más largas o más frecuentes.

## Referencias

- [Para et al., 2024] Para, S., Wirotsasithon, T., Jundee, T., Demissie, M. G., Sekimoto, Y., Biljecki, F., and Phithakkitnukoon, S. (2024). G2viz: An online tool for visualizing and analyzing a public transit system from gtfs data. *Public Transport*, 16(3):893–928.
- [Prommaharaj et al., 2020] Prommaharaj, S., Kattan, L., Shah, S., and Farooq, B. (2020). Visualizing public transit system operation with gtfs data: A case study of calgary, canada. *Heliyon*, 6(5):e04036.
- [Wu et al., 2023] Wu, J., Du, B., Gong, Z., Wu, Q., Shen, J., Zhou, L., and Cai, C. (2023). A gtfs data acquisition and processing framework and its application to train delay prediction. *International Journal of Transportation Science and Technology*, 12(1):201–216.