

Dashboard

Filtros Generales

Paradas:

Todas

Ruta:

Todas

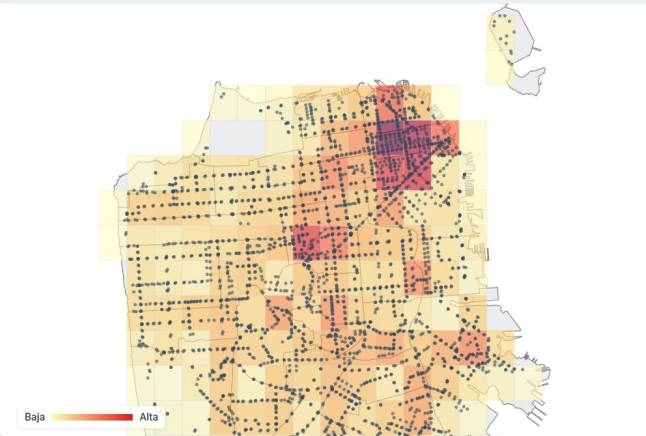
Top 10

Ver Top 10 Rutas

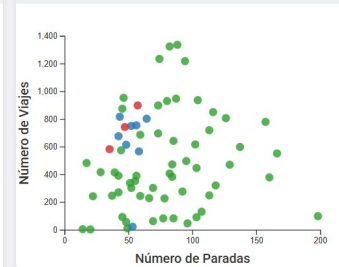
Ver Top 10 Paradas

Hipótesis 3: Mapa de Calor de Densidad de Paradas

☒ Mostrar Densidad de Paradas



Hipótesis 1: Paradas vs. Viajes

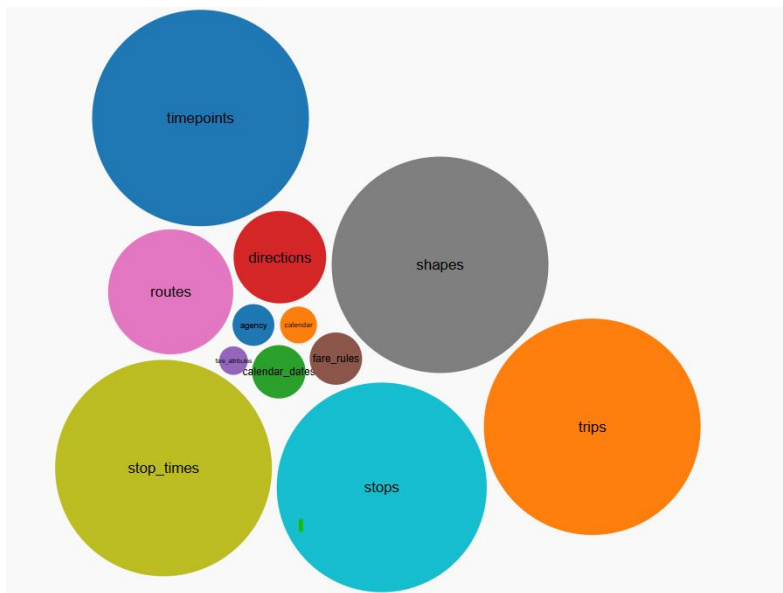


Hipótesis 2: Llegadas por Hora (Laborables)



<https://kemely2018.github.io/DASHBOARD/>

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



Usuarios

Kemely

Descargas

muni_gtfs-current

Ordenar

Ver

Extraer todo

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.

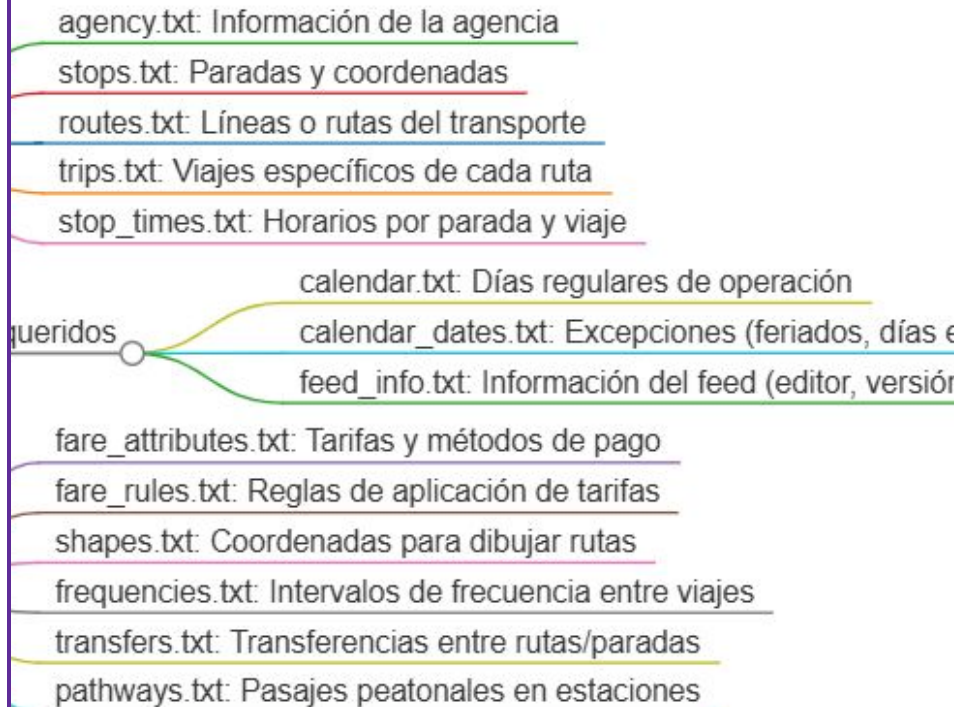
OBJETIVOS

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.



HIPÓTESIS 1

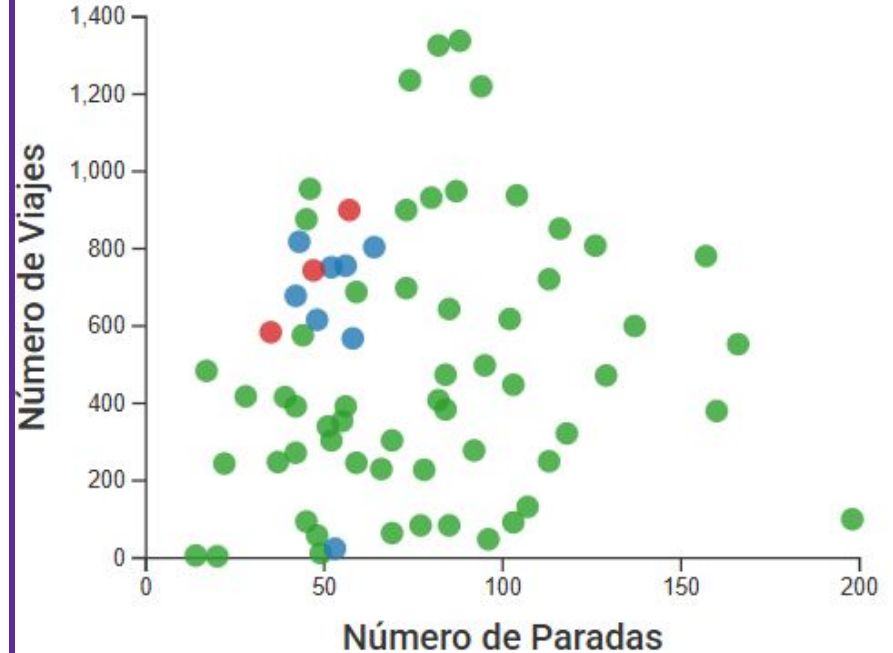
¿Las rutas con mayor número de paradas operan con menos viajes diarios?

Las rutas con más paradas suelen tener menos viajes al día.

Las rutas cortas (pocas paradas) reciben más autobuses/tranvías y salen con mayor frecuencia.

Esto indica que SFMTA da prioridad a las líneas más cortas y concurridas para enviar más vehículos

Hipótesis 1: Paradas vs. Viajes



HIPÓTESIS 2

¿La frecuencia de servicio varía significativamente entre horas punta de días laborables y fines de semana?

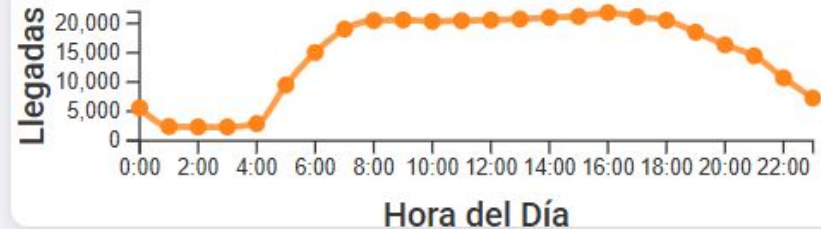
En días de semana, hay un pico muy marcado de llegadas entre las 7 – 9 AM y otro entre las 4 – 6 PM.

Fuera de esas horas, la cantidad de llegadas bajan notablemente.

En fines de semana, la curva es más “plana” (menor diferencia entre horas), sin picos tan pronunciados.

Esto muestra que SFMTA ajusta más buses/tranvías a las horas de mayor demanda entre semana, y controla mejor la oferta en fines de semana.

Hipótesis 2: Llegadas por Hora (Laborables)



HIPÓTESIS 3

¿La densidad de paradas es mayor en el centro de San Francisco y disminuye hacia la periferia?

En el centro de San Francisco (zona de Market St y alrededores) hay muchas paradas muy juntas (alta densidad).

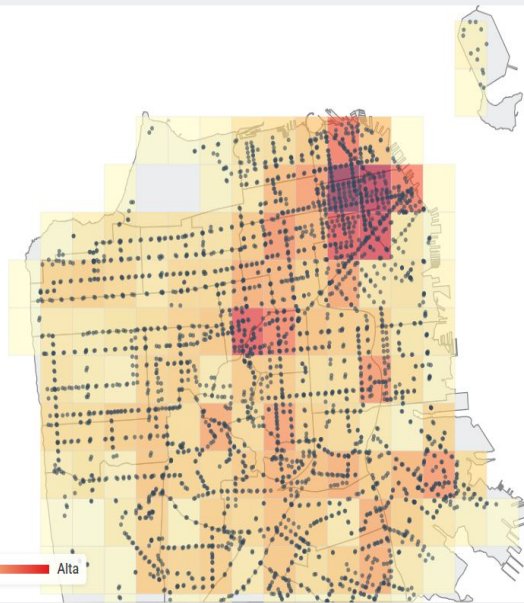
A medida que nos alejamos al oeste, sur o barrios periféricos, las paradas quedan más separadas y escasas.

Algunas zonas exteriores casi no tienen paradas cercanas (por ejemplo, Twin Peaks, Visitacion Valley).

Esto revela brechas de cobertura: áreas que podrían beneficiarse de nuevas paradas o rutas adicionales.

Hipótesis 3: Mapa de Calor de Densidad de Paradas

☒ Mostrar Densidad de Paradas



Objetivo Específico	¿Cómo se satisface?
1. Explorar la estructura del dataset GTFS	Al “desglosar” archivos GTFS (<code>routes.txt</code> , <code>stop_times.txt</code> , <code>stops.txt</code> , <code>calendar.txt</code>), se identifican las variables clave (paradas, viajes, horarios).
2. Limpiar y preparar los datos	En cada hipótesis se corrigen duplicados y valores faltantes (paradas sin coordenadas, trips caducados, rutas mal etiquetadas), asegurando calidad de datos.
3. Identificar rutas, tipos de transporte y actividad	Se extrae <code>route_type</code> , <code>route_id</code> , <code>num_stops</code> y <code>num_trips</code> (Hipótesis 1), y se asigna actividad horaria por <code>trip_id</code> (Hipótesis 2), y paradas por zona (Hipótesis 3).
4. Analizar frecuencias y cobertura	Frecuencia = viajes diarios (Hipótesis 1) y llegadas por hora (Hipótesis 2); Cobertura = densidad de paradas por bin geográfico (Hipótesis 3).
5. Crear visualizaciones interpretables	Scatterplot (Hipótesis 1), Line Chart con hover-line (Hipótesis 2), Heatmap + puntos sobre mapa (Hipótesis 3), con leyendas, tooltips y brushing interactivo.

CONCLUSIONES

- Validar Hipótesis 1 nos permite entender la priorización de rutas y asignación de vehículos.
- Los resultados de Hipótesis 2 muestran picos horarios en horas punta de laborables vs. fines de semana, orientando a optimizar flota según demanda.
- La verificación de Hipótesis 3 revela zonas en la periferia con baja densidad de paradas, sugiriendo áreas de mejora en cobertura geográfica.
- En conjunto, este análisis apoya la toma de decisiones para planificar rutas, ajustar frecuencias y reducir brechas de servicio en SFMTA.