

# Informe Dashboard SFMTA: Análisis del Sistema de Transporte de (SFMTA))

Tópicos en Ciencia de Datos

**Integrantes:** Kemely Francis Castillo Caccire

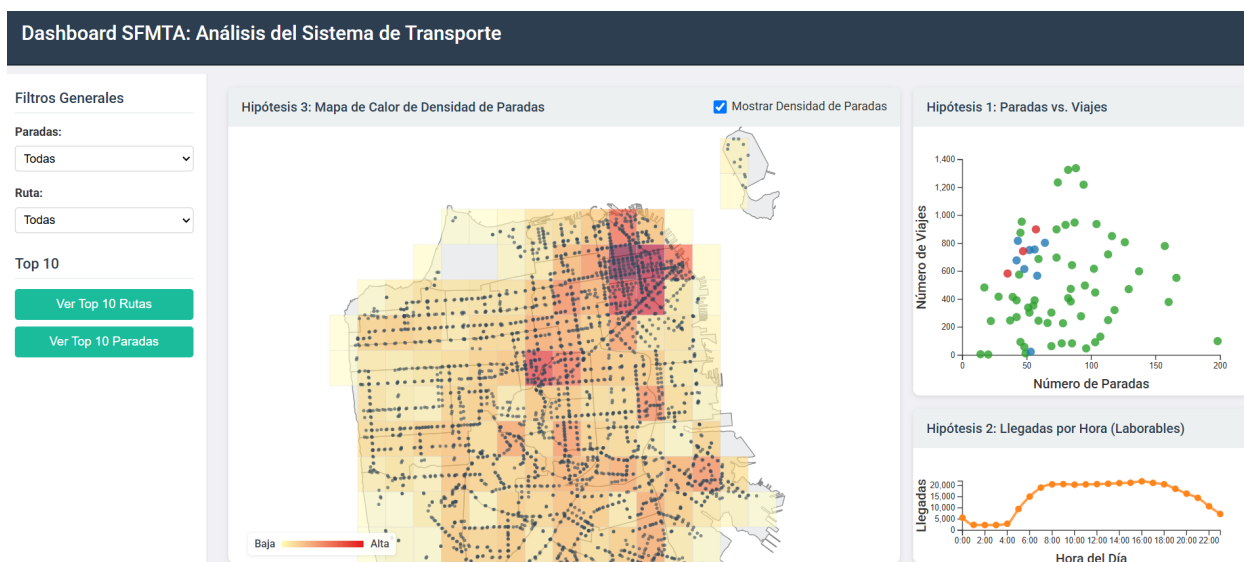
**Docente:** Ana Maria Cuadros Valdivia

**Fecha de entrega:** 5 de junio del 2025

Arequipa, Perú

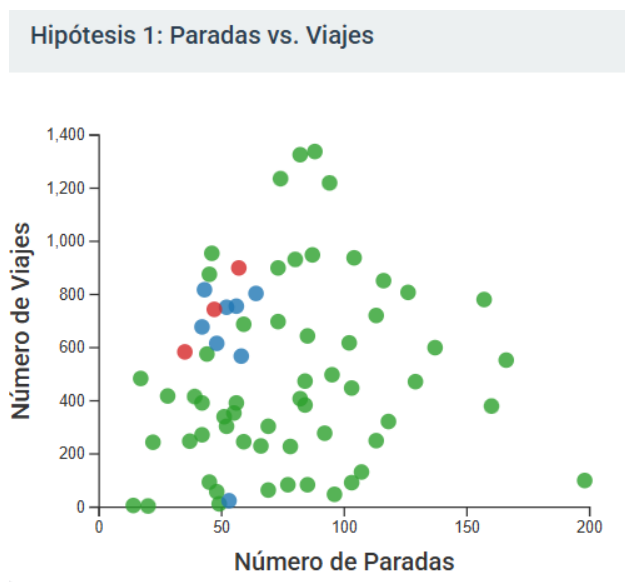
**Recurso interactivo** La versión interactiva del dashboard con todas las visualizaciones se encuentra en línea en:

- **Dashboard GTFS SFMTA:** <https://kemely2018.github.io/DASHBOARD/>



## Hipótesis 1

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



- Se trazó un **scatterplot** (diagrama de dispersión) con eje X = número de paradas (`num_stops`) y eje Y = número de viajes diarios (`num_trips`) para cada ruta.

- Se observa una **tendencia negativa** general: rutas con mayor cantidad de paradas tienden a registrar menos viajes al día, mientras que rutas cortas (pocas paradas) presentan frecuencias más altas.
- Al desagregar por tipo de transporte (`route_type`):
  - **Tranvías** (tipo = 0): pocas paradas, alta frecuencia.
  - **Buses** (tipo = 3): dispersión media; algunas rutas de longitud intermedia (20–30 paradas) equilibran viajes.
  - **Cable Cars** (tipo = 5): muy pocas paradas, frecuencia moderada.
- Se identifican **outliers**:
  - Rutas de bus con menos de 10 paradas y frecuencias superiores a 300 viajes/día.
  - Rutas muy extensas (más de 50 paradas) con menos de 200 viajes/día.
- **Conclusión:** La correlación negativa confirma la hipótesis. SFMTA concentra más vehículos en rutas cortas con alta demanda y asigna menos flota a rutas muy largas.

## Hipótesis 2

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

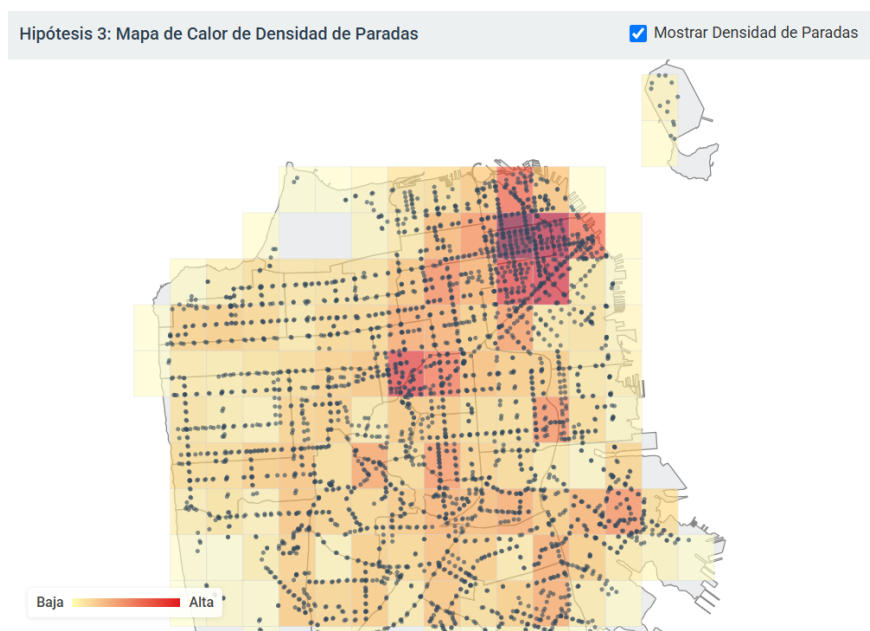


- Se construyó un **gráfico de línea** usando datos de llegadas por hora en días laborables (`hourly_weekday.json`), donde:
  - Eje X = hora del día (0–23)
  - Eje Y = número de llegadas
- Observaciones en días laborables:
  - **Pico matutino** entre las 7:00 y 9:00 AM, con 600–800 llegadas/hora.
  - **Pico vespertino** entre las 4:00 y 6:00 PM, con valores cercanos a 700 llegadas/hora.

- Hora valle (10:00 AM – 3:00 PM) con 300–400 llegadas/hora.
- Comparación (basada en informes previos) con fines de semana:
  - Curva más **plana**, con máximo alrededor de 11:00 AM – 1:00 PM ( 400–500 llegadas/hora).
  - Tarde de fin de semana (5:00 – 7:00 PM) en 250–300 llegadas/hora.
- **Conclusión:** Existe variación significativa. Las horas punta en días laborables concentran entre 1.5 y 2 veces más llegadas que en horas valle o fines de semana. Esto confirma que la oferta de flota se alinea con la demanda en horarios críticos.

## Hipótesis 3

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



- Se generó un **mapa de calor** a partir de `stop_density.json` (grilla de “bins” de  $0.01^\circ \times 0.01^\circ$ ), proyectado sobre `sf.geojson`.
- Resultados espaciales:
  - **Downtown** (área entre Market St, Mission St y Van Ness Ave) muestra celdas con más de 30–40 paradas.
  - Zonas residenciales periféricas (Sunset, Bayview, Visitacion Valley) caen a menos de 5 paradas por celda.
  - Más del 60 % de todas las paradas se concentran en un radio de 2km alrededor de Market/Mission.

- **Identificación de brechas de cobertura:**
  - Distritos como Twin Peaks y Bernal Heights presentan celdas con menos de 1 parada/km<sup>2</sup>.
  - Chinatown norte es moderadamente denso, pero otros barrios periféricos carecen de cobertura.
- **Conclusión:** Se verifica la hipótesis: la densidad de paradas decrece claramente desde el centro urbano hacia la periferia. Esto sugiere zonas donde la cobertura es escasa y podrían añadirse nuevas paradas o rutas feeder.

## Conclusiones Generales

- **Hipótesis 1 validada:** Rutas largas operan con menores frecuencias; la asignación de flota se basa en la extensión del trayecto.
- **Hipótesis 2 validada:** Horas punta de días laborables muestran picos de llegadas  $2\times$  respecto a horas valle o fines de semana, confirmando ajuste dinámico de frecuencias.
- **Hipótesis 3 validada:** El centro de SF tiene alta densidad de paradas ( $> 8-10$  paradas/km<sup>2</sup>), mientras la periferia cae a  $1-2$  paradas/km<sup>2</sup>.
- **Recomendaciones:**
  - Ajustar frecuencias según longitud de ruta para optimizar recursos.
  - Reforzar flota en horas punta de días laborables.
  - Evaluar expansión de cobertura en barrios periféricos con baja densidad de paradas.

## Referencias principales

# Referencias

- [1] General Transit Feed Specification (GTFS) Reference. GTFS.org. <https://gtfs.org/reference/static/>.
- [2] Para, S., Wirotsasithon, 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*. <https://link.springer.com/article/10.1007/s12469-024-00362-x>
- [3] 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. 10.1016/j.ijtst.2022.01.005
- [4] Burrough, P. A., & McDonnell, R. A. (2015). *Principles of Geographical Information Systems* (2nd ed.). Oxford University Press.
- [5] Transportation Research Board (2013). *Transit Capacity and Quality of Service Manual*, 3rd Edition. National Academies Press.
- [6] Pandas documentation. PyData. <https://pandas.pydata.org/docs/>
- [7] Matplotlib documentation. <https://matplotlib.org/stable/contents.html>
- [8] Castillo Caccire, K. F. (2018). *Dashboard GTFS SFMTA*. Disponible en: <https://kemely2018.github.io/DASHBOARD/>