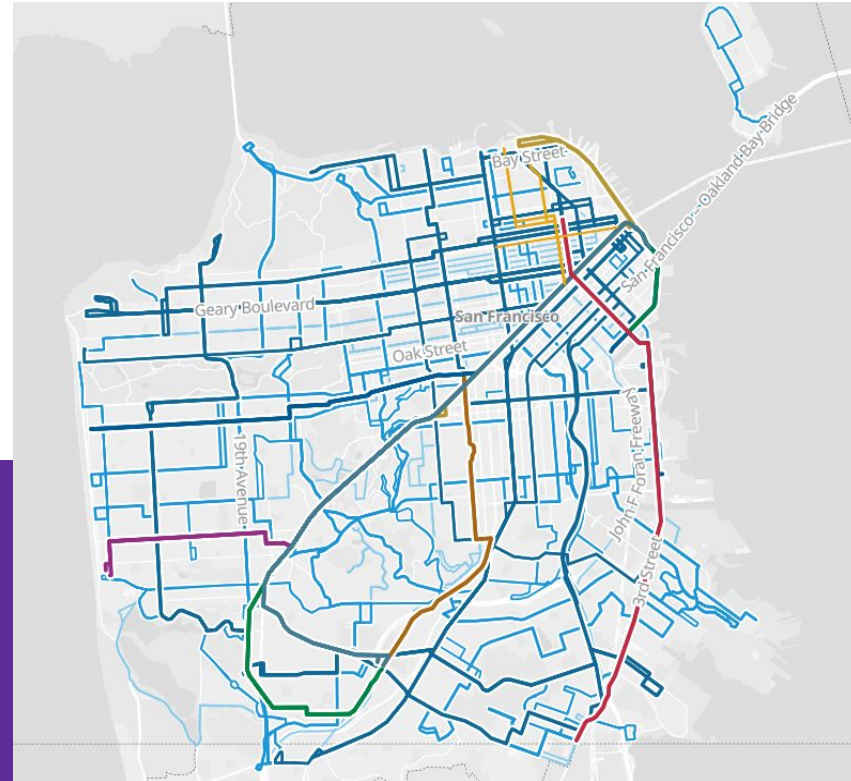


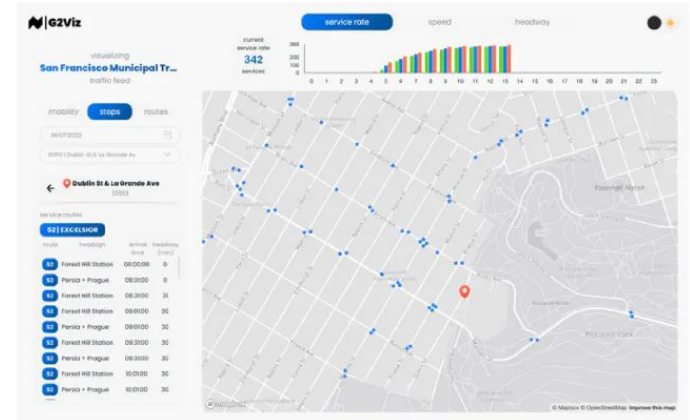
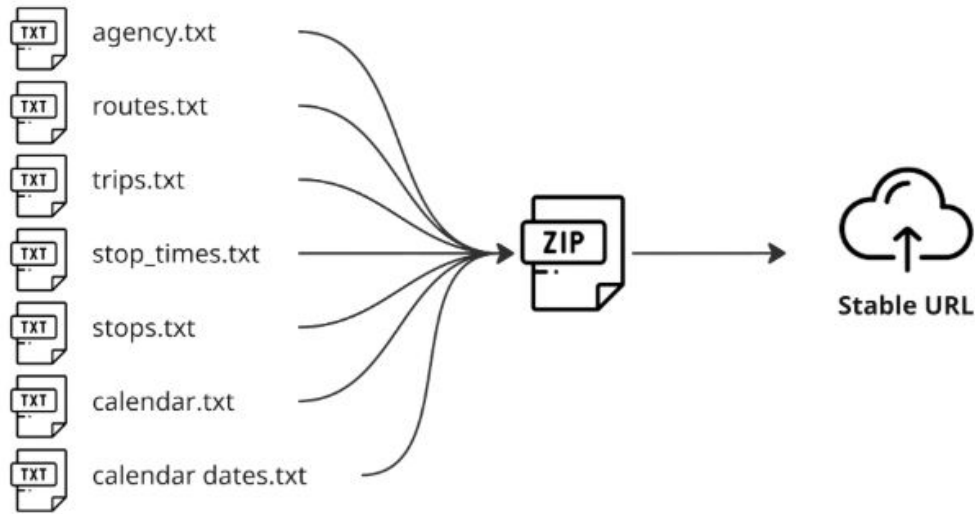
Pipeline

Análisis y Visualización del Sistema de Transporte Público de San Francisco (SFMTA) a partir de Datos GTFS Estáticos



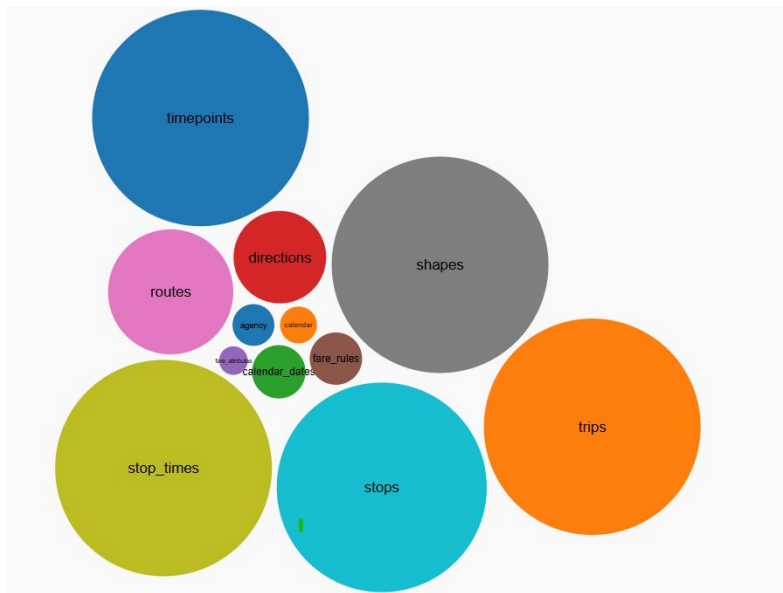
<https://www.transit.land/operators/o-9q8y-sfmta>

Descripción del Dataset



- El transporte público es clave para ciudades sostenibles.
- El formato **GTFS** permite estandarizar rutas, horarios y paradas.
- Herramientas como **G2Viz** muestran el potencial de visualizar estos datos.
- San Francisco (SFMTA) publica GTFS: ¡una gran oportunidad para explorar





Nombre	Tipo	Tamaño comprimido
agency	Documento de texto	1
calendar	Documento de texto	1
calendar_dates	Documento de texto	1
directions	Documento de texto	1
fare_attributes	Documento de texto	1
fare_rules	Documento de texto	1
routes	Documento de texto	2
SFMTA_Transit_Data_License_Agree...	Documento de texto	3
shapes	Documento de texto	470
stop_times	Documento de texto	7,081
stops	Documento de texto	65
timepoints	Documento de texto	627
trips	Documento de texto	134

Problema

Los datos de transporte en formato GTFS están repartidos en muchos archivos distintos; para convertirlos en información útil (rutas, horarios, paradas) primero hay que combinarlos y limpiarlos, un proceso que requiere conocimientos técnicos y herramientas especializadas (Wu et al., 2023). Además, como señalan Para et al. (2024), la falta de visualizaciones accesibles hace difícil descubrir patrones clave—por ejemplo, dónde hay más paradas o qué rutas se usan más—y, en consecuencia, se desaprovecha gran parte del valor de la información del transporte público.

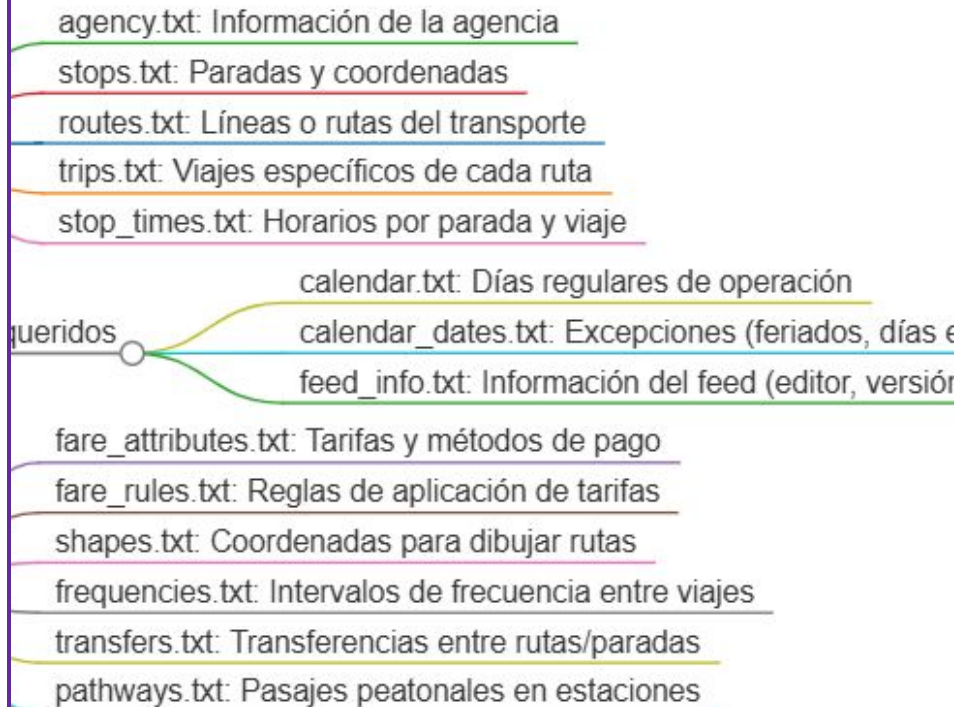
OBJETIVOS

Objetivo General

Analizar y visualizar el sistema de transporte público de San Francisco (SFMTA) a partir de datos GTFS estáticos para comprender mejor su estructura y funcionamiento.

Objetivos Específicos

- Explorar la estructura del dataset GTFS.
- Limpiar y preparar los datos.
- Identificar rutas, tipos de transporte y su actividad.
- Analizar frecuencias y cobertura.
- Crear visualizaciones interpretables.



Descripción del Dataset

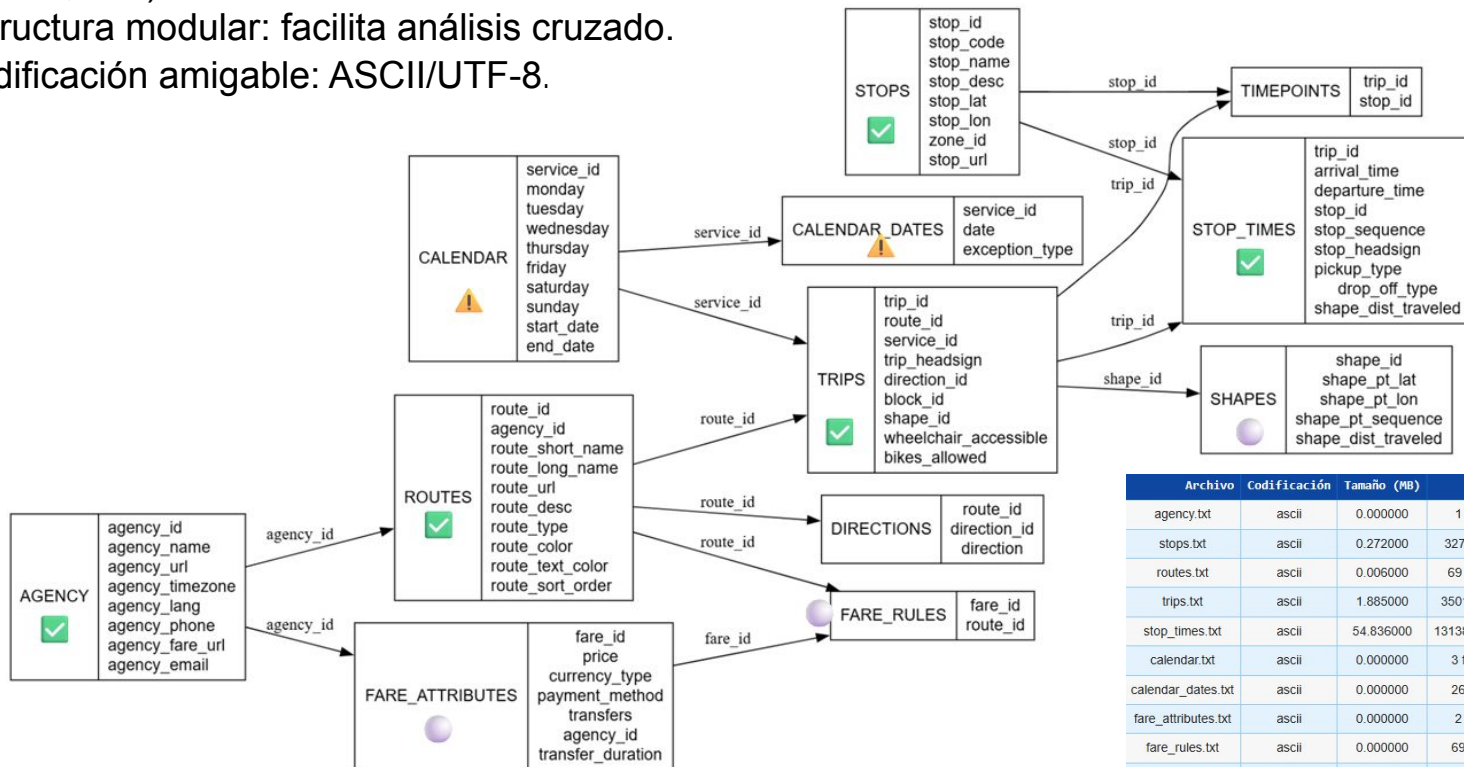
Dataset: muni_gtfs-current.zip

Contiene 12 archivos principales (routes.txt, trips.txt, etc.).

Estructura modular: facilita análisis cruzado.

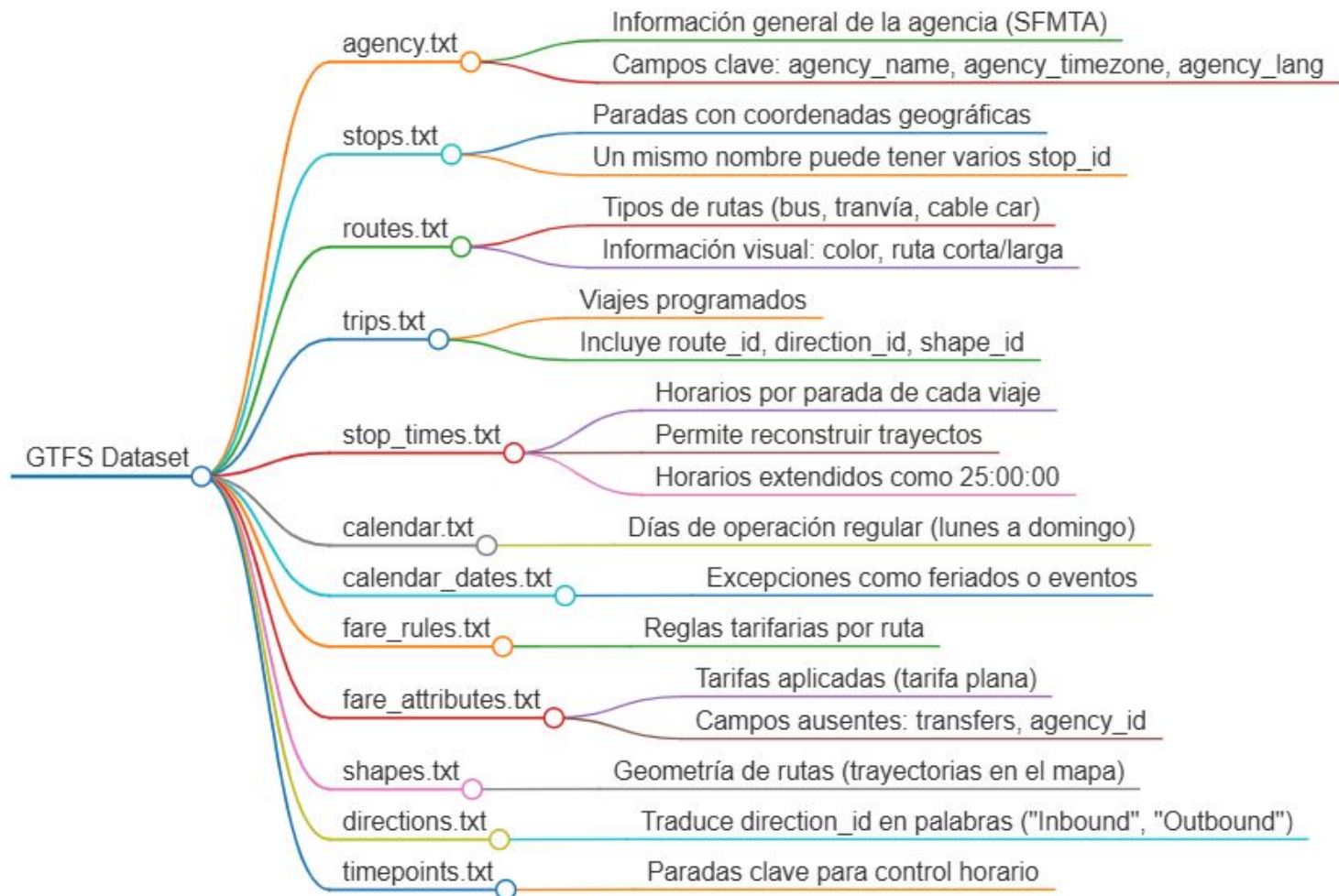
Codificación amigable: ASCII/UTF-8.

Modelo relacional de archivos de texto incluidos en un feed GTFS.



Archivo	Codificación	Tamaño (MB)	Granularidad
agency.txt	ascii	0.000000	1 filas x 8 columnas
stops.txt	ascii	0.272000	3278 filas x 8 columnas
routes.txt	ascii	0.006000	69 filas x 10 columnas
trips.txt	ascii	1.885000	35017 filas x 9 columnas
stop_times.txt	ascii	54.836000	1313881 filas x 9 columnas
calendar.txt	ascii	0.000000	3 filas x 10 columnas
calendar_dates.txt	ascii	0.000000	26 filas x 3 columnas
fare_attributes.txt	ascii	0.000000	2 filas x 7 columnas
fare_rules.txt	ascii	0.000000	69 filas x 2 columnas
shapes.txt	ascii	1.798000	46249 filas x 5 columnas
directions.txt	ascii	0.002000	138 filas x 3 columnas
timepoints.txt	ascii	4.527000	249904 filas x 2 columnas

Exploración Archivos



Calidad de Datos ,Limpieza y Transformación

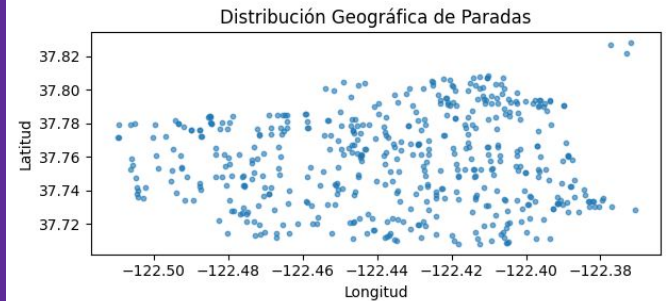
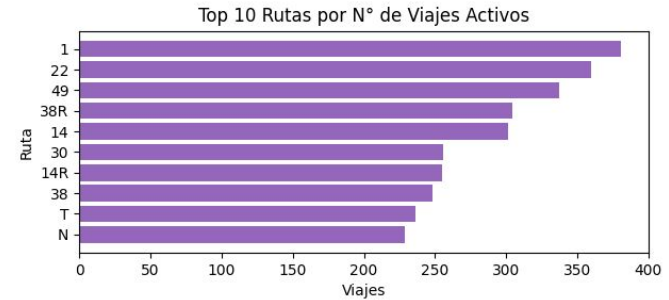
Tabla	Estado de Calidad	Observación
agency	Completa (sin nulos)	Datos listos para análisis
stops	Todas filas con nulos, requiere limpieza	Clave para ubicación, limpieza crítica
routes	Mayoría con nulos, revisar columnas	Nulos en algunas filas, evaluar columnas
trips	Completa (sin nulos)	Datos completos, buen nivel de calidad
stop_times	Todas filas con nulos, requiere limpieza	Tabla muy grande, limpieza imprescindible
calendar	Completa (sin nulos)	Datos completos, pocas filas
calendar_dates	Completa (sin nulos)	Datos completos, pocas filas
fare_attributes	Mayoría con nulos, revisar columnas	Pocas filas completas, evaluar columnas
fare_rules	Completa (sin nulos)	Datos completos, buen nivel de calidad
shapes	Completa (sin nulos)	Datos completos, buen nivel de calidad
directions	Completa (sin nulos)	Datos completos, buen nivel de calidad
timepoints	Completa (sin nulos)	Datos completos, buen nivel de calidad

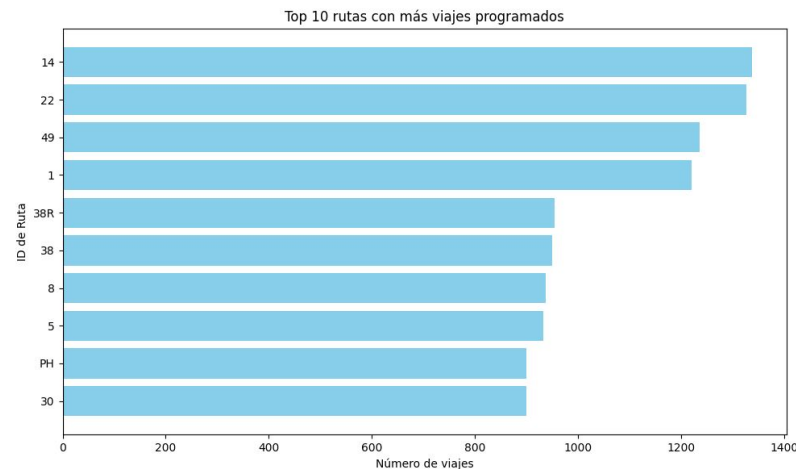
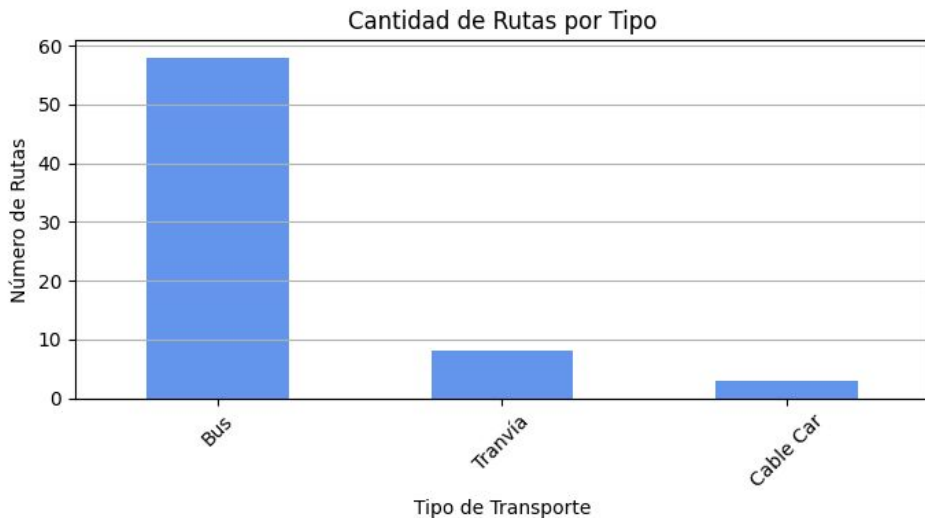
Archivo	Columnas afectadas	Descripción del problema
stops.txt	stop_desc, zone_id	Muchas filas sin descripción o sin zona tarifaria asignada.
stop_times.txt	stop_headsign	Este campo suele estar vacío si no hay información adicional.
routes.txt	route_sort_order	Campo vacío que podría usarse para ordenar rutas en visualizaciones.
fare_attributes.txt	transfers, agency_id	Faltan valores clave para modelar transferencias o no aplican al caso de una sola agencia.

	Tabla	Filas antes	Columnas antes	Filas después	Columnas después	Columnas eliminadas
0	stops	3278	8	3278	6	zone_id, stop_url
1	trips	35017	9	35017	5	wheelchair_accessible, bikes_allowed, block_id, direction_id

Visualización

Transporte Público





Pregunta

Objetivo específico

Archivo(s) GTFS

Propósito analítico

¿Cuántas rutas existen y de qué tipo son (bus, tranvía, cable car)?

Identificar la cantidad de rutas, los tipos de transporte y cuándo están activos

`routes.txt`

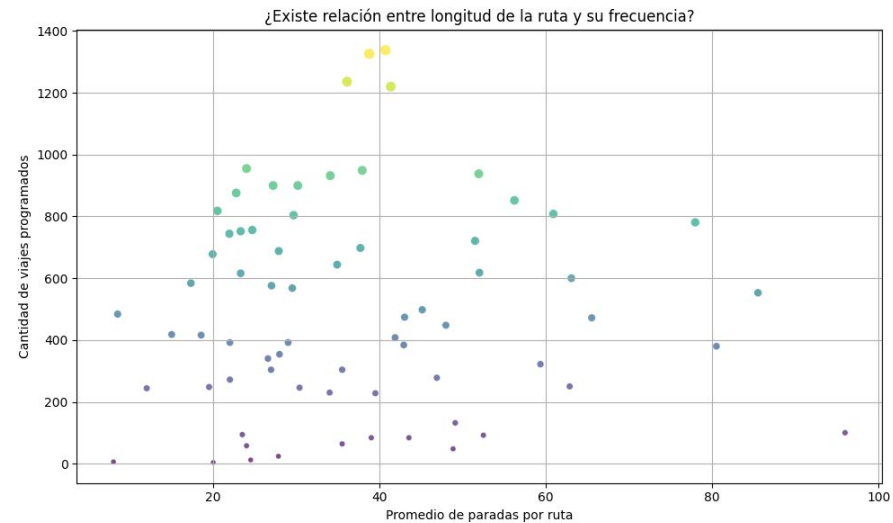
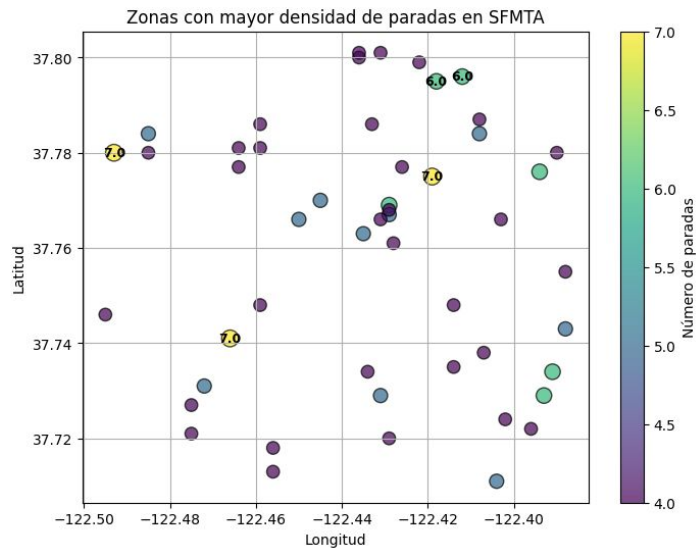
Conocer la variedad de medios operados por SFMTA y su distribución en el sistema

¿Cuáles son las rutas con más viajes programados por día?

Analizar la frecuencia de los servicios

`trips.txt`

Detectar rutas de alta frecuencia; refleja demanda o prioridad operativa



Pregunta

¿Qué zonas tienen más cobertura o mayor densidad de paradas?

¿Las rutas más largas (con más paradas) son también las más frecuentes?

Objetivo específico

Analizar la cobertura geográfica del sistema

Analizar la frecuencia de los servicios y su estructura

Archivo(s) GTFS

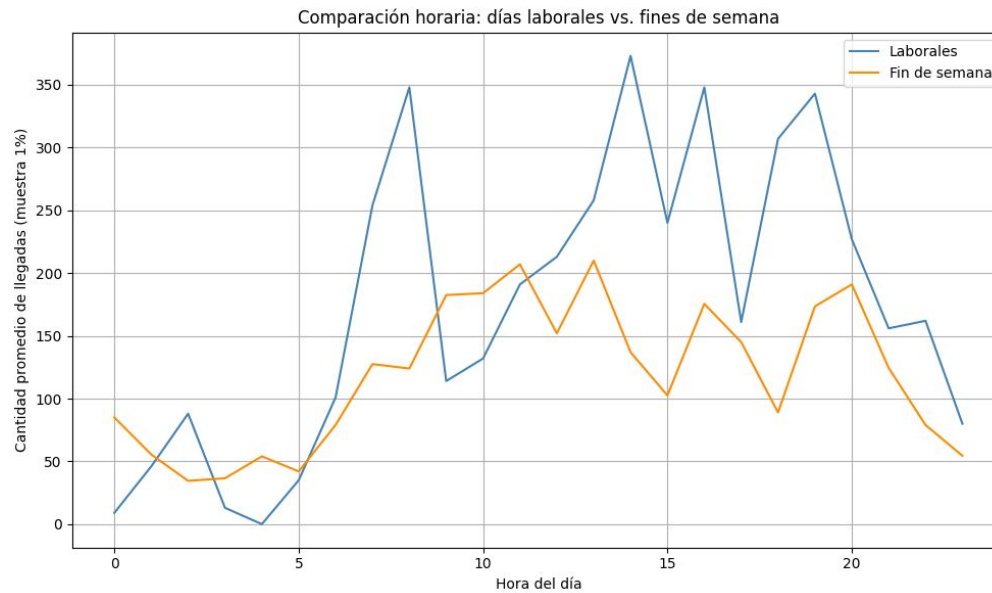
`stops.txt`

`stop_times.txt`,
`trips.txt`

Propósito analítico

Evaluar la accesibilidad y distribución espacial de paradas

Contrastar longitud y frecuencia de rutas para entender decisiones operativas



Pregunta

Objetivo específico

Archivo(s) GTFS

Propósito analítico

¿Qué días y franjas horarias tienen más servicio?

Crear visualizaciones para interpretar patrones de uso del sistema

`calendar.txt`,
`trips.txt`,
`stop_times.txt`

Descubrir diferencias entre días y horarios; identificar horas pico

REFERENCIAS

- Para, S., Wirotasithon, T., Jundee, T., Demissie, M. G., Sekimoto, Y., Biljecki, F. & Phithakkitnukoon, S. (2024). *G2Viz: An online tool for visualizing and analyzing a public transit system from GTFS data*. **Public Transport**, **16** (3), 893–928.
- Wu, J., Du, B., Gong, Z., Wu, Q., Shen, J., Zhou, L. & 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.
- Referencia - General Transit Feed Specification. (s. f.). <https://gtfs.org/es/documentation/schedule/reference/>