



PONTIFICIA
UNIVERSIDAD
CATÓLICA
DE CHILE

El impacto de los eventos en el tiempo de viaje del transporte público de Santiago



Integrantes: Israel Aaron Pino Saavedra, Benjamín Alejandro Pardo Albornoz, Nicolas Ignacio Martinez Bravo, Zhongzhe Cheng

Curso: IMT-2200 Introducción a la Ciencia de Datos

Profesor: Rodrigo A. Carrasco

Fecha: 2025/09/11



Contexto y motivación:

El transporte público constituye un elemento esencial en la rutina diaria de millones de personas que habitan en las grandes ciudades. En un mundo cada vez más interconectado y acelerado, garantizar la predictibilidad y eficiencia de este sistema resulta clave para elevar la calidad de vida de la población.

Santiago de Chile, al igual que otras metrópolis en constante expansión, enfrenta importantes desafíos en materia de movilidad urbana. La congestión vehicular, los extensos tiempos de desplazamiento y las limitaciones en la eficiencia del transporte colectivo son problemas recurrentes que impactan directamente en el bienestar de sus ciudadanos. El Sistema de Transporte Metropolitano, cubre una superficie de alrededor de 680 km² del área del Gran Santiago, donde residen unos 7 millones de personas, a través de sus tres modos de transporte: Metro de Santiago, Tren Central Alameda-Nos y buses urbanos.

En este escenario surge una hipótesis central: cómo afectan distintas variables al tiempo de viaje y cuál es la capacidad de predicción a partir de la data recolectada. Así, la presente investigación busca examinar el tiempo de viaje, analizando el posible efecto que tendrían eventos climáticos y medio ambientales (lluvia, heladas, olas de calor e índice contaminación), eventos masivos planificados (deportivos, protestas, desfiles o musicales), fechas especiales, cortes de energía eléctrica, huelgas personal transporte, fallas técnicas, entre otros eventos.

A través de una metodología multidisciplinaria que contempla el análisis exploratorio de datos, la interpretación de resultados y la validación de la hipótesis, se pretende aportar evidencia sobre cómo estos distintos eventos inciden en la movilidad urbana y su capacidad de predicción para que sirva de insumo para usuarios, organismos estatales y empresas privadas relacionadas con la finalidad de tomar decisiones para aumentar predictibilidad de tiempos de viaje, disminuir tiempos a usuarios y costos del sistema.

Objetivos:

Los objetivos científicos del presente trabajo es, usando tiempos de viajes en el transporte público de Santiago, inferir que eventos son los de mayor impacto y que tan predecibles son. Para ello trabajaremos con la meta de contestar ¿Cuáles son y qué predictibilidad tienen los eventos con mayor impacto en el tiempo de traslado en el transporte público de Santiago?

Analizaremos eventos como algunas condiciones meteorológicas, eventos masivos planificados, fechas especiales y otros, afectan la operación y demanda del sistema de transporte público tomando y analizando tanto cantidades de pasajeros transportados diarios como tiempos de viaje.



Tendremos los objetivos específicos de:

1. Cuantificar el efecto de variables meteorológicas (lluvia, temperatura, contaminación) en el tiempo promedio de viaje de buses y metro.
2. Evaluar el impacto de eventos masivos (protestas, conciertos, partidos de fútbol) en la variación de los tiempos de viaje y en el flujo de pasajeros.
3. Comparar cómo cambia el uso de buses y metro en fechas especiales (feriados, huelgas, cortes de luz).

Se realizará este trabajo para que sirva como insumo por un lado, para poder informar con tiempo cambios en tiempos de traslado a usuarios y por otro lado para que las autoridades a cargo tomen las medidas preventivas para asegurar la eficiencia y correcto funcionamiento de este servicio. Con lo anterior se podrá beneficiar a usuarios con mayor información y menor tiempo de traslado, y al estado y empresas privadas con menores costos debido a planificación.

¿Cuál es la audiencia objetivo de su análisis?

Para determinar la audiencia objetivo debemos hacer una distinción:

- Audiencia objetivo entendida como aquellas personas o grupos a quienes va dirigido el proyecto, técnicamente.
- Audiencia objetivo entendida como personas, grupos o instituciones que finalmente se verán beneficiadas.

Respecto a la primera, el proyecto va dirigido al Ministerio de Transporte, pues éste es el órgano principal encargado de supervigilar el transporte metropolitano.

También va dirigido a las empresas de transporte público entre ellas las empresas de buses que conforman la **Red Metropolitana de Movilidad** compuesta por nueve empresas concesionarias:

- BUSES VULE
- SUBUS CHILE
- VOY SANTIAGO SPA
- METBUS
- REDBUS URBANO
- SANTIAGO TRANSPORTE URBANO
- METROPOL
- GRAN AMERICAS
- CONECTA



Por último agregar al metro de santiago y a empresas privadas/concesionadas que estén directamente relacionadas a este transporte.

Respecto a la segunda, los **potenciales beneficiados** son todos los usuarios del metro de santiago y/o transantiago, los cuales son millones de personas.

Datos:

Se trabajará principalmente con datos estructurados y semi estructuras.

Algunos de los datos considerados en forma preliminar a utilizar son:

Variable	Tipo	Formato	Volumen (Filas Columnas o Espacio de almacenamiento)	Fuente pública, privada u otro método	Origen
Fecha del viaje	Date time (Estructurado)	Fecha (csv)	GTFS_20250830.zip (7.959 KB)	Pública	GTFS (Agosto)
Hora inicial viaje	Date time (Estructurado)	Hora (csv)	GTFS_20250830.zip (7.959 KB)	Pública	
Hora final viaje	Date time (Estructurado)	Hora (csv)	GTFS_20250830.zip (7.959 KB)	Pública	
Tiempo de traslado	Date time (Estructurado)	Hora (csv)	GTFS_20250830.zip (7.959 KB)	Pública	
Cantidad de lluvia día	Float (Estructurado / Semi-Estructurado)	JSON	365 ítem	Pública	Open-Meteo API
Temperatura mínima día	Float (Estructurado / Semi-Estructurado)	JSON	365 ítem	Pública	
Temperatura máxima día	Float (Estructurado / Semi-Estructurado)	JSON	365 ítem	Pública	
Material particulado (MP2.5)	Float (Estructurado / Semi-Estructurado)	JSON	365*24 ítem (está por hora)	Pública	
Número pasajeros día sistema	Numérico (Estructurado)	Protobuf (convertible a JSON o CSV para análisis.)	entre 200 KB y 2 MB por actualización.	Se debe solicitar acceso	Posible GTFS - RT (solicitud enviada)
Número pasajeros día buses	Numérico (Estructurado)	Protobuf (convertible a JSON o CSV para análisis.)	entre 200 KB y 2 MB por actualización.	Se debe solicitar acceso	



Número pasajeros día metro	Numérico (Estructurado)	Protobuf (convertible a JSON o CSV para análisis.)	entre 200 KB y 2 MB por actualización.	Se debe solicitar acceso	
Fecha especial	Categorico (Semi-Estructurado)	JSON o diccionario	Bajo volumen: 100–1000 filas/año, solo KBs o unos MBs	Web-Scraping	Páginas web (wikipedia, T13, CHV, etc.)
Evento masivo	Categorico (Semi-Estructurado)	JSON o diccionario	Bajo volumen: 100–1000 filas/año, solo KBs o unos MBs	Web-Scraping	
Corte de luz masivo en Santiago	Categorico (Semi-Estructurado)	JSON o diccionario	Bajo volumen: 100–1000 filas/año, solo KBs o unos MBs	Web-Scraping	

Preguntas de investigación:

Algunos ejemplos de preguntas de investigación son:

1. ¿Qué variables meteorológicas afectan más la operación (lluvia, temperatura)? ¿En qué porcentaje desvían el tiempo promedio de los viajes?
2. ¿Qué variación de densidad de uso hay entre métodos de transporte (metro - bus) según la fecha?
3. ¿Qué eventos sociales afectan más al funcionamiento del transporte?
4. ¿Con qué eventos cambia el uso de transporte de buses a metro?
5. ¿Qué tan predecible son las fallas en el metro? ¿Se puede saber cuántas habrá al año?
6. ¿Se puede ajustar la cantidad de buses/viajes metro según fecha o pronóstico climático?

Diseño tentativo: Métodos computacionales y Estadísticos

1. Uso de Python 3 con distintas librerías que facilitan extracción, transformación y carga de datos (ETL).
2. Extracción de información web por URL, API y Scraping. Uso de librería **request** para extracción de datos ya sea open data o APIs.
3. Uso de **BeautifulSoup** para obtención de información publicada en páginas web sin descarga oficial.
4. Extracción de archivos mediante carga en **pandas** (read_csv, read_json)
5. Los datos serán almacenados en estructuras tabulares por **pandas DataFrames**
6. Limpieza y cruce de datos de transporte y clima: GTFS y .JSON. Identificación y eliminación de registros duplicados o inconsistentes.



7. Análisis exploratorio de Datos: Tiempos de trayecto, gráficos de dispersión, boxplots de atrasos vs clima, mapas de rutas. Medir el volumen de pasajeros durante el día. Detección de outliers y valores atípicos en viajes.
8. **Modelos tentativos:** Uso de **Matplotlib** y **Seaborn** para visualizaciones: mapas de Santiago con rutas afectadas, gráficos comparativos por clima. Heatmap por tiempo de viaje.
9. Se utilizarán estadísticos descriptivos como: media, mediana, moda, desviación estándar, rango y percentiles relevantes.
10. Coeficientes de correlación para medir fuerza y dirección de relación entre variables.
11. Métodos de modelización predictiva como modelos de regresión, se trabajará con el que mejor permita hacer predicciones en el área de estudio.
12. Página web donde se visualizará resumen de información del proyecto (Github Pages).

Conclusión:

Se trabajará en el proyecto recolectando, limpiando, procesando, analizando y modelando datos con el objetivo de dar respuesta a la pregunta que nos hemos planteado, ¿Cuáles son y qué predictibilidad tienen los eventos con mayor impacto en el tiempo de traslado en el transporte público de Santiago?. La información se visualizará en página web y en video de presentación del informe. Es un proceso no lineal en el cual se iteran algunos pasos debido al mejor entendimiento o posibles errores en el trabajo de los datos, lo anterior lo muestra la imagen 1 (obtenida de una presentación de cátedra).

Por los objetivos que nos han planteado de trabajo, no abarca la etapa de evaluación y validación. Tampoco habrá toma de decisión directamente relacionada con el trabajo realizado, pero podrá servir como insumo para la toma de decisiones y/o investigaciones futuras.

Imagen 1:

