
1. Introducción

El transporte público es una parte clave del funcionamiento de cualquier ciudad moderna. Es fundamental para que las personas se desplacen, para reducir el tráfico vehicular y para minimizar el impacto ambiental. A medida que crecen las ciudades y las necesidades de movilidad, también crece la importancia de analizar cómo funciona el sistema de transporte. Para ello, contar con datos abiertos y estructurados es esencial, ya que permiten entender su desempeño de forma objetiva.

En ese sentido, el formato GTFS (General Transit Feed Specification) se ha convertido en un estándar muy utilizado. Este formato organiza la información del transporte público en múltiples archivos de texto, donde se detalla la operación de rutas, horarios, paradas y agencias. Gracias a su estructura relacional, estos datos pueden ser analizados con herramientas de programación, como Python, y visualizados para identificar patrones importantes sobre cómo funciona el sistema.

Sin embargo, a pesar de que muchas agencias publican datos GTFS, su análisis profundo no es tan común. Generalmente, estos datos se usan para planificar recorridos o mostrar horarios en apps, pero tienen un gran potencial para el análisis exploratorio de datos. Algunos estudios como [Para et al., 2024] y [Prommaharaj et al., 2020] han mostrado cómo usar estos datos para generar visualizaciones que revelan patrones operativos, diferencias entre zonas o días, y otros aspectos útiles para la toma de decisiones.

Este proyecto parte del análisis detallado del dataset GTFS de la SFMTA (San Francisco Municipal Transportation Agency). Primero se hizo una limpieza de los datos, identificando estructuras clave, valores faltantes y relaciones entre archivos. Luego, se generaron visualizaciones a partir de variables como rutas, paradas, horarios y frecuencias. Gracias a este proceso se pudo entender mejor cómo están organizados los datos, y se formularon preguntas clave para orientar el análisis final.

El objetivo es aplicar un enfoque claro para entender cómo está organizado y cómo funciona el sistema de transporte de San Francisco, utilizando únicamente datos GTFS estáticos. Además, se espera que este enfoque pueda adaptarse fácilmente a otras ciudades o proyectos similares.

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., Wirotasithon, 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.