



UNIVERSIDAD DE BUENOS AIRES
FACULTAD DE CIENCIAS EXACTAS Y NATURALES

Fronteras dentro de una ciudad: caso de estudio de escuelas en zonas cercanas al tren Sarmiento, CABA.

Tesis de Licenciatura en Ciencias de Datos

Manuel Graña

Directora: Inés Caridi

Codirectora: Paula Mascías

Buenos Aires, 2025

FRONTERAS DENTRO DE UNA CIUDAD: CASO DE ESTUDIO DE ESCUELAS EN ZONAS CERCANAS AL TREN SARMIENTO, CABA.

En las ciudades, ciertas infraestructuras funcionan como fronteras físicas o simbólicas que pueden limitar el acceso equitativo a oportunidades fundamentales, como la educación. En esta tesis se estudia el caso del ferrocarril Sarmiento como posible barrera urbana que afecta la movilidad escolar, entendida como los desplazamientos cotidianos que realizan los estudiantes entre sus hogares y sus establecimientos educativos. Se analiza si esta infraestructura, que atraviesa la ciudad en dirección este-oeste, condiciona la elección de escuelas por parte de los alumnos de nivel primario en la Ciudad Autónoma de Buenos Aires (CABA).

Para ello, se utilizaron datos abiertos de movilidad provenientes del sistema SUBE, que registra los viajes realizados en transporte público. A partir de esta información se reconstruyeron trayectorias escolares de alumnos del Área Metropolitana de Buenos Aires, y se diseñó una métrica denominada Barrier Score que permite comparar la cantidad de cruces observados de una frontera urbana con la cantidad esperada bajo un modelo nulo espacialmente restringido, que preserva la distribución de distancias.

El análisis propuesto busca aportar evidencia sobre los patrones de segregación educativa en CABA, así como herramientas metodológicas que permitan identificar barreras a la movilidad escolar. Los resultados pueden ser de utilidad para diseñar políticas públicas orientadas a mejorar la equidad en el acceso a la educación.

Palabras claves: Segregación urbana, movilidad, acceso a la educación, tren Sarmiento, datos SUBE, fronteras.

BOUNDARIES WITHIN A CITY: A CASE STUDY OF SCHOOLS NEAR THE SARMIENTO RAILWAY, BUENOS AIRES

In cities, certain infrastructures function as physical or symbolic boundaries that can limit equitable access to fundamental opportunities, such as education. This thesis examines the case of the Sarmiento railway as a potential urban barrier affecting school mobility, understood as the daily journeys students make between their homes and educational institutions. The study analyzes whether this infrastructure, which crosses the city in an east–west direction, influences the school choices of primary-level students in the Autonomous City of Buenos Aires (CABA).

To address this question, we used open mobility data from Argentina’s SUBE smart-card system, which records all public transport trips. Based on these data, we reconstructed student trajectories across the Metropolitan Area of Buenos Aires and applied a custom metric, the Barrier Score, which compares the observed number of crossings over a given infrastructure with the number expected under a spatially constrained null model that preserves trip distances.

This analysis aims to shed light on patterns of educational segregation in Buenos Aires and to develop methodological tools that help identify barriers to school mobility. The insights obtained can inform public policies that seek to improve equity in access to education and reduce spatial inequalities within cities.

Keywords: Urban segregation, mobility, educational access, Sarmiento railway, SUBE data, barriers.

AGRADECIMIENTOS

A mi mamá, mi primera y más importante maestra. Gracias no solo por orientar y guiar esta investigación, sino por inculcarme desde siempre el inmenso valor de la educación pública y la importancia de defenderla. Sos mi docente preferida de la vida, mi guía y mi ejemplo a seguir en todo.

A mi directora, Inés. Su generosidad, paciencia y sabiduría fueron fundamentales. Gracias por estar disponible en cada momento, por cada consejo, por abrirme puertas y conectarme con personas que enriquecieron mi perspectiva. Más allá de la tesis, gracias por inspirarme a seguir el camino de la ciencia y por todo lo que haces para que jóvenes investigadores como yo podamos dar nuestros primeros pasos.

A Zoe, que me amó siempre y me banco en cada momento que la carrera se hizo difícil. Gracias por acompañarme y estar siempre. Estar al lado tuyo hace que todo valga la pena.

A mi papá, por despertar mi curiosidad desde niño, acercándome a la belleza de la matemática, y por esa bendita insistencia para que aprendiera a programar. Esa visión tuya sembró la semilla de esta carrera que hoy, gracias a vos, termino con orgullo.

A mi hermano, con quien compartí incontables partidas de ajedrez que, aunque casi nunca gané, me enseñaron mucho más que estrategias. Gracias por el cariño de siempre, por las risas que alivianaron cualquier carga y por ser un compañero de vida excepcional.

A mis amigos del corazón y amantes de Shrek que son mi segunda familia. Con ellos crecí y compartí todos los momentos de mi vida, fueron y son la mejor compañía que puedo pedir. Gracias por cada momento y a cada uno por ser un pilar y en conjunto sostenerme para que la vida sea más fácil y linda.

A mis amigos y compañeros de facultad, con quienes compartí aulas, apuntes y ansiedades desde el primer día de Análisis I. Una mención especial para Bruno, mi compañero de estudio y también de vida. Esas jornadas infinitas de estudio no solo dieron sus frutos académicos, sino que forjaron una amistad que hizo todo este viaje posible y mucho más valioso. Llegar juntos a este momento vale muchísimo para mí.

A mi educación pública.

Índice general

1..	Introducción	1
1.1.	Contexto y motivación	1
1.2.	Marco teórico	2
1.2.1.	Segregación urbana	2
1.2.2.	Segregación educativa	3
1.2.3.	Fronteras urbanas	3
1.2.4.	Movilidad urbana y datos masivos	4
2..	Objetivos de la Investigación	5
3..	Datos de escuelas y análisis exploratorio preliminar	7
3.1.	Datos estructurales e informativos de las escuelas	7
3.2.	Indicadores de desempeño educativo y vulnerabilidad	10
3.3.	Análisis exploratorio mediante aprendizaje automático	12
3.3.1.	Importancia de variables con Random Forest	12
3.3.2.	Agrupamiento de escuelas con K-Means	14
3.4.	Conclusión parcial del análisis exploratorio sobre datos de escuelas primarias de CABA	15
4..	Datos de movilidad y reconstrucción de trayectorias	17
4.1.	Origen y características de los datos SUBE	17
4.2.	Construcción de pares casa-escuela	18
4.3.	Reconstrucción de la red de trayectorias y análisis de cruces	20
5..	Metodología del Barrier Score (BS)	23
5.1.	Definición del Barrier Score	23
5.2.	Modelo nulo de referencia	24
5.2.1.	Algoritmo del modelo nulo	26
5.3.	Barrier Score Direccional (BSD)	30
5.4.	Barrier Score por Distancia (B(d))	31
6..	Resultados	33
6.1.	Análisis de los cruces observados sobre el tren Sarmiento	33
6.2.	Comparación del efecto barrera: el tren Sarmiento frente a otras infraestructuras	34
6.3.	Análisis direccional del efecto barrera	36
6.4.	Análisis del efecto barrera por distancia de viaje	39
6.5.	Resultados preliminares: validación estadística del efecto barrera	41
6.6.	Síntesis de los hallazgos sobre el efecto barrera	43
7..	Conclusiones y Trabajo a Futuro	45
7.1.	Conclusiones principales	45
7.2.	Trabajo a futuro y perspectivas	46

1. INTRODUCCIÓN

1.1. Contexto y motivación

La Ciudad Autónoma de Buenos Aires (CABA) no es diferente a otras grandes ciudades del mundo: desde hace mucho tiempo, se observan en ella patrones de **segregación social y espacial** [1, 2, 3], lo que refiere a la concentración de las personas por alguna característica, como nacionalidad, grupo socioeconómico, cultura, religión, entre otras. Aunque las ciudades son lugares que pueden impulsar la diversidad y la innovación por la gran cantidad de gente que concentran [4, 5] esto puede ser afectado cuando el espacio se fragmenta.

Esta fragmentación, muchas veces causada por las mismas infraestructuras que construimos, puede terminar dividiendo barrios, aumentando las desigualdades y fomentando la segregación [6, 7, 8, 9, 10]. Como ya mostraban los estudios clásicos de sociología, cuando la gente vive separada según su clase social o etnia, se crean circuitos cerrados donde las oportunidades no se reparten por igual, algo que se nota mucho en la educación [11]. Investigaciones más recientes han agregado que a esto también influyen otros factores, como el tipo de vivienda disponible para cada grupo [12], la forma en que están conectadas las calles de la ciudad [13], o el "capital cultural" que cada persona trae desde su casa [14, 15].

En CABA, el análisis de datos abiertos evidencia la persistencia de la brecha norte-sur: las escuelas públicas del norte tienden a mostrar, mejores indicadores de desempeño que las del sur, lo que se refleja también en menores tasas de repitencia y sobreedad [16]. Esta marcada diferencia en los perfiles escolares plantea interrogantes sobre los patrones de asistencia: ¿En qué medida los estudiantes asisten a establecimientos dentro de sus propias zonas socioeconómicas, y cómo podría esto relacionarse con la segregación que otros estudios han señalado [17, 18]?

Además de dónde vive la gente, la ciudad también está atravesada por **barreras físicas**. Las autopistas y las vías de tren que vemos son un buen ejemplo: pueden ser verdaderos obstáculos que hacen más difícil que la gente se relacione o que llegue a lugares importantes como escuelas o centros de salud [19, 7, 20, 21, 22, 23]. Desde la década de 1960, distintos urbanistas ya advertían cómo las vías de alto tráfico podían reducir las oportunidades para crear y mantener lazos sociales entre barrios divididos, afectando la cohesión social [24]. Esta idea sigue siendo central hoy en día, cuando se busca una planificación urbana que promueva ciudades más inclusivas y resilientes [25, 26]. En nuestra ciudad, **el tren Sarmiento**, con su recorrido de este a oeste, es un claro ejemplo de cómo una infraestructura puede cortar la continuidad del tejido urbano, generando una división muy marcada entre el **norte** y el **sur** [27]. Aunque estas ideas sobre el impacto de las barreras son importantes en la teoría urbana, no siempre se ha contado con mediciones empíricas sólidas que muestren la relación entre infraestructuras de transporte y la conectividad social o el acceso a servicios. Si bien existen algunos estudios cualitativos o a pequeña escala [28, 29], la investigación cuantitativa, muchas veces limitada por la falta de datos georreferenciados sobre interacciones, se ha enfocado más en medir la segregación socioeconómica usando datos demográficos o, más recientemente, datos de movilidad [30, 31, 32]. Solo algunos trabajos han intentado vincular explícitamente la segregación con la

presencia de barreras urbanas [9, 33, 34]. Sin embargo, todavía no contamos con una evaluación empírica rigurosa sobre cómo una barrera física específica como el tren Sarmiento podría condicionar las trayectorias desde la casa de los y las estudiantes a su escuela en CABA (**movilidad escolar**), un vacío que este trabajo busca empezar a llenar.

La creciente disponibilidad de registros masivos de movilidad, como los datos de transacciones de la tarjeta SUBE un miércoles de 2019 [35, 36], nos brinda una oportunidad valiosa para investigar esta cuestión. De hecho, trabajos que han utilizado datos de telefonía móvil o tarjetas inteligentes han demostrado cómo el análisis de los flujos cotidianos de personas puede ayudar a comprender las desigualdades que estas experimentan en su día a día [37, 38]. En esta línea, metodologías que emplean *random walks* sobre redes de transporte multicapa han permitido estimar la probabilidad real de cruce entre zonas urbanas segmentadas [18].

Este trabajo se alinea con esfuerzos recientes por cuantificar el efecto de las barreras urbanas para separar grupos [19]. Para ello, nuestro enfoque combina tres elementos: el análisis espacial de los datos educativos, la reconstrucción detallada de las trayectorias casa-escuela a partir de datos SUBE, y la aplicación de un índice direccional que determina cuantitativamente el efecto de una barrera en la ciudad al que llamamos **Barrier Score**. El propósito es describir y cuantificar cómo la presencia del tren Sarmiento se relaciona con los patrones de viaje de los estudiantes de nivel primario, y analizar si funciona como barrera que genera que los alumnos no crucen el tren para ir a la escuela, contribuyendo a la segregación de estudiantes en CABA.

Confiamos en que estos hallazgos, además de contribuir al debate académico sobre la desigualdad urbana, puedan ofrecer insumos concretos para el diseño de políticas públicas que busquen mejorar la conectividad entre zonas separadas por barreras y promover una educación más equitativa.

1.2. Marco teórico

Este trabajo se apoya en cuatro ejes conceptuales, los cuáles brindan la teoría previa para poder abordar el objeto de estudio y comprender el problema:

1.2.1. Segregación urbana

Entendemos la **segregación urbana** como ese patrón persistente de separación espacial entre distintos grupos sociales dentro de una ciudad. Cuando estos grupos se distribuyen de manera desigual, terminan viviendo en entornos cotidianos muy diferentes. La formulación clásica de este concepto no solo considera dónde vive la gente, sino también cómo esa ubicación afecta su acceso a servicios, espacios públicos y redes sociales. En este sentido, la segregación es un proceso relacional: no se trata solo de contar cuántas personas de un tipo viven juntas, sino de entender cómo esa distribución territorial moldea sus oportunidades de vida [3]. En la práctica, esto puede generar un circuito que se retroalimenta, donde las ventajas (o desventajas) de la segregación urbana se acumulan y se transmiten entre generaciones, lo que le da un carácter "estructural".

En las últimas dos décadas, el avance de las herramientas computacionales nos ha permitido mirar la segregación con otros ojos. Se puede modelar a la ciudad como una red de nodos y aristas donde los nodos podrían representar objetos de estudio como viviendas de personas o escuelas y las aristas las calles que unen esos puntos de la ciudad. Entonces, la "distancia social" ya no depende solo de la cercanía física; lo que realmente importa es la

posibilidad efectiva de interacción entre diferentes territorios. Métodos basados en *random walks*, medidas de accesibilidad y simulaciones con agentes han mostrado que dos barrios pueden estar "lejos" socialmente aunque los separen pocos metros, si las conexiones viales o de transporte son deficientes, o si existe una fuerte percepción de peligro entre ellos. Estos enfoques han ayudado a identificar umbrales de mezcla –es decir, niveles mínimos de interrelación necesarios para romper con la segmentación– y a describir cómo pequeños cambios en la conectividad pueden generar transformaciones importantes en la estructura social de la ciudad [13, 38, 18].

1.2.2. Segregación educativa

La **segregación educativa** es la separación y aislamiento de los grupos sociales en sectores, áreas e instituciones del sistema. Esto genera que los alumnos que asisten a una misma escuela tiendan a provenir de entornos socioeconómicos muy similares. Como resultado, la población estudiantil dentro de cada establecimiento se vuelve muy homogénea, limitando la interacción y la mezcla entre chicos y chicas de diferentes orígenes sociales [39]. Lejos de ser un espacio neutral, la escuela tiende a reflejar y reproducir las jerarquías de la sociedad que la contiene. Bourdieu y Passeron, por ejemplo, argumentaron que el capital cultural –ese conjunto de disposiciones, lenguajes y saberes que se adquieren en el hogar– se alinea con los códigos que la institución educativa valora y utiliza para evaluar. Quienes llegan a la escuela con ese capital "incorporado" suelen tener una ventaja [14]. A esto hay que sumarle la capacidad económica de las familias para elegir ciertos establecimientos, costear apoyos extraescolares o activar sus redes de contactos. El resultado es la conformación de *circuitos diferenciados de escolaridad*: trayectorias educativas que se separan desde temprano y que tienden a mantenerse separadas a lo largo de la vida. Para el contexto latinoamericano, Tedesco ha mostrado cómo las crisis macroeconómicas pueden agudizar este fenómeno, impactando con mayor dureza en las escuelas y hogares con menos recursos [15].

En términos urbanos, la segregación educativa puede manifestarse espacialmente: las escuelas consideradas de mayor "prestigio" tienden a concentrarse en zonas de altos ingresos, mientras que aquellas con indicadores de desempeño más bajos suelen agruparse en barrios más vulnerables [40, 41]. La literatura sugiere que esto puede llevar a una especie de "doble cierre": se hipotetiza que los alumnos no solo tenderían a asistir a colegios cercanos a sus viviendas sino que, además, estos colegios podrían reforzar la homogeneidad social de su grupo de origen [42].

1.2.3. Fronteras urbanas

Las ciudades no solo se definen por la suma de sus barrios, sino también por las líneas que las dividen y conectan. Infraestructuras como autopistas, ríos canalizados o vías férreas introducen discontinuidades en el tejido urbano. Estas pueden funcionar como **barreras materiales**, difíciles de cruzar físicamente, pero también como **barreras simbólicas**, cargadas de significados asociados al peligro, la marginalidad o simplemente "el otro lado" [43].

Al limitar la interacción y el movimiento entre diferentes zonas, estas infraestructuras se convierten en mecanismos que pueden generar y perpetuar la segregación urbana, aislando a comunidades y reforzando las divisiones existentes [8]. En Buenos Aires, el tren Sarmiento es un caso paradigmático de esta dinámica: su traza longitudinal frag-

menta la ciudad, interrumpe calles y obliga a realizar rodeos que pueden encarecer los desplazamientos y desalentar el contacto entre quienes viven a orillas opuestas. De hecho, investigaciones recientes han documentado cómo infraestructuras lineales similares tienden a reducir la densidad de lazos sociales y contribuyen a perpetuar la segmentación residencial y educativa en diversas ciudades del mundo [19, 22].

1.2.4. Movilidad urbana y datos masivos

La irrupción del *big data* de transporte ha abierto nuevas posibilidades para analizar el acceso a las oportunidades urbanas. Fuentes de datos como los registros de tarjetas inteligentes (SUBE), la geolocalización de teléfonos móviles o los rastros GPS de vehículos públicos ofrecen una visión de la dinámica urbana con una escala y frecuencia antes impensables. A diferencia de los censos residenciales, que ofrecen una foto estática de dónde vive la gente, estos datos permiten realizar **estimaciones sobre la movilidad** de las personas: por dónde se mueven, con qué frecuencia y en qué horarios [44]. Esto, a su vez, posibilita aproximar la exposición real de distintos grupos a servicios como escuelas, empleos o centros de salud.

Sin embargo, es crucial subrayar que la reconstrucción de trayectorias a partir de estos datos no es una ciencia exacta y está llena de desafíos. La tarea de inferir un viaje completo, con su origen, destino y ruta, a partir de registros dispersos (como el pago en una validadora SUBE) es una tarea de modelado compleja que se basa en hipótesis y asunciones. La granularidad temporal y espacial de los datos, así como la falta de información sobre el propósito de cada viaje, imponen limitaciones significativas [45]. Por lo tanto, si bien estos datos nos permiten analizar patrones agregados de movilidad, no se debe asumir que se puede reconstruir de manera inequívoca el recorrido de una persona. Los resultados obtenidos siempre son estimaciones que dependen de la metodología y de la calidad de los datos de origen.

Adicionalmente, el uso de datos de movilidad individual conlleva una enorme responsabilidad ética en cuanto a la protección de la privacidad. Estos registros pueden revelar patrones de vida. Por ello, cualquier investigación en esta área debe realizarse sobre datos que hayan sido previamente anonimizados y agregados de forma robusta, siguiendo protocolos estrictos que impidan la re-identificación de las personas. La investigación nunca debe implicar el acceso o la filtración de información personal identificable. En este trabajo, como se detallará más adelante, se utilizan exclusivamente fuentes de datos públicas que ya han sido sometidas a procesos de anonimización y enmascaramiento, y se adhiere a las recomendaciones internacionales sobre el tratamiento seguro y ético de la información personal [46, 47].

A pesar de estas limitaciones y responsabilidades, el estudio de la movilidad a través de datos masivos se convierte en una especie de "laboratorio empírico" para investigar, a escala agregada, cómo se entrelazan las infraestructuras urbanas y las dinámicas socio-económicas.

2. OBJETIVOS DE LA INVESTIGACIÓN

El propósito fundamental de esta tesis es investigar la relación entre las barreras físicas urbanas y la segregación educativa en la Ciudad Autónoma de Buenos Aires (CABA). En particular, se busca evaluar de manera empírica si el tren Sarmiento, una de las infraestructuras más importantes que atraviesa la ciudad, actúa como una frontera que condiciona la movilidad escolar de los alumnos de nivel primario.

Para alcanzar este propósito general, la investigación se estructura en torno a los siguientes objetivos:

En primer lugar, se busca **reconstruir y caracterizar las trayectorias casa-escuela** de los estudiantes de nivel primario. Utilizando un extenso conjunto de datos de transporte público (tarjeta SUBE), se inferirán los desplazamientos cotidianos en transporte público, generando un mapa de la movilidad escolar en CABA. Este primer paso es esencial para poder analizar cómo los estudiantes interactúan con la geografía de la ciudad.

En segundo lugar, el trabajo se propone **medir cuantitativamente el impacto del tren Sarmiento como barrera**. A partir de las trayectorias reconstruidas, se analizará la propensión de los estudiantes a cruzar esta infraestructura para asistir a sus escuelas. El análisis no se limitará a contar cuántos cruzan, sino que desarrollará una métrica que compare los cruces observados con un escenario esperado si la barrera no existiera, permitiendo así aislar y cuantificar el "efecto barrera" de la vía.

Estos objetivos buscarán, en última instancia, dar respuesta a las siguientes **preguntas guía** que orientan la investigación:

- ¿Existe evidencia de segregación educativa en la Ciudad Autónoma de Buenos Aires, manifestada en los patrones de movilidad escolar en transporte público?
- ¿Actúa el tren Sarmiento como una frontera efectiva que limita el acceso a escuelas para los alumnos de nivel primario en CABA?

3. DATOS DE ESCUELAS Y ANÁLISIS EXPLORATORIO PRELIMINAR

Antes de abordar el análisis de la movilidad escolar y el efecto barrera del tren Sarmiento, se realizó un análisis exploratorio preliminar de los datos de las **escuelas públicas de nivel primario** en la Ciudad Autónoma de Buenos Aires (CABA). El objetivo principal de esta etapa fue caracterizar la oferta educativa y detectar posibles patrones espaciales. Un eje central de este análisis es la conocida brecha socioeconómica entre el norte y el sur de la ciudad y la diferencia entre el nivel académico que se mostrará posteriormente. Para facilitar la comparación a lo largo de este capítulo, se agruparon las comunas de CABA en dos grandes zonas: "Norte" y "Sur". Como se ilustra en la Figura 3.1, se consideraron como Comunas del Norte a la 12, 13, 14 y 15; y como Comunas del Sur a la 4, 7, 8 y 9. Esta división, aunque simplificada, permite una primera aproximación a las disparidades territoriales. Este análisis inicial sienta las bases para comprender el contexto en el que se insertan las trayectorias casa-escuela que se estudiarán posteriormente.

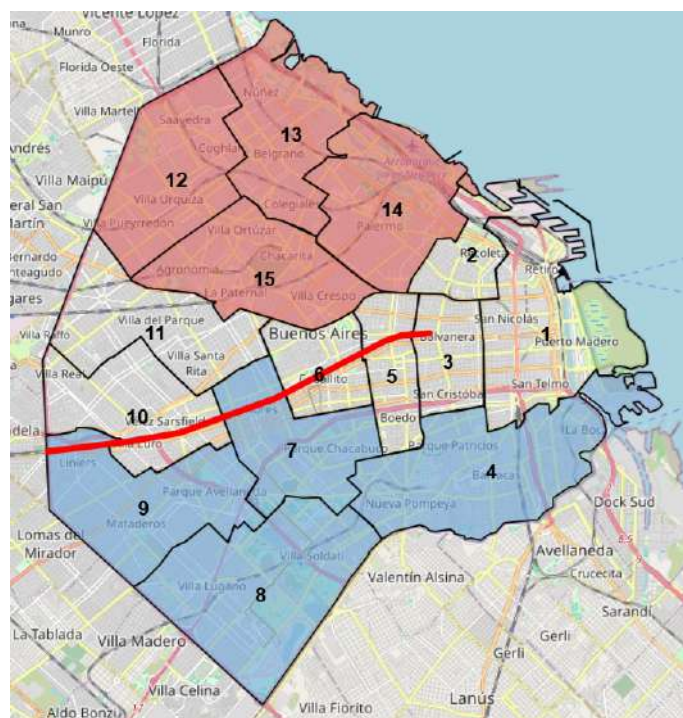


Fig. 3.1: Mapa de la Ciudad Autónoma de Buenos Aires con la división de comunas utilizada para el análisis exploratorio. En color rojo se destacan las Comunas del Norte (12, 13, 14, 15) y en color azul las Comunas del Sur (4, 7, 8, 9). El resto de las comunas se muestran en un color neutro. En rojo se marca la traza del tren Sarmiento

3.1. Datos estructurales e informativos de las escuelas

La información fundamental sobre los establecimientos educativos se obtuvo a partir de los datos abiertos provistos por el Gobierno de la Ciudad de Buenos Aires (GCBA).

Específicamente, se recurrió al conjunto de datos denominado *"Escuelas públicas y privadas de CABA"* [48], disponible en el portal BA Data. Para esta investigación, se filtraron los registros correspondientes a escuelas de gestión estatal y nivel primario. La extracción de los datos se realizó mediante técnicas de *web scraping* para asegurar la versión más actualizada disponible al momento del análisis (febrero de 2025).

Este conjunto de datos proporciona una rica caracterización de cada establecimiento, incluyendo:

- **Identificadores y contacto:** Nombre de la escuela, Clave Única de Establecimiento (CUE), domicilio, correo electrónico y teléfono.
- **Ubicación geográfica:** Coordenadas de latitud y longitud, fundamentales para el análisis espacial.
- **Características de infraestructura:** Información sobre elementos como la cantidad de bibliotecas, laboratorios, o la disponibilidad de acceso a internet en el establecimiento.
- **Matrícula:** Número total de alumnos inscritos en el nivel primario.

A modo ilustrativo, la Tabla 3.1 presenta las primeras cinco filas del conjunto de datos procesado, mostrando una selección de las variables más relevantes para este estudio.

Tab. 3.1: Primeras cinco filas del conjunto de datos de escuelas primarias estatales de CABA (selección de variables relevantes).

Nombre Escuela	Comuna	Dirección	Latitud	Longitud	Matrícula
Escuela Primaria N°1 DE 1	1	Av. de Mayo 500	-34,608	-58,373	350
Escuela Primaria N°2 DE 1	1	Suipacha 100	-34,603	-58,378	280
Escuela Primaria N°1 DE 14	14	Av. Cabildo 2000	-34,560	-58,455	420
Escuela Primaria N°5 DE 4	4	Av. Patricios 1050	-34,639	-58,369	510
Escuela Primaria N°7 DE 9	9	Av. Directorio 4500	-34,641	-58,485	300

Nota: Los datos presentados son ficticios y sirven únicamente como ejemplo de la estructura de la tabla. Las variables de infraestructura no se muestran pero fueron analizadas.

Un análisis preliminar de las variables de infraestructura (como la presencia de bibliotecas, laboratorios o acceso a internet) no reveló diferencias sistemáticas o patrones geográficos marcados entre las escuelas de distintas zonas de la ciudad. Si bien existen variaciones individuales entre establecimientos, no se identificó una tendencia clara que sugiera una disparidad significativa en la dotación de estos recursos estructurales básicos entre el norte y el sur de CABA a nivel agregado para las escuelas primarias estatales.

En contraste, una de las diferencias más notorias y espacialmente correlacionadas se observa en la **matrícula** total por establecimiento. Al agrupar la matrícula total por comuna, como se visualiza en la Figura 3.2, se evidencia que las escuelas ubicadas en las comunas del sur de la ciudad tienden a tener, en promedio, una matrícula mayor que aquellas localizadas en las comunas del norte. Esto sucede específicamente con la comuna 8 (comuna más al sur de CABA) donde se observa una diferencia bastante mayor, sobretodo con la comuna 15 (comuna más al norte de CABA) que presenta la menor matrícula promedio de todas las comunas.

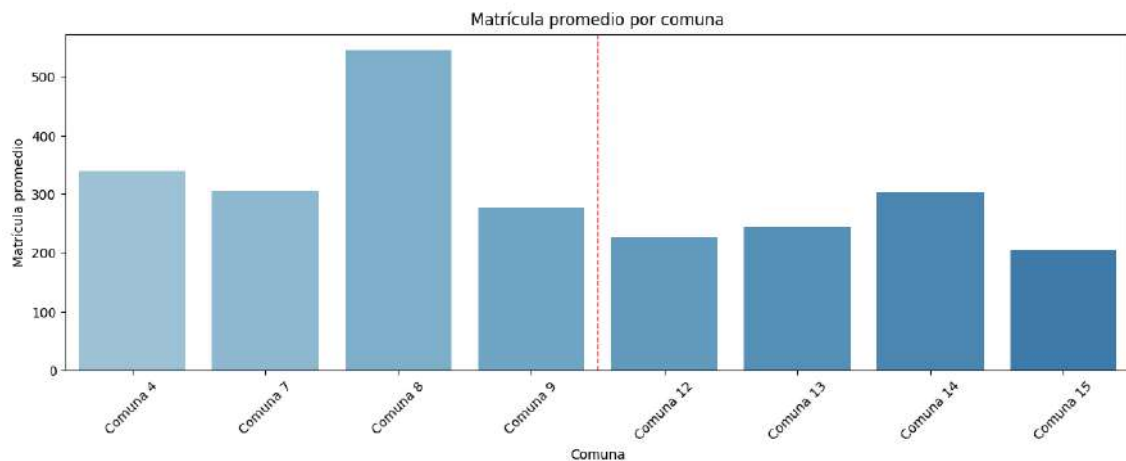


Fig. 3.2: Distribución de la matrícula total de alumnos en escuelas primarias estatales por Comuna en CABA. A la izquierda las comunas del sur y a la derecha las del norte.

Para resaltar esta diferencia, la Figura 3.3 compara la matrícula total acumulada en las escuelas primarias estatales de las Comunas del Norte (12, 13, 14, 15) con la de las Comunas del Sur (4, 7, 8, 9).

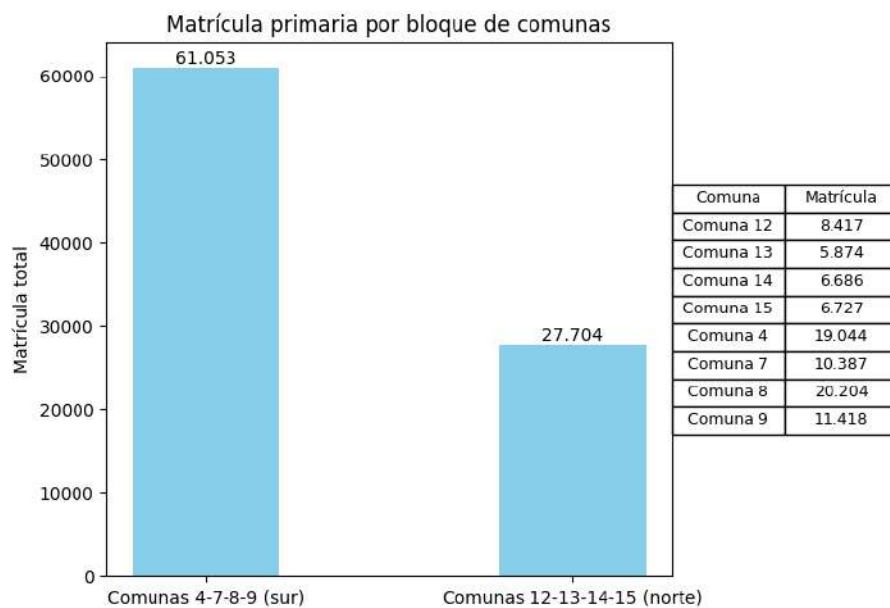


Fig. 3.3: Comparación de la matrícula total acumulada en escuelas primarias estatales entre las Comunas del Sur (4, 7, 8, 9) y las Comunas del Norte (12, 13, 14, 15)

Se observa que el número total de alumnos inscritos en escuelas primarias estatales es mayor en el conjunto de comunas del Sur que en las del Norte, esto sucede principalmente por la comuna 8 que tiene una gran cantidad de alumnos inscritos en sus escuelas en comparación a la comuna 15 que tiene menos de la mitad. Este hallazgo resulta particularmente interesante si se considera que, según datos poblacionales generales como los del "Censo Nacional de Población", la densidad poblacional o el número total de habitantes

no es necesariamente mayor en estas comunas del sur en la misma proporción. Una posible hipótesis para explicar esta mayor concentración de matrícula estatal en el sur, que requeriría un análisis más profundo y datos adicionales, podría estar relacionada con una menor oferta o un menor acceso a escuelas de gestión privada en las comunas del sur, lo que derivaría en una mayor presión sobre el sistema estatal en dichas zonas.

Si bien las características estructurales básicas parecen ser relativamente homogéneas, la diferencia en la matrícula entre el norte y el sur invita a explorar si existen también disparidades en otros aspectos, como los indicadores de desempeño educativo, lo cual se abordará en la siguiente sección.

3.2. Indicadores de desempeño educativo y vulnerabilidad

Más allá de las características estructurales, resulta fundamental analizar los indicadores de desempeño educativo y vulnerabilidad para comprender las posibles disparidades en la calidad y las condiciones de la oferta escolar en CABA. Para ello, se recurrió a la plataforma de indicadores educativos de *Argentinos por la Educación* [16], que compila y presenta datos de diversas fuentes oficiales. Para este estudio, se utilizaron los datos correspondientes al año 2019 para el nivel primario, con el fin de mantener la consistencia temporal en el análisis.

Una dimensión clave del análisis son los resultados de las pruebas Aprender (anteriormente ONE - Operativo Nacional de Evaluación), evaluaciones estandarizadas que miden los logros de aprendizaje de todos los estudiantes de sexto grado (11-12 años) de nivel primario de la provincia y ciudad de Buenos Aires. Los datos obtenidos de [16] para las pruebas Aprender 2019, agregados por comuna, detallan el porcentaje de estudiantes que alcanzan los niveles de desempeño: *avanzado*, *satisfactorio*, *básico* y *por debajo del básico*, en las áreas de Lengua y Matemática (y otras áreas, aunque el enfoque aquí se centra en estas dos para ilustrar los patrones).

Un análisis consistente a través de las áreas evaluadas revela un patrón geográfico claro: las escuelas ubicadas en las comunas del sur de la ciudad tienden a presentar, en promedio, resultados de desempeño inferiores en comparación con aquellas situadas en las comunas del norte. A modo ilustrativo, la Figura 3.4 muestra dos indicadores contrastantes extraídos de [16] para el año 2019: el porcentaje de estudiantes con desempeño *por debajo del básico* en Matemática y el porcentaje de estudiantes con desempeño *avanzado* en Lengua, ambos graficados por comuna. En el primer caso (Matemática por debajo del básico), Figura 3.4A), se observa una mayor proporción de estudiantes en este nivel en las comunas del sur. Inversamente, en el segundo caso (Lengua avanzado, Figura 3.4B), las comunas del norte concentran un mayor porcentaje de alumnos con los más altos logros.

Adicionalmente a los resultados de las pruebas estandarizadas, se analizaron indicadores de trayectoria escolar que pueden reflejar situaciones de vulnerabilidad educativa, también obtenidos de [16] para el año 2019 y agregados por comuna:

- La **tasa de sobreedad** refiere al porcentaje de estudiantes cuya edad supera la teóricamente esperada para el grado que están cursando. Una alta tasa de sobreedad puede indicar dificultades en la progresión regular a través del sistema educativo.
- La **tasa de repitencia** indica el porcentaje de estudiantes que deben volver a cursar un mismo grado debido a no haber alcanzado los objetivos de aprendizaje esperados.

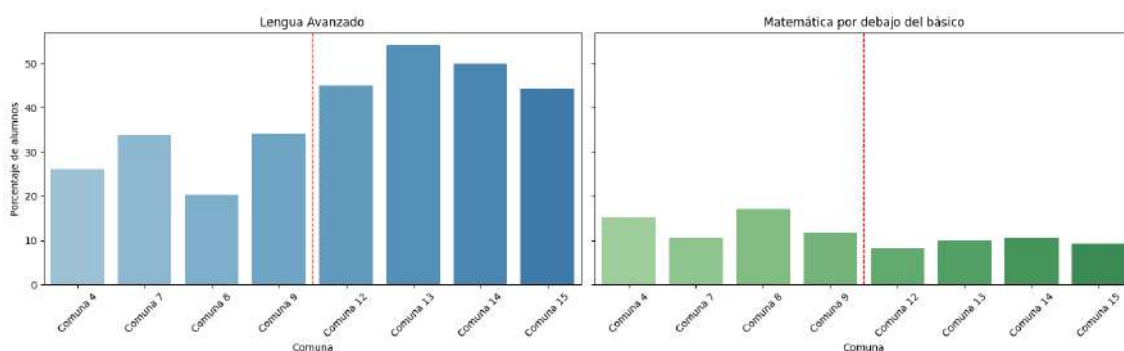


Fig. 3.4: Porcentaje de estudiantes con desempeño *avanzado* en lengua (izquierda) y porcentaje de estudiantes con desempeño *por debajo del básico* en Matemática (derecha). En cada gráfico se presentan a la izquierda las comunas del sur y a la derecha las del norte. (Datos 2019, Fuente: [16]).

El análisis de estos indicadores, presentados en la Figura 3.5, también muestra una marcada disparidad territorial. Tanto la **tasa de sobreedad** (Figura 3.5A) como la tasa de repitencia (Figura 3.5B) parecen ser mas elevadas en las comunas del sur de la ciudad en comparación con las del norte, a excepción de la comuna 15 que presenta mayor tasa que las otras comunas agrupadas como norte.

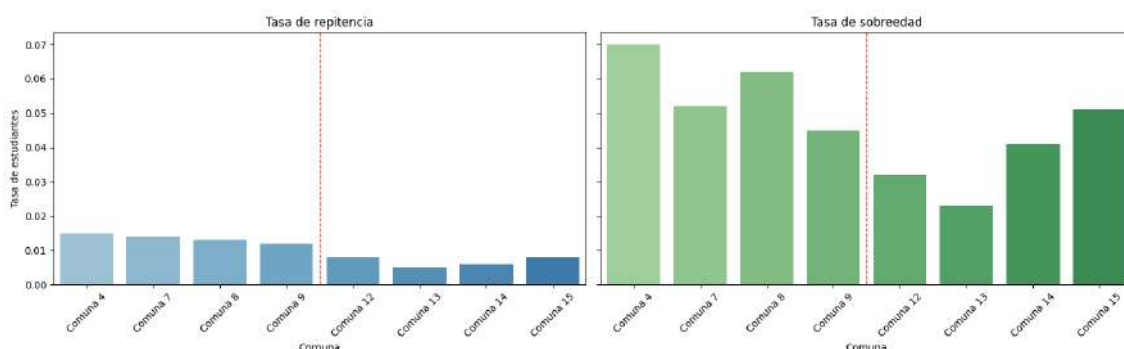


Fig. 3.5: Tasa de repitencia (izquierda) y sobreedad(derecha) en escuelas primarias estatales por comuna. En cada gráfico se presentan a la izquierda las comunas del sur y a la derecha las del norte. (Datos 2019, Fuente: [16]).

En conjunto, tanto los resultados de las pruebas de aprendizaje como los indicadores de trayectoria escolar, todos correspondientes al año 2019 y obtenidos de la plataforma de *Argentinos por la Educación* [16], sugieren una posible segmentación norte-sur en términos de resultados y condiciones educativas en el nivel primario estatal de CABA. Las comunas del sur no solo presentan una mayor matrícula por establecimiento (como se vio en la Sección 3.1), sino también, en promedio, menores niveles de desempeño en pruebas estandarizadas y mayores tasas de sobreedad y repitencia.

3.3. Análisis exploratorio mediante aprendizaje automático

Para profundizar en la caracterización de las diferencias entre las escuelas y explorar patrones multidimensionales, se aplicaron dos técnicas de **aprendizaje automático** sobre el conjunto de datos de escuelas primarias estatales de CABA. El objetivo principal no fue construir modelos predictivos, sino utilizar estas herramientas como métodos exploratorios para identificar las variables más distintivas entre zonas geográficas y para descubrir agrupamientos naturales de escuelas basados en algunas de sus características.

3.3.1. Importancia de variables con Random Forest

Se entrenó un modelo de clasificación basado en **Random Forest** [49] con el propósito de predecir si una escuela se encuentra al norte o al sur del tren Sarmiento, utilizando como predictoras el conjunto completo de variables disponibles: matrícula, los distintos niveles de desempeño en las pruebas Aprender 2019 (Lengua, Matemática, Cs. Sociales, Cs. Naturales), las tasas de repitencia y sobreedad, y las variables estructurales (disponibilidad de bibliotecas, internet, etc.) y se consideró la ubicación de las escuelas en relación al tren Sarmiento, definiendo una variable categórica "Norte/Sur del tren" para cada establecimiento. Esto se realizó sobre el conjunto de datos de todas las escuelas primarias de CABA.

Random Forest es un algoritmo de aprendizaje automático supervisado que combina múltiples árboles de decisión para realizar tareas de clasificación o regresión. Un **árbol de decisión** es un modelo que segmenta el espacio de características en regiones mediante decisiones binarias secuenciales (por ejemplo, "¿La matrícula de la escuela es mayor a 400?"), el árbol divide esta pregunta en "Sí" o "No" y dependiendo de la escuela sigue una rama u otra. Luego sigue preguntando algo más generando una estructura jerárquica que permite llegar a una predicción en función de las variables de entrada. Cada decisión dentro del árbol busca generar grupos lo más homogéneos posible según la variable de interés, que en este caso es si la escuela se ubica al norte o al sur del tren.

En un Random Forest, cada árbol del conjunto se entrena sobre una muestra aleatoria de los datos de entrenamiento, seleccionada con reemplazo (técnica conocida como *bootstrap*), y utilizando un subconjunto aleatorio de variables en cada división del árbol. Esta estrategia introduce diversidad entre los árboles del conjunto.

La predicción final del Random Forest se obtiene mediante votación entre árboles: cada árbol emite una clase como predicción, y la clase más frecuente entre todos los árboles determina el resultado final del bosque.

El árbol de decisión representado en la Figura 3.6 es un ejemplo simplificado que ilustra cómo el modelo realiza una clasificación a partir de variables predictoras. A partir de preguntas jerárquicas sobre el porcentaje de estudiantes con nivel avanzado en Lengua, la tasa de repitencia y la cantidad de matrícula, el árbol conduce a una predicción sobre si una escuela se encuentra al norte o al sur del tren Sarmiento. En cada bifurcación se evalúa una condición, y las ramas marcan la dirección que toma el algoritmo según si se cumple ("Sí") o no ("No"). En este ejemplo, el 62 % de las escuelas quedan clasificadas como "Norte" y el otro 28 % como sur.

Es importante aclarar que este árbol no representa el modelo completo, sino una instancia individual dentro del conjunto generado por el algoritmo de *Random Forest* [49]. Cada árbol del bosque es distinto, ya que se entrena sobre una muestra aleatoria de los datos (bootstrap) y, en cada división del árbol, se selecciona un subconjunto aleatorio de

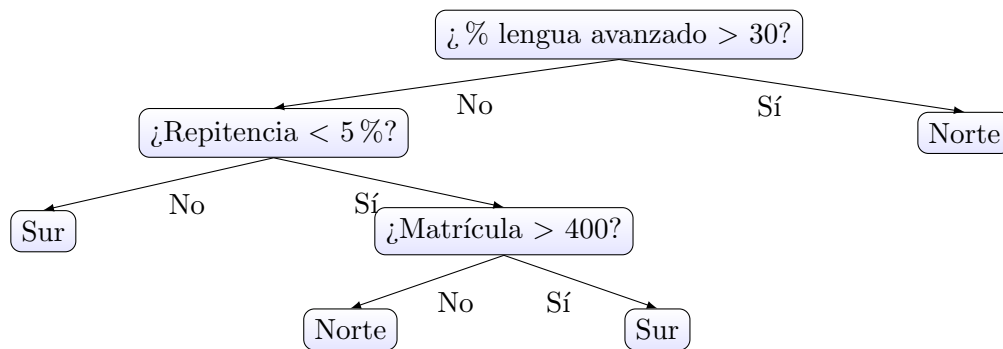


Fig. 3.6: Ejemplo simplificado de árbol de decisión para predecir si una escuela está al norte o al sur del tren Sarmiento.

variables para evaluar las posibles particiones. Esta diversidad estructural es precisamente lo que permite al conjunto ofrecer predicciones más estables y generalizables, al combinar múltiples perspectivas sobre el problema.

Si bien el modelo alcanzó una alta precisión en la clasificación, el interés principal de este ejercicio radicó en evaluar la importancia de cada variable (feature importance) en la tarea de discriminación entre escuelas del norte y del sur. La Figura 3.7 presenta la jerarquía de **importancia de las variables** según el criterio de "permutation importance" proporcionado por el modelo.

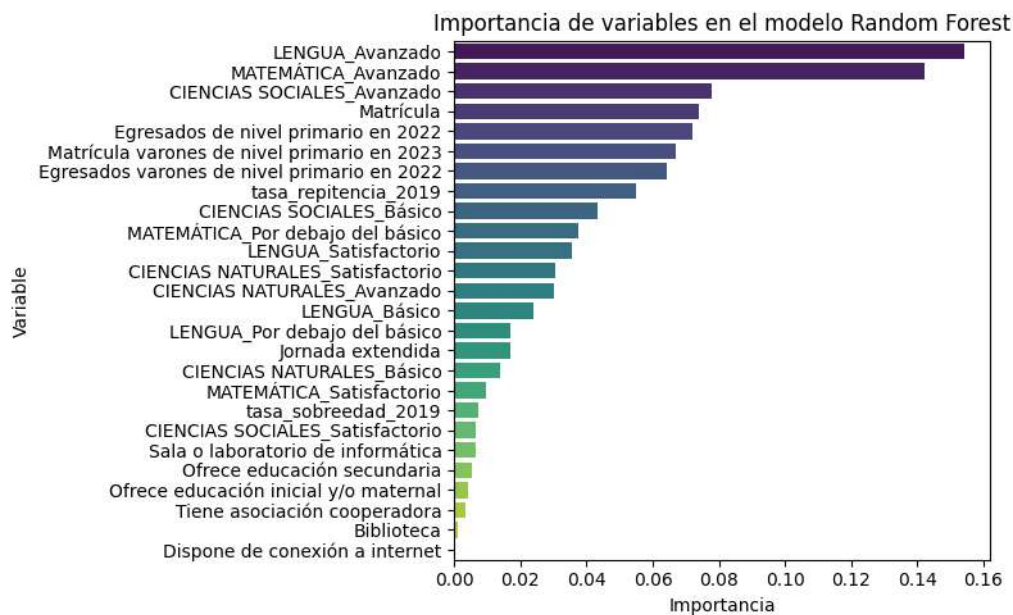


Fig. 3.7: Importancia de las variables en la clasificación de escuelas como "Norte del Tren" o "Sur del Tren" según el modelo Random Forest. Variables con mayor puntuación son consideradas más influyentes por el modelo para realizar la distinción.

Como se observa en la Figura 3.7, las variables con mayor poder discriminatorio son consistentemente aquellas relacionadas con el desempeño educativo (porcentaje de alumnos en niveles *avanzado* o *por debajo del básico* en distintas asignaturas), la matrícula, y las tasas de repitencia y sobreedad. Por el contrario, las variables estructurales bási-

cas (como la disponibilidad de ciertas infraestructuras o equipamientos) mostraron una importancia considerablemente menor. Este resultado refuerza los hallazgos de las secciones anteriores: mientras que la dotación estructural parece ser relativamente homogénea, son las dimensiones de desempeño, trayectoria escolar y escala (matrícula) las que más fuertemente diferencian a las escuelas situadas a uno u otro lado del tren Sarmiento.

3.3.2. Agrupamiento de escuelas con K-Means

Para investigar cómo las escuelas se agrupan naturalmente en función de una combinación de sus características, se aplicó el algoritmo de agrupamiento **K-Means**. Se seleccionó un subconjunto de variables representativas de las dimensiones más relevantes identificadas previamente: Matrícula, porcentaje de alumnos con desempeño *avanzado* en Matemática, porcentaje de alumnos con desempeño *básico* en Ciencias Sociales y la tasa de repitencia. Se exploraron soluciones con $k = 2$ y $k = 4$ clusters.

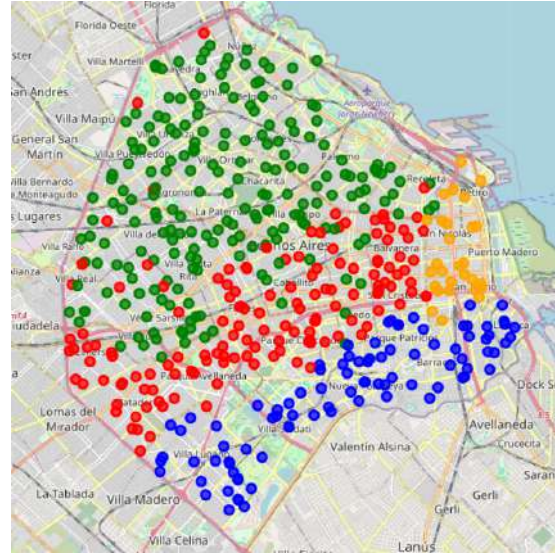
K-Means es un algoritmo de aprendizaje no supervisado que permite identificar patrones en los datos sin necesidad de etiquetas previas [50]. Su objetivo es dividir el conjunto de observaciones en k grupos (o clústeres) tales que los elementos dentro de un mismo grupo sean similares entre sí, y distintos de los elementos de otros grupos. El procedimiento comienza seleccionando aleatoriamente k centroides (centros provisionales de cada clúster), y luego itera entre dos pasos: (1) asignar cada observación al centro más cercano, y (2) recalculan los centroides como el promedio de los puntos asignados a cada grupo. Este proceso se repite hasta que las asignaciones dejan de cambiar significativamente. El resultado permite observar si existen agrupamientos naturales entre las escuelas según su perfil académico y estructural, lo cual puede aportar a la identificación de patrones territoriales o brechas en el sistema educativo.

Los resultados del agrupamiento se visualizaron espacialmente para identificar patrones geográficos.

- **Con $k = 2$ clusters (Figura 3.8a):** El algoritmo tendió a separar las escuelas en dos grandes grupos. Un clúster agrupó predominantemente a escuelas situadas en el extremo sur de la ciudad, particularmente en las comunas 4 y 8. El otro cluster englobó a la mayoría de las escuelas restantes, incluyendo establecimientos tanto al norte como al sur del tren, pero con características promedio distintas al primer grupo.
- **Con $k = 4$ clusters (Figura 3.8b):** La solución con cuatro cluster reveló una segmentación espacial más matizada:
 1. Un primer cluster persistió en agrupar escuelas del extremo sur (similar al caso $k = 2$).
 2. Un segundo cluster se conformó por escuelas situadas también al sur del tren Sarmiento, pero geográficamente al norte del primer clúster del extremo sur, sugiriendo un gradiente dentro de la zona sur.
 3. Un tercer clúster agrupó escuelas localizadas casi exclusivamente al norte del tren Sarmiento.
 4. Un cuarto y último cluster se concentró en la zona este de la ciudad, abarcando áreas del centro, Puerto Madero y Retiro, con características que las diferenciaban de los otros tres grupos.



(a) Mapa de escuelas coloreadas por clúster ($k = 2$).



(b) Mapa de escuelas coloreadas por clúster ($k = 4$).

Fig. 3.8: Visualización espacial de los clústeres de escuelas obtenidos con el algoritmo K-Means para $k = 2$ (izquierda) y $k = 4$ (derecha), basados en variables de matrícula, desempeño y tasa de repitencia. Cada punto representa una escuela, coloreada según el clúster al que fue asignada.

Estos resultados de K-Means sugieren que, incluso considerando una combinación de múltiples variables, emergen agrupamientos de escuelas con una clara correspondencia geográfica. La segmentación norte-sur de la ciudad es evidente, y la solución con $k = 4$ incluso insinúa una heterogeneidad dentro de la zona sur y una particularidad en la zona céntrica/este. Estos patrones espaciales podrían reforzar la idea de una ciudad educativa-mente fragmentada.

3.4. Conclusión parcial del análisis exploratorio sobre datos de escuelas primarias de CABA

El análisis exploratorio de los datos de las escuelas primarias estatales de la Ciudad Autónoma de Buenos Aires, presentado en este capítulo, ha revelado patrones consistentes y significativos que apuntan a una marcada diferenciación entre los establecimientos ubicados al norte y al sur del tren Sarmiento, y más ampliamente, entre las zonas más al norte y al sur de la ciudad.

Inicialmente, se observó que si bien las características estructurales básicas de las escuelas (como la disponibilidad de bibliotecas o acceso a internet) no muestran disparidades sistemáticas a gran escala, la matrícula por establecimiento sí presenta una diferencia, siendo mayor en las escuelas de las comunas del sur.

Posteriormente, el análisis de los indicadores de desempeño educativo y vulnerabilidad (basados en los resultados de las pruebas Aprender 2019, y las tasas de sobreedad y repitencia del mismo año) confirmó y profundizó esta brecha. De manera consistente, las escuelas de las comunas del sur tienden a registrar porcentajes más altos de estudiantes en

los niveles más bajos de desempeño y menores porcentajes en los niveles avanzados, junto con tasas de sobreedad y repitencia más elevadas en comparación con las escuelas de las comunas del norte.

Finalmente, la aplicación de técnicas de aprendizaje automático reforzó estos hallazgos desde una perspectiva multivariada. El modelo Random Forest identificó a las variables de desempeño, matrícula y trayectoria escolar como las más importantes para distinguir entre escuelas del norte y del sur del tren. A su vez, el agrupamiento mediante K-Means evidenció la formación de clusters de escuelas con una clara correspondencia geográfica, reflejando la segmentación norte-sur e incluso sugiriendo gradientes internos dentro de estas zonas.

En conjunto, estos resultados preliminares sugieren la existencia de una división educativa en CABA que se alinea con el conocido gradiente socioeconómico norte-sur[51]. Las escuelas no son homogéneas a lo largo del territorio; por el contrario, presentan perfiles diferenciados que, al menos en términos de resultados agregados y condiciones de trayectoria escolar, podrían implicar experiencias y oportunidades educativas distintas para los alumnos que asisten a ellas.

Habiendo caracterizado estas diferencias entre los establecimientos educativos, surge la pregunta fundamental que guiará la siguiente etapa de esta investigación: ¿Se observa un patrón de movilidad escolar que refuerce esta segmentación territorial? Específicamente, ¿los estudiantes tienden a asistir a escuelas ubicadas en su misma zona (norte o sur), y en qué medida el tren Sarmiento actúa como una barrera que condiciona estas elecciones? Estas interrogantes serán abordadas en los capítulos subsiguientes mediante el análisis de las trayectorias casa-escuela reconstruidas a partir de datos masivos de movilidad.

4. DATOS DE MOVILIDAD Y RECONSTRUCCIÓN DE TRAYECTORIAS

Para investigar los patrones de asistencia escolar y evaluar el impacto del tren Sarmiento como barrera, es imprescindible analizar la **movilidad** cotidiana de los estudiantes. En esta tesis, nos centramos en los viajes realizados en transporte público, utilizando para ello un extenso conjunto de datos derivados de las transacciones de la tarjeta SUBE. El objetivo principal de esta sección es reconstruir las trayectorias desde la casa de los estudiantes a su escuela e identificar aquellas que cruzan el tren Sarmiento.

4.1. Origen y características de los datos SUBE

Los datos de movilidad utilizados en esta investigación provienen de un repositorio público que contiene una tabla de datos con el registro de 14.922.654 viajes pagados con la tarjeta SUBE de un día miércoles representativo de noviembre de 2019 para la Región Metropolitana de Buenos Aires (RMBA) [35]. Este repositorio, desarrollado en el marco de una iniciativa del Banco Interamericano de Desarrollo, no solo ofrece el conjunto de transacciones SUBE sino que también aporta la matriz Origen-Destino final y los procedimientos metodológicos (scripts de PostgreSQL + PostGIS) empleados para su adquisición.

La obtención original de los datos SUBE se realizó mediante un pedido de acceso a la información pública ¹. Es fundamental destacar las medidas tomadas para preservar la privacidad de los usuarios en la divulgación de estos datos. Según lo informado en [35], se omitió la identificación directa de la tarjeta SUBE, enmascarando este campo. Adicionalmente, se implementaron procesos de agregación espacial y temporal: no se proporcionan los minutos exactos de las transacciones (solo la hora) y las coordenadas geográficas de las transacciones fueron truncadas a tres decimales. Este truncamiento implica un margen de error espacial de aproximadamente 100 metros, lo que dificulta la identificación precisa de puntos de origen o destino exactos, pero permite un análisis a nivel agregado. El archivo principal de transacciones ('transacciones.csv'), disponible en el directorio 'data/' del repositorio mencionado, constituye el insumo fundamental para nuestro análisis.

Cada registro en el conjunto de datos de transacciones representa una validación de la tarjeta SUBE en un medio de transporte público. La Tabla 4.1 muestra un extracto de las primeras filas de esta tabla, ilustrando las variables más relevantes para la presente investigación.

De las variables disponibles, las más cruciales para nuestro estudio son:

- **id_tarjeta**: Un identificador único (aunque enmascarado) para cada tarjeta SUBE, lo que permite rastrear las transacciones secuenciales realizadas por un mismo usuario a lo largo del día.
- **modo**: Indica el medio de transporte utilizado (COL para colectivo, TRE para tren, SUB para subterráneo).

¹ Expediente EX-2020-32945006-DNAIP#AAIP

Tab. 4.1: Ejemplo de estructura de los datos de transacciones SUBE (selección de variables relevantes).

id_tarjeta	modo	lat	lon	sexo	id_linea	id_tarifa	hora
100001	COL	-34,570	-58,508	M	4	1	16
100001	COL	-34,570	-58,508	M	4	1	7
100002	TRE	-34,602	-58,494	F	6	5	8
100003	COL	-34,566	-58,512	F	84	1	19
100004	SUB	-34,584	-58,750	M	21	1	22
100005	COL	-34,584	-58,750	F	3	5	6
100005	TRE	-34,584	-58,750	F	5	2	15
100006	COL	-34,616	-58,636	F	19	4	15
100006	COL	-34,616	-58,636	F	13	4	19
100007	SUB	-34,584	-58,472	M	8	11	8

- **lat** y **lon**: Coordenadas de latitud y longitud (truncadas) del lugar donde se realizó la validación del pasaje (aproximado de la parada de colectivo o la estación de tren/subte).
- **id_tarifa**: Identificador del tipo de tarifa aplicada en la transacción. Esta variable es particularmente importante, ya que un valor específico (**id_tarifa** = 11, según la codificación provista con los datos) permite identificar las transacciones correspondientes a tarifas subsidiadas para **estudiantes de nivel primario**. Este filtro es esencial para aislar la movilidad del grupo de interés de esta tesis.
- **hora**: La hora (sin minutos) en que se registró la transacción.
- Otras variables como **sexo** o **id_linea** (que no corresponde directamente a la línea de colectivo, tren o subte sino que es un identificador) también pueden ser útiles para enriquecer el análisis o para filtrar los datos.

Estos datos de transacciones, a pesar de las agregaciones y enmascaramientos por privacidad, ofrecen una oportunidad sin precedentes para reconstruir secuencias de viaje y, a partir de ellas, inferir las trayectorias casa-escuela de los estudiantes de nivel primario que utilizan el transporte público en CABA. El proceso de transformación de estas transacciones en trayectorias significativas se detallará en la siguiente sección. Se lograron reconstruir 5425 trayectorias, las cuales forman los datos con los que se trabajó y concluyeron los resultados.

4.2. Construcción de pares casa-escuela

Una vez filtrados los registros de la tarjeta SUBE para identificar aquellos correspondientes a estudiantes de nivel primario durante un día lectivo típico (como se detalló en la Sección 4.1), el siguiente paso consistió en reconstruir las trayectorias individuales **casa-escuela**. Dado que un mismo estudiante puede realizar múltiples viajes en transporte público a lo largo del día, y que un solo desplazamiento hacia la escuela puede implicar el uso de varios medios de transporte (transbordos), fue necesario establecer criterios para

inferir el **origen** (casa aproximada) y el **destino** (escuela aproximada) de cada estudiante. Es importante destacar que el destino no corresponde con la dirección de una escuela específica, se podría haber buscado la escuela más cercana a este punto y asignarla como su escuela pero se prefirió trabajar con el par (lat,lon) y definirlo como escuela aproximada.

Para ello, en primer lugar, los viajes filtrados se agruparon por el identificador único de la tarjeta SUBE ("id_tarjeta") y se ordenaron cronológicamente según la hora de validación del viaje. Esto permitió obtener una secuencia de todos los usos del transporte público realizados por cada estudiante durante el día analizado. A modo de ejemplo, la Tabla 4.2 ilustra cómo se verían los registros de viajes para un mismo "id_tarjeta" una vez agrupados y ordenados.

Tab. 4.2: Ejemplo de secuencia de viajes en transporte público para un mismo "id tarjeta" (estudiante) durante un día, ordenados cronológicamente.

id_tarjeta	Nro. Viaje	Hora	id_linea	Latitud	Longitud
123456	1	07	121	-34,620	-58,400
123456	2	08	34	-34,605	-58,380
123456	3	12	34	-34,590	-58,370
123456	4	13	121	-34,600	-58,375

Nota: Los datos son ficticios y sirven para ilustrar la estructura.

A partir de estas secuencias de viajes por estudiante, se aplicaron los siguientes criterios heurísticos para inferir el par origen-destino casa-escuela:

1. **Determinación del Origen (Casa aproximada):** Se consideró que la ubicación geográfica (latitud, longitud) donde el estudiante realiza el pago del primer viaje en transporte público del día corresponde a una aproximación de la localización de su casa. Esta asunción se basa en la premisa de que el primer desplazamiento matutino suele iniciarse desde la vivienda para dirigirse a la escuela.
2. **Determinación del Destino (Escuela aproximada):** Identificar el destino escolar requirió un análisis de las secuencias de viaje para detectar posibles transbordos. Se definió una **secuencia de viaje de ida** como un conjunto de uno o más viajes consecutivos realizados por el mismo estudiante, donde cada viaje subsiguiente ocurre en la misma hora o, como máximo, una hora después del viaje anterior. Esta ventana temporal de una hora se estableció debido a que los datos de la tarjeta SUBE disponibles para este estudio no contaban con la granularidad de los minutos en el registro de validación, solo la hora.

Bajo esta definición:

- Si un estudiante realiza un único viaje en la mañana (por ejemplo, a las 7:00 hs) y su siguiente viaje es mucho más tarde (por ejemplo, a las 12:00 hs o 17:00 hs), se asume que ese primer viaje lo llevó directamente a la escuela. La ubicación de pago de ese único viaje matutino se considera entonces el origen (su casa) y la ubicación del viaje siguiente se considera el destino (su escuela).
- Si un estudiante realiza múltiples viajes en secuencia horaria (e.g., un viaje a las 7:00 hs seguido de otro a las 7:00 hs o a las 8:00 hs), se considera que estos forman parte de un mismo desplazamiento hacia la escuela (implican

transbordos). En este caso, el destino (escuela) se infiere a partir del primer viaje que rompe esta secuencia matutina. Es decir, el primer viaje que el estudiante realiza después de haber completado su secuencia de viajes de ida a la escuela se interpreta como el inicio de su viaje de regreso a casa. Por lo tanto, la ubicación de pago de este primer viaje de la "tarde" o "mediodía" (el que sigue a la secuencia de ida) se toma como la localización aproximada de la escuela. Se asume que el estudiante inicia este viaje desde las cercanías de su establecimiento educativo.

Por ejemplo, en la Tabla 4.2, para el "id_tarjeta" 123456:

- El Origen (casa) se establecería en la ubicación de pago del Viaje 1 (Lat: -34,620, Lon: -58,400), realizado a las 07 hs.
- Los Viajes 1 (07 hs) y 2 (08 hs) forman una secuencia de ida.
- El Viaje 3, realizado a las 12 hs, rompe esta secuencia matutina. Por lo tanto, la ubicación de pago del Viaje 3 (Lat: -34,590, Lon: -58,370) se consideraría la aproximación del Destino (Escuela).

Es importante reconocer que esta metodología de inferencia de pares origen-destino se basa en asunciones sobre los patrones típicos de movilidad escolar y está sujeta a las limitaciones de los datos disponibles (principalmente la granularidad horaria y la ausencia de información sobre el propósito del viaje). Se consideró que para dos viajes con el mismo registro horario de transacción, el que aparece primero en el dataset es el viaje que se hizo antes. Sin embargo, ofrece un enfoque sistemático y aplicable a gran escala para aproximar las trayectorias casa-escuela a partir de los registros de transporte público.

4.3. Reconstrucción de la red de trayectorias y análisis de cruces

Una vez inferidos los pares **origen-destino** (casa-escuela aproximados) para cada estudiante, el paso siguiente fue reconstruir la trayectoria espacial de cada uno de estos desplazamientos. El objetivo principal de esta reconstrucción no es replicar con exactitud el o los medios de transporte específicos utilizados por cada alumno (información no disponible en los datos SUBE), sino obtener una representación geográfica del camino más directo y factible entre la casa y la escuela.

Para ello, se utilizó la biblioteca de Python OSMnx [52]. Esta herramienta permite descargar datos de redes viales directamente desde OpenStreetMap (OSM) —una base de datos geográfica colaborativa y abierta— y modelarlas como una red o grafo sobre la cual se pueden realizar cálculos donde los nodos son las esquinas de las cuadras y las aristas las calles. Es un grafo direccional donde las aristas apuntan en la dirección que se puede recorrer esa calle.

Para cada par origen-destino (casa-escuela aproximados) de un estudiante, se procedió a calcular el camino más corto sobre la red de calles de la Ciudad Autónoma de Buenos Aires. El concepto de "camino más corto" es fundamentalmente el mismo que utilizan aplicaciones de navegación como Google Maps o Waze cuando se les pide una ruta entre dos puntos: el algoritmo busca la secuencia de calles conectadas que minimiza la distancia total a recorrer desde el punto de partida hasta el punto de llegada.

Es crucial destacar que la red de calles obtenida a través de OSMnx es una red direccional. Esto significa que cada calle tiene un sentido de circulación definido (mano única o doble mano). Al calcular el camino más corto, el algoritmo respeta esta direccionalidad, asegurando que las trayectorias generadas sean realistas y no incluyan tramos en contramano.

Además, es importante remarcar que este proceso de ruteo sobre la red de calles es una aproximación del viaje real. No simula el recorrido exacto de las líneas de colectivo o subterráneo que el estudiante pudo haber tomado, ni considera factores como la frecuencia del servicio, los tiempos de espera en paradas o las preferencias individuales de ruta. Sin embargo, para los fines de esta tesis, esta aproximación se considera adecuada y suficiente. El interés principal radica en determinar si una trayectoria, en su recorrido general desde el origen al destino, cruza o no una barrera física específica como el tren Sarmiento, más que en modelar los detalles modales de cada viaje. La ruta más corta en la red vial proporciona una base geométrica consistente para realizar este análisis de intersección a gran escala.

El resultado de este proceso es un conjunto de trayectorias geográficas (líneas) que conectan la casa aproximada de cada estudiante con su escuela aproximada. La Figura 4.1 muestra una visualización de un subconjunto de estas trayectorias reconstruidas sobre un mapa de CABA, ilustrando cómo se distribuyen espacialmente y cómo algunas de ellas intersectan el tren Sarmiento.

El principal objetivo de esta reconstrucción detallada de trayectorias es poder identificar inequívocamente aquellas que cruzan el tren Sarmiento. Esto se logra mediante un análisis de intersección espacial entre las geometrías de las trayectorias y la traza del tren. Contar el número de estas trayectorias "transversales" (aquellas con la casa de un lado del tren y la escuela del otro) es fundamental para el cálculo posterior del Barrier Score intrudido en la próxima sección y para evaluar empíricamente el rol del Sarmiento como frontera en la movilidad escolar.

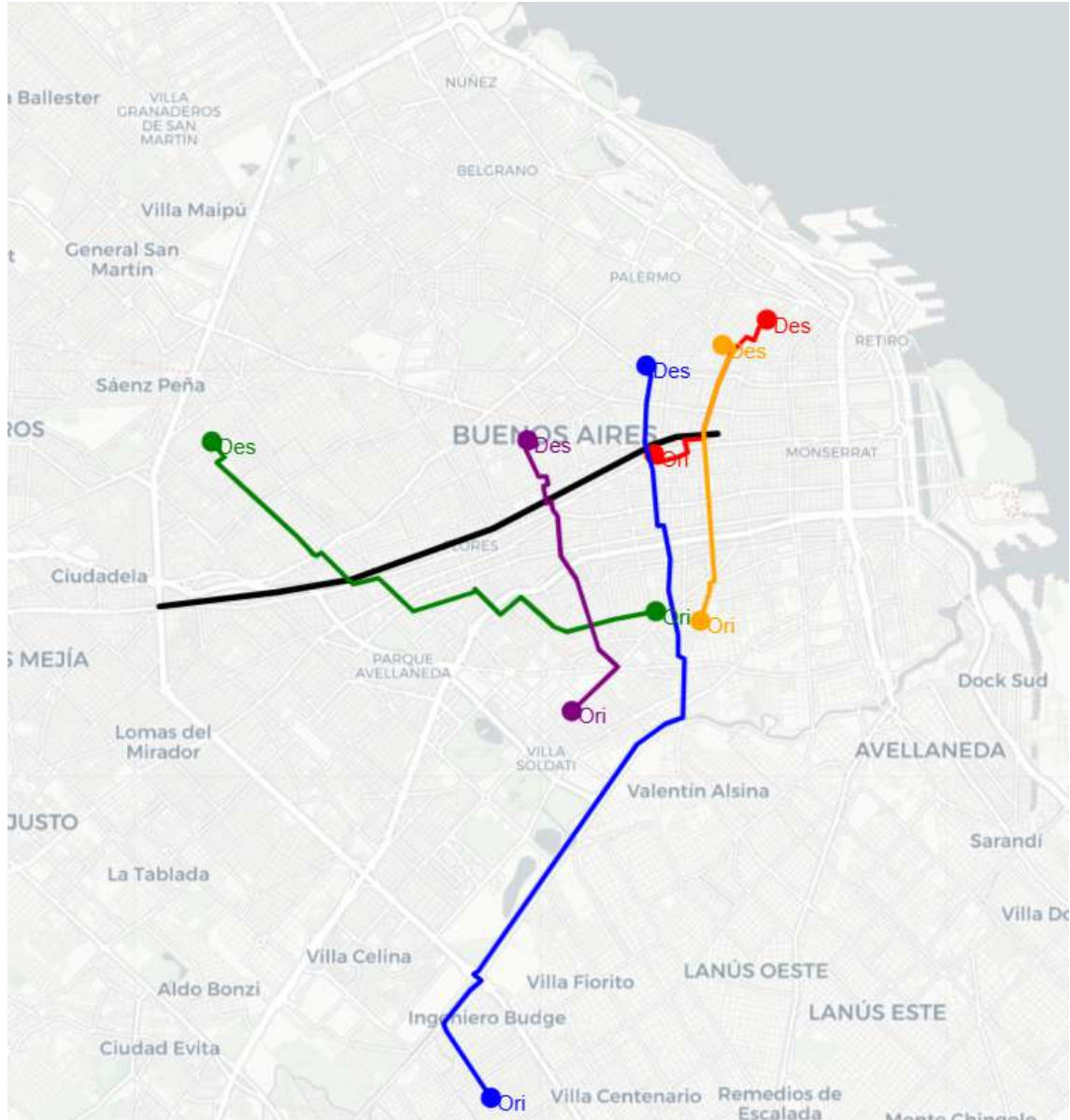


Fig. 4.1: Visualización de un subconjunto de trayectorias casa-escuela reconstruidas en CABA. Cada línea representa el camino más corto estimado sobre la red de calles entre la casa (Ori) y la escuela (Des) de un estudiante. Se puede observar cómo estas trayectorias se distribuyen por la ciudad e interactúan con infraestructuras como el tren Sarmiento (marcado para referencia). En el ejemplo la trayectoria roja se superpone con la naranja en un tramo apareciendo la naranja por encima.

5. METODOLOGÍA DEL BARRIER SCORE (BS)

Para cuantificar el efecto de una infraestructura lineal, como el tren Sarmiento, en tanto barrera para la movilidad escolar, en esta tesis se adopta la metodología del **Barrier Score** (BS) propuesta por Aiello et al. [19]. Esta métrica permite evaluar el impacto de una barrera comparando el número de trayectorias observadas que la cruzan con el número esperado en un escenario contrafactual generado por un **modelo nulo**.

5.1. Definición del Barrier Score

Siguiendo la definición presentada en [19] para una barrera i , el Barrier Score (BS) se calcula como la diferencia relativa entre el número de **cruces esperados** en un modelo nulo (E_i) y el número de **cruces observados** empíricamente (O_i), normalizada por estos últimos:

$$BS_i = \frac{E_i - O_i}{O_i} \quad (5.1)$$

Donde:

- O_i representa el número de cruces observados de la barrera i en los datos reales (trayectorias casa-escuela).
- E_i representa el número de cruces esperados de la misma barrera i , generados a partir de un modelo nulo que simula un escenario donde las trayectorias son indiferentes a la presencia de la barrera.

Con esta formulación, el Barrier Score se interpreta como el incremento porcentual relativo de cruces que se esperarían sobre la barrera i si esta no ejerciera su efecto restrictivo, en comparación con los cruces que efectivamente ocurren. Por ejemplo:

- Un **$BS_i = 0$** (cuando $E_i = O_i$) indica que el número de cruces esperados es igual al observado. Esto sugiere que, bajo las suposiciones del modelo nulo, la barrera i no presenta un efecto neto de obstaculización ni de atracción para las trayectorias analizadas.
- Un **BS_i positivo** (cuando $E_i > O_i$) indica un efecto barrera. Por ejemplo, un $BS_i = +0.62$ (o $+62\%$) significa que, en un escenario sin el efecto restrictivo de la barrera (según el modelo nulo), se esperarían un 62% más de cruces sobre la barrera i de los que realmente se observan. Valores positivos mayores indican un efecto barrera más pronunciado.
- Un **BS_i negativo** (cuando $E_i < O_i$) indicaría que se observan más cruces de los que el modelo nulo predice. Esto podría interpretarse como que la infraestructura, en lugar de una barrera, actúa como un canalizador de flujos, o podría señalar aspectos de la movilidad no capturados adecuadamente por el modelo nulo. Aiello et al. [19] mencionan que, aunque generalmente positivos, los Barrier Scores pueden volverse ligeramente negativos en algunas circunstancias.

■ **Caso $O_i = 0$ (Cero cruces observados):**

- Si $E_i > 0$ (se esperaban cruces, pero no se observó ninguno), el BS es matemáticamente indefinido debido a la división por cero. Esta situación representa el máximo efecto barrera empíricamente observable (una barrera "absoluta" en los datos). Para fines de análisis o visualización, estos casos requieren un tratamiento especial, como ser reportados por separado o mediante la asignación de un valor indicativo de barrera muy alta, si el contexto lo permite y se justifica. (Es importante notar que Aiello et al. [19] no detallan explícitamente el manejo de este caso en la definición principal, pero en la práctica, infraestructuras con $O_i = 0$ y $E_i > 0$ serían inequívocamente barreras fuertes).
- Si $E_i = 0$ y $O_i = 0$, no hay flujo esperado ni observado, por lo que no se puede evaluar un efecto barrera mediante esta métrica.

La robustez del BS y su interpretación dependen críticamente de la adecuada construcción del modelo nulo, el cual debe preservar las propiedades fundamentales de la red de movilidad original mientras elimina cualquier correlación con la ubicación específica de la barrera.

5.2. Modelo nulo de referencia

El cálculo del Barrier Score (BS), tal como se definió en la sección anterior, depende crucialmente de la comparación entre los cruces observados (O_i) de una barrera y los cruces esperados (E_i). Estos últimos se obtienen a partir del azar de un modelo nulo que crea nuevos pares orígenes-destino de manera "aleatoria".

Un **modelo nulo** es una representación simplificada o aleatorizada de un sistema (en nuestro caso, la red de movilidad escolar) que se construye bajo ciertas restricciones o hipótesis específicas. Su propósito principal es servir como punto de referencia o línea base contra la cual comparar los patrones observados en los datos reales. En el estudio de barreras urbanas, el objetivo del modelo nulo es simular cómo se distribuirían las trayectorias y, por ende, cuántos cruces de una barrera específica ocurrirían, "si el efecto particular de esa barrera no existiera" o si las decisiones de movilidad fueran indiferentes a su presencia. Es decir, el modelo nulo intenta capturar un escenario contrafactual donde la barrera no impone una fricción adicional ni una atracción particular al movimiento.

Este modelo busca responder a la pregunta: ¿qué pasaría con los cruces si las trayectorias no estuvieran influenciadas por la presencia del tren, pero sí por otros patrones de comportamiento realistas? La estrategia central es aleatorizar los pares origen-destino, es decir, mantener los orígenes (las casas de los estudiantes) pero reasignar aleatoriamente los destinos (las escuelas) bajo ciertas reglas que reflejen la realidad.

La construcción de este modelo nulo es una decisión metodológica propia de esta tesis. Si bien se inspira en enfoques similares utilizados en la literatura, como el de Aiello et al. [19] para redes sociales, el diseño específico ha sido adaptado conscientemente para el problema de la movilidad escolar. El objetivo es crear un escenario aleatorio que, sin embargo, sea plausible y respete las lógicas de comportamiento observadas.

Una de las propiedades más importantes a preservar en este proceso es la distribución de las distancias de viaje. En la realidad, los estudiantes no eligen escuelas a distancias completamente arbitrarias; existe una fuerte tendencia a asistir a establecimientos más cercanos, y la probabilidad de elegir una escuela disminuye a medida que aumenta la

distancia. Por lo tanto, una regla fundamental de nuestro modelo nulo es que, al reasignar una escuela a un estudiante, se debe respetar la distancia de su viaje original. Si el modelo generara viajes con distancias muy diferentes a las observadas, la comparación de cruces no sería significativa, ya que no estaría reflejando un comportamiento de movilidad realista. El modelo nulo debe, ante todo, generar patrones espacialmente plausibles para poder aislar, por contraste, el único efecto que queremos medir: el de la barrera.

Para asegurar que se respetan las tendencias empíricas de elección de distancia, es crucial considerar la distribución de probabilidad de conexión en función de la distancia origen-destino. Esta distribución representa la propensión observada de los estudiantes a realizar viajes de una determinada longitud. En este trabajo, se estimó esta distribución a partir de las trayectorias casa-escuela reales. Para ello se dividió la distancia de los viajes en bins de 1km y se observó la cantidad de viajes que hay en ese bin, es decir cuántos viajes hay cuya distancia es de 0-1km, cuántos de 1-2km, y así sucesivamente hasta el viaje de mayor distancia. En la Figura 5.1 se muestra el histograma de viajes según su distancia.

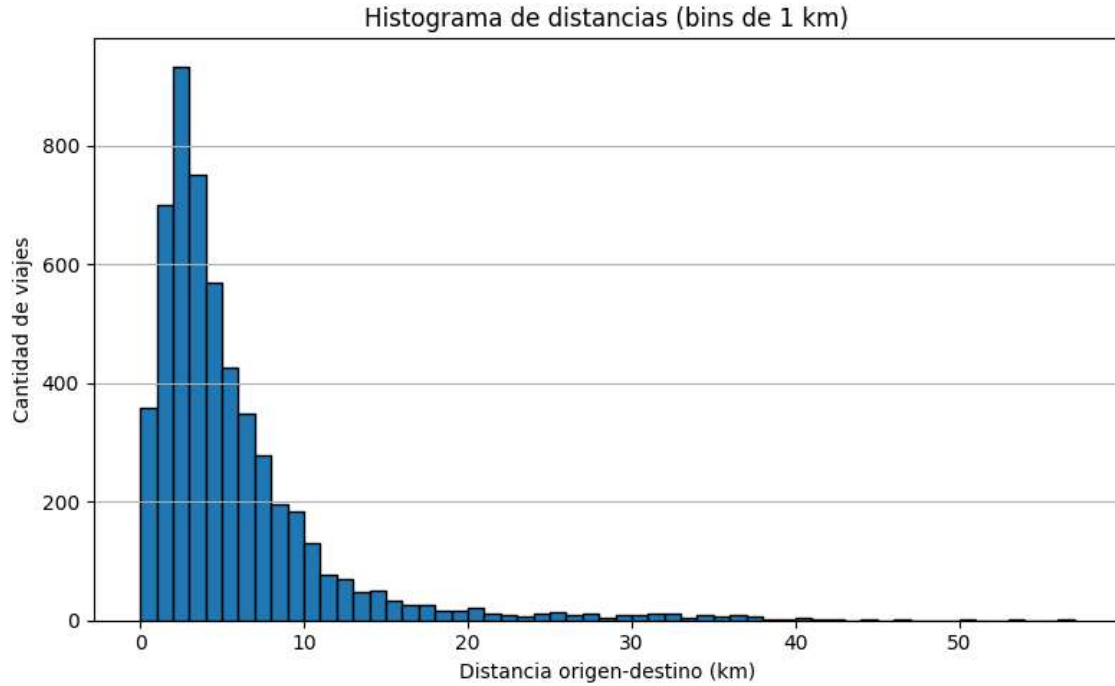


Fig. 5.1: Histograma de cantidad de viajes por bin de 1km

Luego, para cada bin de distancia d , se calculó la probabilidad de conexión $P(d)$ como:

$$P(d) = \frac{\text{Número de viajes reales cuya distancia cae en el bin } d}{\text{Número total de pares origen-destino}}$$

Esta probabilidad $P(d)$ captura la preferencia observada por ciertas longitudes de viaje, normalizada por la disponibilidad de escuelas a esas distancias. La Figura 5.2 muestra la forma típica de esta distribución empírica para las trayectorias casa-escuela en CABA.

Como se puede apreciar en la Figura 5.2, esta distribución es típicamente una función decreciente: la probabilidad de conexión es más alta para distancias cortas y disminuye a medida que la distancia entre la casa y la escuela potencial aumenta. Se observa que la

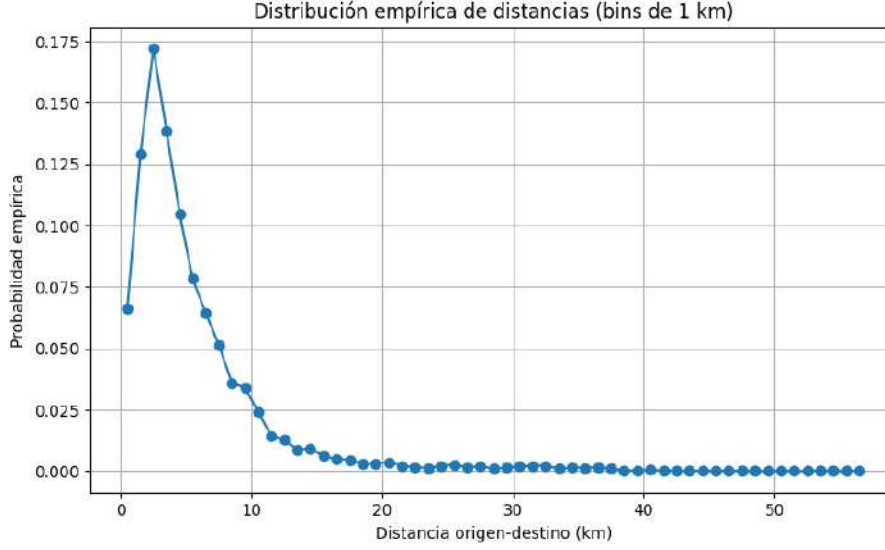


Fig. 5.2: Distribución empírica de la probabilidad de conexión $P(d)$ en función de la distancia origen-destino para las trayectorias casa-escuela en CABA. La distancia se agrupa en bins de 1 km. Se observa una función decreciente, indicando que la probabilidad de que un estudiante se conecte con una escuela disminuye a medida que aumenta la distancia.

mayor cantidad de viajes se concentran en el tercer bin (2-3km) mientras que hay menor cantidad de viajes en los bins anteriores (0-1km y 1-2 km). Este es un patrón intuitivo y ampliamente documentado en estudios de movilidad, a menudo asociado a leyes de decaimiento por distancia o de gravedad espacial [53]. El algoritmo del modelo nulo que se describirá a continuación utiliza esta distribución empírica $P(d)$ para guiar la selección aleatoria de nuevas escuelas, asegurando así que los viajes generados en el escenario nulo respeten las preferencias de distancia observadas en la realidad.

5.2.1. Algoritmo del modelo nulo

Para lograr esta asignación controlada de nuevas escuelas a cada casa, el algoritmo opera en dos fases principales:

Fase 1: Identificación de escuelas candidatas (Función ‘DestinosPosibles’)

Primero, para cada estudiante (es decir, para cada origen de viaje), necesitamos identificar un conjunto de **escuelas candidatas** a las que podría ser reasignado. El criterio principal para que una escuela sea candidata es que su distancia desde la casa del estudiante sea parecida a la distancia del viaje que ese estudiante realizó en la realidad.

Para manejar esto de forma práctica, las distancias se agrupan en “categorías” o “bins”. Por ejemplo, podríamos tener un bin para distancias entre 0 y 1 km, otro para 1-2 km, y así sucesivamente.

- Para cada estudiante, miramos la distancia de su viaje real y vemos a qué categoría (bin) pertenece.

- Luego, para ese mismo estudiante, miramos todas las escuelas en los posibles destinos y calculamos la distancia desde su casa a cada una de ellas.
- Una escuela se considera "candidata válida" para ese estudiante si la distancia desde su casa hasta ella cae en la misma categoría de distancia (o en una categoría muy cercana, por ejemplo, la inmediatamente anterior o posterior) que la de su viaje real.

Al final de esta fase, para cada estudiante (origen), tendremos una lista personalizada de escuelas que son destinos "plausibles" en términos de distancia.

Fase 2: Reasignación aleatoria de escuelas (Función 'RecableoMatching')

Una vez que tenemos las listas de escuelas candidatas para cada estudiante, procedemos a la "reasignación". Este es un proceso iterativo:

- Tomamos a los estudiantes (orígenes) uno por uno, al azar.
- Para el estudiante seleccionado, miramos su lista de escuelas candidatas válidas que aún no hayan sido "asignadas" a otro estudiante en esta simulación nula (no vamos a asignar más de un origen a cada destino).
- Si hay escuelas candidatas disponibles, elegimos una al azar de esa lista y creamos una nueva trayectoria "nula" desde la casa del estudiante hasta esa escuela elegida. Esa escuela se marca como "asignada" y se elimina de las posibles para ser elegida por algún próximo origen.
- Si para un estudiante no quedan escuelas candidatas válidas y disponibles (quizás todas las cercanas ya fueron asignadas a otros en el modelo nulo), ese estudiante podría quedar sin una trayectoria asignada en esta simulación particular del modelo nulo.
- Este proceso se repite hasta que todos los estudiantes han sido considerados o no quedan más asignaciones posibles.

Al final, obtenemos un nuevo conjunto de trayectorias casa-escuela. Estas trayectorias son "nulas" porque, si bien respetan los orígenes y las distancias aproximadas de los viajes reales, la elección específica de la escuela destino fue **aleatoria** (dentro de las candidatas válidas). Sobre este conjunto de trayectorias nulas es que luego contamos cuántas cruzarían el tren Sarmiento (o cualquier otra barrera) para obtener el número de "cruces esperados" (E_i).

A continuación, se presenta el pseudocódigo detallado del algoritmo empleado para generar el modelo nulo.

Algoritmo 1: Generar modelo nulo

```

1 Función DestinosPosibles( $\mathcal{O} : \text{Origenes}, \mathcal{D} : \text{Destinos}$ ):
2   foreach origen  $o \in \mathcal{O}$  do
3     calcular  $\text{bin}(o)$  con la distancia real  $d(o)$ 
4     foreach destino  $d \in \mathcal{D}$  do
5       calcular  $\text{bin}(d|o)$  con la distancia  $d(o, d)$ 
6       if  $|\text{bin}(o) - \text{bin}(d|o)| \leq 1$  then añadir  $d$  a la lista válida de  $o$ 
7   return diccionario  $\mathcal{L}$ : destinos válidos por origen
8 Función RecableoMatching( $\mathcal{O}, \mathcal{D}, \mathcal{L}$ ):
9   while  $\mathcal{O} \neq \emptyset$  y  $\exists o \in \mathcal{O} : \mathcal{L}[o] \cap \mathcal{D}_{\text{libre}} \neq \emptyset^a$  do
10    elegir al azar un origen  $o \in \mathcal{O}$ 
11     $C \leftarrow \mathcal{L}[o] \cap \mathcal{D}$ 
12    if  $C \neq \emptyset$  then
13      elegir al azar un destino  $d \in C$ 
14      asignar arista  $o \rightarrow d$ 
15      quitar  $d$  de  $\mathcal{D}$ 
16    else
17      marcar  $o$  como sin destino válido
18    quitar  $o$  de  $\mathcal{O}$ 
19    recalcular  $\text{distancia\_km}$  para todas las aristas asignadas
20  return Pares Origen - Destino
21

```

^a Esto es lo mismo que pedir que haya orígenes disponibles para asignar y que además exista un destino posible para al menos uno de los orígenes no asignados

Una vez ejecutado el algoritmo es necesario reconstruir la trayectoria casa - nueva escuela para cada origen-destino de la red nula, el algoritmo devuelve los nuevos pares origen-destino a los que hay que calcular la nueva trayectoria. Esto se hace como se comentó anteriormente en la Sección 4.3 . Esto es necesario para poder determinar las trayectorias que cruzan las barreras para el cálculo del BS. Además, el modelo nulo surge del azar de las iteraciones por lo que para ser constante con el trabajo hay que replicar varios modelos nulos y luego promediar los resultados con el fin de no obtener el resultado sesgado por el azar de un solo modelo nulo.

Para ilustrar visualmente el efecto del proceso de recableado del modelo nulo, la Figura 5.3 presenta un ejemplo esquemático con un pequeño conjunto de 10 trayectorias. A la izquierda se muestran las trayectorias originales observadas, y a la derecha, las mismas 10 trayectorias después de haber sido procesadas por el algoritmo del modelo nulo. Se puede apreciar cómo los puntos de origen se mantienen, pero los destinos han sido reasignados aleatoriamente a nuevas escuelas que respetan aproximadamente la distancia original del viaje, resultando en un patrón de conexiones diferente.

Este proceso de generación de trayectorias nulas es fundamental, ya que al comparar el número de veces que estas trayectorias recableadas cruzan una barrera (los "cruces esperados", E_i) con los cruces de las trayectorias reales (los "cruces observados", O_i), podemos aislar el efecto atribuible específicamente a la presencia de la barrera, descontando los patrones que se explicarían por la simple distribución espacial de hogares, escuelas y



(a) Diez trayectorias casa-escuela originales observadas.



(b) Las mismas diez trayectorias después del recableado del modelo nulo.

Fig. 5.3: Ejemplo ilustrativo del proceso de recableado del modelo nulo para 10 trayectorias. (A) Conexiones originales. (B) Conexiones generadas por el modelo nulo, donde los orígenes se preservan pero los destinos se reasignan aleatoriamente manteniendo restricciones de distancia.

las distancias de viaje típicas. Para asegurarse que la función de probabilidad de conexión $P(d)$ se mantiene similar en el modelo nulo y en el real se muestra la Figura 5.4 donde se grafican sobre los mismos ejes ambas funciones.

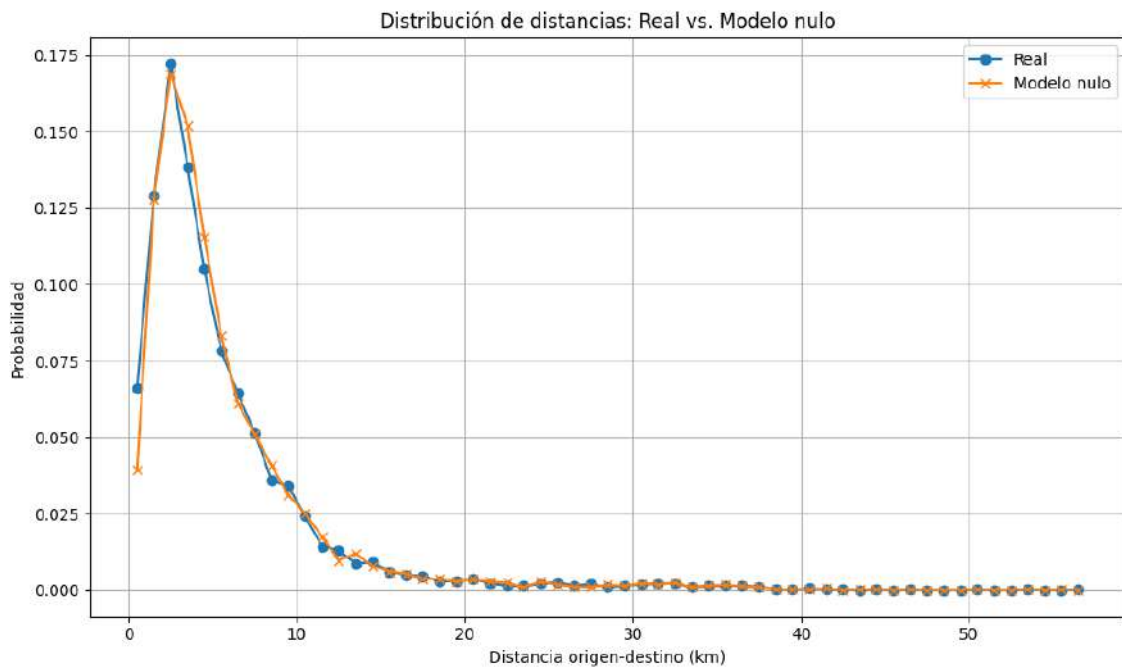


Fig. 5.4: Probabilidad de conexión según bin de distancia para el modelo nulo y el real.

5.3. Barrier Score Direccional (BSD)

El Barrier Score (BS) definido anteriormente proporciona una medida global del efecto barrera de una infraestructura. Sin embargo, en muchos contextos urbanos, y particularmente al estudiar barreras que dividen zonas con características socioeconómicas diferentes, el impacto de una barrera puede no ser simétrico. Es decir, una infraestructura podría obstaculizar más el movimiento en una dirección que en la otra. Para capturar esta posible asimetría, se introduce en este trabajo el concepto de **Barrier Score Direccional (BSD)**.

Es importante destacar que esta extensión direccional no se explora en el trabajo original de Aiello et al. [19] que inspira nuestra métrica del BS. Esto se debe a que su análisis se centra en lazos sociales en redes como Twitter, donde las conexiones (aristas) representan relaciones mutuas (seguimiento recíproco) y, por lo tanto, no poseen una noción intrínseca de dirección del cruce de una barrera física en el mismo sentido que una trayectoria de movilidad. Una trayectoria casa-escuela, en cambio, tiene un origen y un destino claros, lo que permite analizar la direccionalidad del cruce.

El cálculo del Barrier Score Direccional sigue la misma lógica que el BS general, pero se aplica de forma separada para cada una de las dos posibles direcciones de cruce de una barrera lineal. Para una barrera i que divide el espacio en dos lados, Lado A y Lado B, se calculan dos BSD:

1. **BSD del Lado A al Lado B ($BSD_{A \rightarrow B}$):**

$$BSD_{A \rightarrow B} = \frac{E_{A \rightarrow B} - O_{A \rightarrow B}}{O_{A \rightarrow B}}$$

Donde:

- $O_{A \rightarrow B}$ es el número de trayectorias observadas que *se originan en el Lado A y cuyo destino se encuentra en el Lado B*, cruzando así la barrera i .
- $E_{A \rightarrow B}$ es el número de trayectorias generadas por el modelo nulo que *se originan en el Lado A y cuyo destino (aleatorizado) se encuentra en el Lado B*, cruzando la barrera i .

2. **BSD del Lado B al Lado A ($BSD_{B \rightarrow A}$):**

$$BSD_{B \rightarrow A} = \frac{E_{B \rightarrow A} - O_{B \rightarrow A}}{O_{B \rightarrow A}}$$

Donde:

- $O_{B \rightarrow A}$ es el número de trayectorias observadas que *se originan en el Lado B y cuyo destino se encuentra en el Lado A*, cruzando la barrera i .
- $E_{B \rightarrow A}$ es el número de trayectorias generadas por el modelo nulo que *se originan en el Lado B y cuyo destino (aleatorizado) se encuentra en el Lado A*, cruzando la barrera i .

En el caso particular del tren Sarmiento en CABA, que actúa como una división aproximada entre el norte y el sur de la ciudad, se calcularán:

- El BSD para los viajes que se originan en la zona Norte y cruzan hacia la zona Sur ($BSD_{N \rightarrow S}$).
- El BSD para los viajes que se originan en la zona Sur y cruzan hacia la zona Norte ($BSD_{S \rightarrow N}$).

El objetivo de esta descomposición direccional es investigar si una frontera urbana ejerce un efecto divisorio con mayor o menor intensidad en una dirección determinada, afectando de forma diferencial la movilidad de los grupos que residen a uno u otro lado. Por ejemplo, si se encontrara que el $BSD_{N \rightarrow S}$ es marcadamente diferente al $BSD_{S \rightarrow N}$, esto podría indicar que el tren Sarmiento no impacta de la misma manera los cruces desde el norte hacia el sur que aquellos desde el sur hacia el norte. Dichas asimetrías en el efecto barrera podrían estar vinculadas a una variedad de factores que sugieren un análisis sociológico complementario.

La interpretación de cada BSD sigue la misma lógica que el BS global: valores positivos indican un efecto barrera en esa dirección específica, y valores cercanos a cero sugieren una permeabilidad similar a la esperada por el modelo nulo en esa dirección. El manejo de casos donde los cruces observados en una dirección particular ($O_{A \rightarrow B}$ o $O_{B \rightarrow A}$) sean cero también sigue las consideraciones discutidas para el BS general.

5.4. Barrier Score por Distancia ($B(d)$)

El Barrier Score (BS) global, definido anteriormente, ofrece una medida agregada del efecto barrera sobre el conjunto total de trayectorias. Sin embargo, el impacto de una infraestructura puede no ser homogéneo para viajes de diferentes longitudes. Además, al comparar conjuntos de viajes de distintas longitudes, los viajes más largos tienen intrínsecamente una mayor probabilidad de cruzar cualquier línea simplemente por recorrer más espacio.

Para abordar estos matices, Aiello et al. [19] proponen un Barrier Score desagregado por distancia, o **Barrier Score por bin** de distancia, denotado como $B(d)$, que normaliza por la longitud de los viajes.

El primer paso es calcular el número promedio de cruces por unidad de distancia para un conjunto de trayectorias E . Siguiendo la Ecuación [1] del paper, esta cantidad, que llamaremos c_E , se define como:

$$c_E = \frac{1}{|E|} \sum_{(i,j) \in E} \frac{c_{ij}}{d_{ij}}$$

Donde:

- E es un conjunto de trayectorias casa-escuela, desde un origen i a un destino j .
- $|E|$ es el número total de trayectorias en el conjunto.
- c_{ij} es el número de veces que la trayectoria (i, j) cruza la barrera en cuestión.
- d_{ij} es la distancia (longitud) de la trayectoria (i, j) .

Esta métrica, c_E , representa la densidad de cruces promedio para un conjunto de viajes, eliminando el sesgo de que los viajes más largos crucen más veces.

Con esta definición, el Barrier Score por bin de distancia, $B(d)$, se calcula comparando esta densidad de cruces para las trayectorias reales y las del modelo nulo, dentro de un bin de distancia específico d . Siguiendo la Ecuación [2] de [19]:

$$B(d) = \frac{c_{E_{nulo}}(d) - c_{E_{real}}(d)}{c_{E_{real}}(d)}$$

Donde:

- d representa un bin de distancia (por ejemplo, el intervalo de 1 a 2 km).
- $c_{E_{real}}(d)$ es el número promedio de cruces por unidad de distancia (calculado con la Ecuación [1]) para el conjunto de trayectorias reales cuya longitud cae dentro del bin d .
- $c_{E_{nulo}}(d)$ es el número promedio de cruces por unidad de distancia para el conjunto de trayectorias del modelo nulo (recableadas) cuyas longitudes originales también pertenecían al bin d .

La interpretación de $B(d)$ es análoga a la del BS global, pero referida a esta densidad de cruces. Un $B(d) > 0$ indica que para los viajes de longitud d , las trayectorias reales tienen una menor densidad de cruces (cruzan menos por cada kilómetro recorrido) que las trayectorias del modelo nulo, lo cual es indicativo de un efecto barrera.

Este enfoque es más robusto porque, como señalan los autores, la normalización por la distancia d_{ij} "igual a las variaciones que pueden sobrerrepresentar la contribución de los lazos más largos dentro del mismo bin". Al calcular $B(d)$ para una secuencia de bins de distancia, se puede construir un perfil detallado y comparable de cómo el efecto barrera emerge, alcanza un pico (generalmente a distancias cortas) y luego disminuye a medida que aumenta la longitud de los viajes.

6. RESULTADOS

En este capítulo se presentan los principales resultados obtenidos a partir del análisis de las trayectorias casa-escuela y la aplicación de la metodología del Barrier Score (BS) para evaluar el impacto del tren Sarmiento y otras infraestructuras como barreras a la movilidad escolar en CABA.

6.1. Análisis de los cruces observados sobre el tren Sarmiento

A partir de la reconstrucción de las trayectorias casa-escuela de los estudiantes de nivel primario que utilizan transporte público, se identificaron un total de **5.425 trayectorias** completas dentro del área de estudio. De este conjunto, se procedió a analizar cuántas de ellas implicaban un cruce del tren Sarmiento.

Se encontró que un total de **607 trayectorias**, lo que representa aproximadamente el 11,2 % del total de viajes reconstruidos (607/5425), cruzan la línea del tren Sarmiento. Este dato inicial ya sugiere una interacción limitada con esta infraestructura en el contexto de la movilidad escolar en transporte público.

Para visualizar la distribución espacial de estos cruces, la Figura 6.1 muestra la traza del tren Sarmiento dividida por tramos entre estaciones principales, con puntos que indican la ubicación geográfica donde cada una de las 607 trayectorias interseca la vía. Los puntos se han desplazado ligeramente de forma perpendicular a la vía para facilitar la visualización de la densidad de cruces en diferentes segmentos.

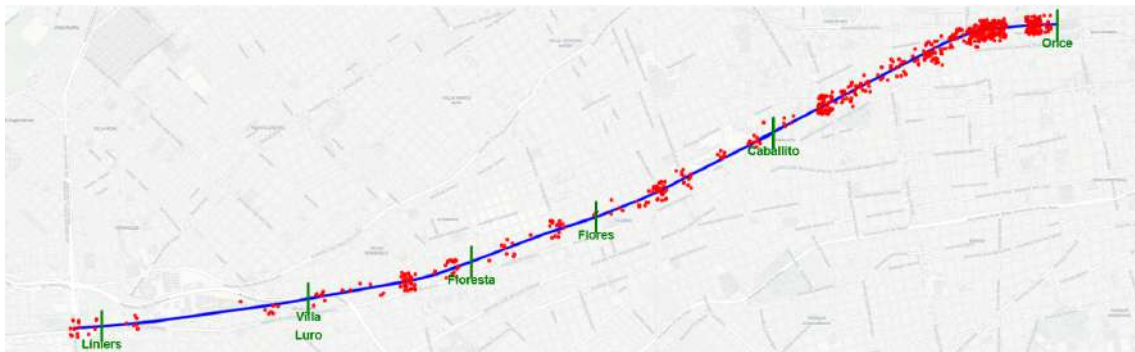


Fig. 6.1: Distribución espacial de los 607 cruces observados sobre la línea del tren Sarmiento por parte de las trayectorias casa-escuela. Cada punto rojo representa la ubicación de un cruce. La línea del tren se muestra segmentada por tramos entre estaciones principales para contextualizar la localización de los cruces. Los puntos han sido ligeramente desplazados para evitar la superposición total y permitir apreciar la densidad.

El análisis de la Figura 6.1 ya sugiere una concentración de los cruces en ciertos tramos de la vía. Para cuantificar esta distribución, se contabilizó el número de cruces por segmento principal del tren Sarmiento dentro de CABA. Los resultados se presentan en la Tabla 6.1.

La Tabla 6.1 evidencia una marcada concentración de los cruces: el **67 %** de todos los cruces observados ocurren en el tramo entre las estaciones de Caballito y Once. Los tramos más occidentales, como Floresta-Flores y Liniers-Villa Luro, presentan un porcentaje de

Tab. 6.1: Distribución porcentual de los 607 cruces observados sobre el tren Sarmiento, por tramo entre estaciones principales.

Tramo del tren Sarmiento	Porcentaje y cantidad de cruces observados
Caballito - Once (3,8km)	67,5 % (410)
Flores - Caballito (2,1km)	12,7 % (77)
Floresta - Flores (1,4m)	5,4 % (33)
Villa Luro - Floresta (2km)	9,7 % (59)
Liniers - Villa Luro (2,3km)	4,7 % (28)
Total de Cruces	100 % (607 cruces)

cruces considerablemente menor. Esta heterogeneidad en la distribución de los cruces podría interpretarse como una primera indicación de que diferentes secciones de la vía férrea podrían presentar distintos grados de "permeabilidad" o, inversamente, actuar como una "frontera más dura" en aquellos segmentos con menor cantidad de cruces, asumiendo una distribución comparable de orígenes y destinos a ambos lados.

Finalmente, se analizó la direccionalidad de estos 607 cruces para determinar si existe un flujo predominante en una dirección (Norte-Sur o Sur-Norte) sobre la barrera del Sarmiento. Los resultados son los siguientes:

- Trayectorias originadas al Sur del tren y con destino al **Norte**: 471 cruces (**77,6 %** del total de cruces).
- Trayectorias originadas al Norte del tren y con destino al Sur: 136 cruces (**22,4 %** del total de cruces).

Estos datos muestran una fuerte asimetría en la dirección de los cruces: la gran mayoría de los estudiantes que cruzan el tren Sarmiento en transporte público para ir a la escuela lo hacen desde la zona Sur hacia la zona Norte. Este patrón sugiere, a priori y antes de aplicar el modelo nulo, que el tren Sarmiento podría estar actuando como una barrera más significativa para los estudiantes que residen al Norte y que viajan al sur. Esta hipótesis será evaluada con mayor rigor mediante el cálculo del Barrier Score Direccional.

Estos hallazgos descriptivos sobre los cruces observados sientan las bases para la aplicación del modelo nulo y el cálculo del Barrier Score, lo que permitirá cuantificar el efecto barrera del tren Sarmiento de una manera más robusta, controlando por la distribución espacial de orígenes, destinos y las distancias de viaje.

6.2. Comparación del efecto barrera: el tren Sarmiento frente a otras infraestructuras

Tras analizar los patrones de cruce observados sobre el tren Sarmiento, el siguiente paso fue cuantificar su efecto barrera mediante el cálculo del Barrier Score (BS), tal como se describió en el Capítulo 5 con la Ecuación 5.1. Para cada infraestructura analizada i , se generaron 20 realizaciones del modelo nulo, y el número de cruces esperados (E_i) se obtuvo promediando los resultados de estas simulaciones.

Con el objetivo de contextualizar la magnitud del efecto barrera del tren Sarmiento, se calculó el BS para esta línea ferroviaria y se la comparó con otras infraestructuras lineales relevantes en CABA: todas las demás líneas de tren de pasajeros que atraviesan la ciudad,

la Autopista Perito Moreno y la Avenida Juan B. Justo (una arteria vial principal que no es autopista). Los resultados de estos cálculos se resumen en la Tabla 6.2. Para todos los casos, el modelo nulo se construyó preservando la distancia de los viajes (considerando un radio de 1 km alrededor de la distancia original para la selección de destinos aleatorios, denotado como "1km" en los datos de cruces esperados).

Tab. 6.2: Comparación del Barrier Score (BS) para el tren Sarmiento y otras infraestructuras lineales en CABA. Los cruces esperados se basan en un modelo nulo que preserva la distancia del viaje (radio de 1 km).

Infraestructura (Barrera)	Cruces Observados (O)	Cruces Esperados (E)	Barrier Score (BS)
Tren Sarmiento (completo)	607	926,79	0,527
Tren Sarmiento (sin tramo Caballito-Once)	211	374,89	0,806
Tren Sarmiento (Caballito - Once)	406	576,58	0,420
Avenida Juan B. Justo	936	943,11	0,008
Autopista Perito Moreno	253	288,68	0,141
Tren Belgrano Sur	607	671,32	0,106
Tren Mitre	529	609,16	0,152
Tren Roca	41	69,05	0,684
Tren San Martín	836	900,16	0,077
Tren Urquiza	187	272,84	0,459

Nota: "Cruces Esperados (E)" es el promedio de 20 realizaciones del modelo nulo.

Los resultados presentados en la Tabla 6.2 son reveladores. El tren Sarmiento (completo) exhibe un Barrier Score alto (**BS = 0,527**) comprado con todas las infraestructuras analizadas que presentan un número considerable de cruces. Un BS de 0,527 para el Sarmiento implica que, según el modelo nulo, se esperarían aproximadamente un 52,7 % más de cruces sobre esta vía de los que realmente ocurren, indicando un gran efecto barrera para la movilidad escolar en transporte público. El Tren Roca muestra un valor aún mayor (**BS = 0,684**), aunque sobre un número bastante menor de cruces observados (41), lo que sugiere un efecto barrera fuerte aunque sobre un flujo reducido.

Es particularmente notable el resultado obtenido al analizar el tren Sarmiento excluyendo su tramo más permeable (Caballito-Once) donde se concentraba el 67 % de los cruces reales. Al aislar los segmentos restantes de la vía, el Barrier Score se incrementa a **0,777**, sugiriendo que todo el tramo anterior (desde Liniers a Caballito) actúa como una barrera aún más pronunciada para las trayectorias casa-escuela. Esto refuerza la idea de que, si bien algunos puntos pueden facilitar el cruce, gran parte de la traza del Sarmiento representa un obstáculo significativo. A su vez, el tramo Caballito-Once por sí solo muestra un BS de **0,386**, considerablemente más bajo.

En comparación, la Avenida Juan B. Justo (**BS = 0,008**) y la Autopista Perito Moreno (**BS = 0,141**) presentan valores de BS mucho menores, sugiriendo un impacto considerablemente más bajo como barreras para la movilidad escolar estudiada. Entre las otras líneas de tren, el Tren Mitre (**BS = 0,152**), el Tren San Martín (**BS = 0,077**) y el Tren Belgrano Sur (**BS = 0,106**) muestran efectos barrera de magnitud moderada, mientras que el Tren Urquiza presenta un BS también alto (0,459), aunque para un número menor cruces observados menor (187).

Para ofrecer una perspectiva visual integrada de estos hallazgos, la Figura 6.2 presenta un mapa de CABA donde se grafican las principales infraestructuras analizadas, coloreadas según la magnitud de su Barrier Score.

La visualización espacial de la Figura 6.2 corrobora los resultados tabulados, destacando visualmente al Tren Sarmiento como la infraestructura con el efecto barrera más intenso para la movilidad escolar en transporte público, según esta métrica y el modelo

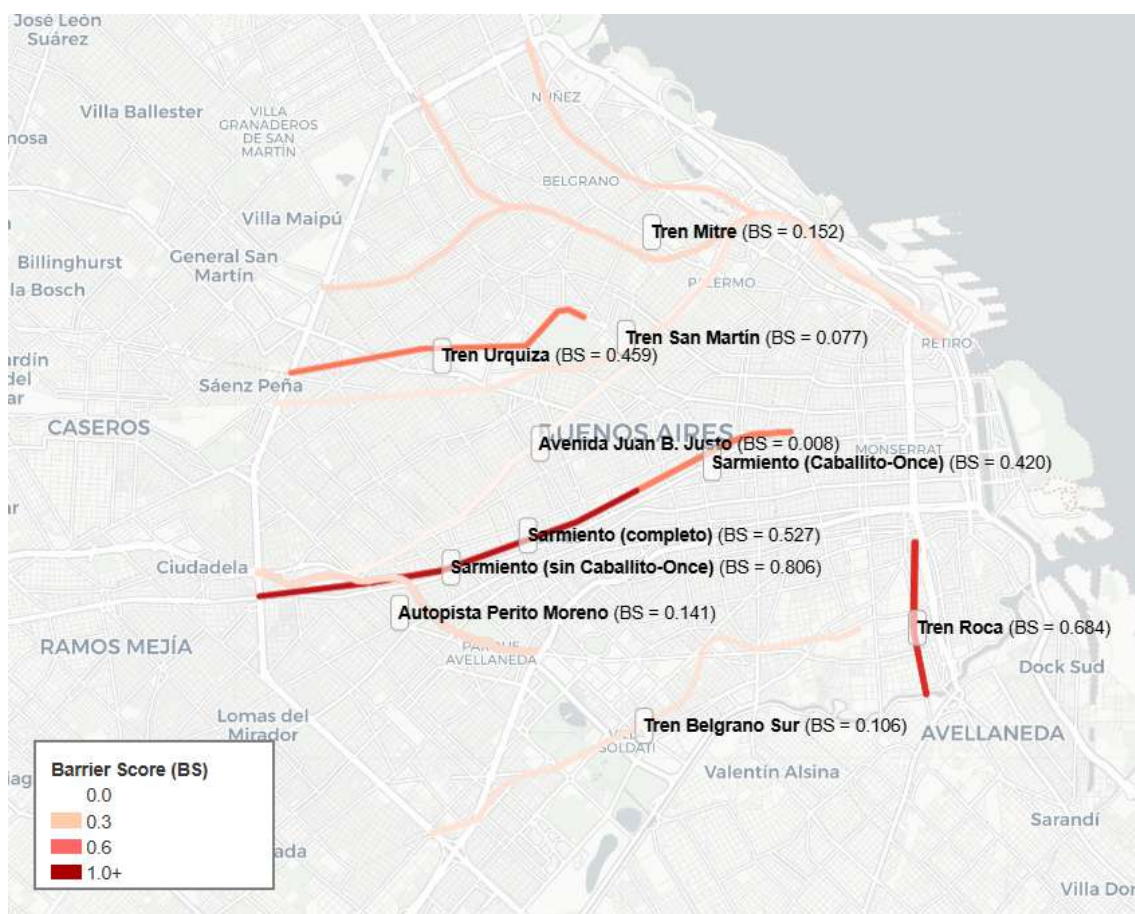


Fig. 6.2: Mapa del Barrier Score (BS) para diferentes infraestructuras lineales en CABA. El color de cada línea representa la magnitud de su BS: tonalidades más rojizas indican un mayor efecto barrera (BS más alto), mientras que colores más claros o neutros indican un efecto barrera menor. Las infraestructuras no analizadas o con datos insuficientes se muestran en un color base.

nulo aplicado. El tren Roca también muestra grandes valores para el BS pero no son muy representativos ya que es una zona de poca movilidad y no se observan muchos cruces.

6.3. Análisis direccional del efecto barrera

Como se discutió en la sección metodológica (Sección 5.3), el impacto de una barrera urbana puede ser asimétrico. Para investigar esta posibilidad, se calculó el Barrier Score Direccional (BSD) para el tren Sarmiento y las demás infraestructuras analizadas. Este cálculo desagrega el efecto barrera según la dirección del cruce.

Para facilitar la nomenclatura, nos referiremos a las dos direcciones de cruce como "Sur → Norte" (S→N) y "Norte → Sur" (N→S). Es importante aclarar que estos términos son una generalización para simplificar la exposición, inspirada en la orientación predominante del tren Sarmiento. Sin embargo, no todas las infraestructuras analizadas siguen una orientación estrictamente Este-Oeste, por lo que el "lado Norte" o "lado Sur" de cada barrera se define funcionalmente en relación con su propia geometría.

Hecha esta aclaración, los resultados del análisis direccional se presentan en la Tabla 6.3.

Tab. 6.3: Barrier Score Direccional (BSD) para diversas infraestructuras en CABA. Se muestran los cruces observados (O), los cruces esperados por el modelo nulo (E, promedio de 20 realizaciones), y el BSD para cada dirección.

Infraestructura	Dirección Sur $\rightarrow$ Norte (S $\rightarrow$ N)			Dirección Norte $\rightarrow$ Sur (N $\rightarrow$ S)		
	$O_{S\rightarrow N}$	$E_{S\rightarrow N}$	$BSD_{S\rightarrow N}$	$O_{N\rightarrow S}$	$E_{N\rightarrow S}$	$BSD_{N\rightarrow S}$
Sarmiento (completo)	472	656,95	0,392	135	269,84	0,999
Sarmiento (Caballito-Once)	324	408,79	0,261	82	167,79	1,046
Sarmiento (sin Caballito-Once)	148	272,68	0,842	53	102,21	0,928
Avenida Juan B. Justo	495	497,53	0,005	441	445,58	0,010
Autopista Perito Moreno	213	229,68	0,078	40	59,00	0,475
Tren Belgrano Sur	552	576,47	0,044	55	94,84	0,724
Tren Mitre	274	324,37	0,184	255	284,79	0,117
Tren Roca	30	51,05	0,702	11	18,00	0,636
Tren San Martín	444	483,89	0,090	392	416,26	0,062
Tren Urquiza	79	131,84	0,669	108	141,00	0,306

Para el tren Sarmiento, se observa una clara diferencia en el impacto según la dirección del cruce.

- Considerando la traza completa del tren, el efecto barrera es bastante más alto para los viajes originados al Norte y con destino al Sur (**BSD = 0,999**) en comparación con aquellos originados al Sur y con destino al Norte (**BSD = 0,393**). Un BSD de 0,999 implica que, bajo el modelo nulo, se esperarían un 99,9 % más de cruces desde el Norte hacia el Sur de los que realmente se registran, lo que sugiere que el Sarmiento actúa como un obstáculo muy considerable para los estudiantes que residen al norte de la vía y que se dirigen a escuelas al sur.
- En su tramo principal (Caballito-Once), el efecto barrera es drásticamente más alto para los viajes originados al Norte y con destino al Sur (**BSD = 1,046**) en comparación con aquellos originados al Sur y con destino al Norte (**BSD = 0,261**).
- Esta tendencia, donde el Sarmiento representa una barrera más fuerte para los cruces Norte $\rightarrow$ Sur, se mantiene también pero con mucha menos diferencia si consideramos los segmentos restantes de la línea (denominados "Sarmiento sin Caballito-Once"), donde el BSD N $\rightarrow$ S es **0,928** frente a un BSD S $\rightarrow$ N de **0,842**.

Estos datos sugieren que, en general, los estudiantes cuyas viviendas se encuentran al norte del Sarmiento cruzan en menor proporción hacia el sur de lo que se esperaría, en comparación con los estudiantes del sur que cruzan hacia el norte. Por lo tanto, el tren Sarmiento parece imponer una restricción más severa a la movilidad escolar para quienes residen en su margen norte.

Un aspecto interesante surge al comparar el BSD máximo con el BS global. Mientras que el BS global más alto para el Sarmiento se observaba en los tramos sin la sección Caballito-Once, el Barrier Score Direccional Norte $\rightarrow$ Sur más elevado (BSD = 1,046) se registra precisamente en el tramo principal Caballito-Liniers. Sin embargo, para la dirección Sur $\rightarrow$ Norte, el BSD más alto sí corresponde a los tramos del Sarmiento sin la sección Caballito-Liniers (BSD = 0,842). Estas diferencias en la localización del máximo efecto

barrera según la dirección del cruce son complejas y podrían estar relacionadas con la combinación de la distribución de la oferta escolar, la permeabilidad diferencial de los cruces en distintos segmentos y las características socioeconómicas de las poblaciones aledañas. Si bien una interpretación exhaustiva de estas dinámicas específicas excede los alcances de la presente tesis, los resultados resaltan la importancia de un análisis direccional.

En cuanto a las otras infraestructuras, se observan diversos patrones. Por ejemplo, el Tren Belgrano Sur también exhibe una fuerte asimetría, actuando como una barrera considerablemente mayor para los viajes Norte $\rightarrow$ Sur (BSD = 0,724) que en la dirección opuesta (BSD = 0,044). Otras líneas como el Tren Urquiza también muestran diferencias direccionales, mientras que la Avenida Juan B. Justo y los trenes Mitre, Roca y San Martín presentan un impacto más simétrico entre direcciones, aunque con magnitudes de efecto barrera variables.

Para complementar la información tabular y ofrecer una perspectiva geográfica de estas asimetrías, la Figura 6.3 visualiza los Barrier Scores Direccionales para las principales infraestructuras. En este mapa, cada línea que representa una barrera se presenta de manera que puedan distinguirse los dos valores direccionales. Por ejemplo, para una barrera con orientación predominante Este-Oeste, la franja superior de la línea podría representar el BSD para los cruces Norte $\rightarrow$ Sur y la franja inferior el BSD para los cruces Sur $\rightarrow$ Norte, cada una coloreada según la magnitud de su respectivo BSD.

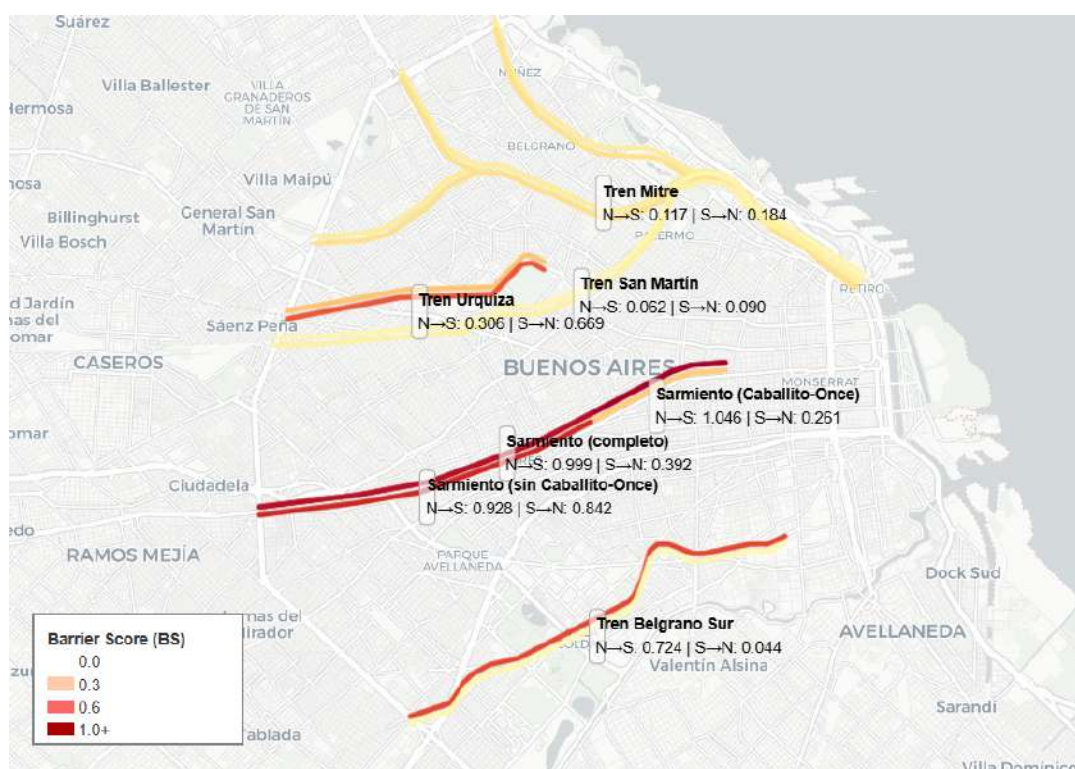


Fig. 6.3: Mapa del Barrier Score Direccional (BSD) para diferentes trenes en CABA. Para cada barrera, se indican dos valores de BSD: uno para la dirección Norte $\rightarrow$ Sur (ej. franja superior/color X) y otro para la dirección Sur $\rightarrow$ Norte (ej. franja inferior/color Y). El color de cada componente direccional representa la magnitud de su BSD, siguiendo una escala donde tonalidades más rojizas indican un mayor efecto barrera y colores más claros o neutros un efecto menor.

La visualización en la Figura 6.3 permite apreciar geográficamente las diferencias direccionales en el efecto barrera. Se puede observar, por ejemplo, cómo el tren Sarmiento presenta una coloración que indica un alto BSD para la dirección Norte→Sur a lo largo de gran parte de su traza, mientras que la coloración para la dirección Sur→Norte, aunque también indicativa de un efecto barrera, es generalmente menos intensa. Este tipo de representación visual ayuda a identificar no solo qué infraestructuras son barreras, sino también si su impacto es predominantemente unidireccional o bidireccional.

En conclusión, el análisis direccional, especialmente para el tren Sarmiento, subraya que su impacto como barrera no es uniforme. La marcada dificultad para los cruces Norte → Sur es un hallazgo central, sugiriendo que las restricciones a la movilidad escolar impuestas por esta infraestructura son experimentadas de manera desigual, lo que podría tener implicaciones significativas para la equidad en el acceso a oportunidades educativas en CABA.

6.4. Análisis del efecto barrera por distancia de viaje

Como se describió en la metodología, el impacto de una barrera puede depender de la longitud de los viajes. Para investigar esta dinámica, se calculó el **Barrier Score por bin** de distancia, $B(d)$, para el tren Sarmiento y las demás infraestructuras analizadas. Este enfoque permite observar cómo varía el efecto barrera para viajes de diferentes longitudes.

Se calcularon los valores de $B(d)$ para los primeros 15 bins de distancia, cada uno con un ancho de **1 km**. De esta manera, se obtuvo el efecto barrera para viajes con longitudes entre 0-1 km, 1-2 km, y así sucesivamente hasta el rango de 14-15 km. Para cada bin y cada barrera, el cálculo se realizó promediando los resultados de 20 realizaciones del modelo nulo.

La Figura 6.4 presenta un resumen visual de estos resultados en forma de *heatmap*. Cada fila corresponde a una infraestructura diferente y cada columna a un bin de distancia. El color de cada recuadro representa la magnitud del Barrier Score ($B(d)$) para esa barrera y ese rango de distancia específicos. La escala de color se ha diseñado para que los tonos rojizos indiquen un efecto barrera positivo (más cruces esperados que observados), mientras que los tonos verdosos indicarían un efecto de atracción (más cruces observados que esperados). La intensidad del color refleja la magnitud del efecto.

Es importante interpretar estos resultados con cierta precaución, especialmente en los bins de distancia donde el número de trayectorias es bajo. Dado que el Barrier Score se calcula como un cociente, sus valores pueden volverse muy sensibles o variables cuando el número de cruces observados (O) es pequeño. En algunos casos de este análisis, particularmente para ciertas barreras o en bins de distancia más largos, el número de cruces observados o esperados es inferior a 10. Si bien estos valores se incluyen por completitud, el análisis debe centrarse en los patrones generales y en aquellos bins donde la cantidad de datos es suficiente para garantizar una estimación más estable del efecto barrera.

El análisis del Barrier Score por bin de distancia ($B(d)$) para el tren Sarmiento y sus diferentes tramos, como se visualiza en la Figura 6.4, revela una dinámica compleja y dependiente de la escala del viaje.

1. **Sarmiento (completo):** Para la línea completa, el efecto barrera no es inmediato. En el primer bin de distancia (0-1 km), el $B(d)$ es ligeramente negativo (-0,06), lo que podría indicar que para viajes muy cortos, la presencia de estaciones actúa más como un punto de atracción que como una barrera. Sin embargo, a partir del bin de

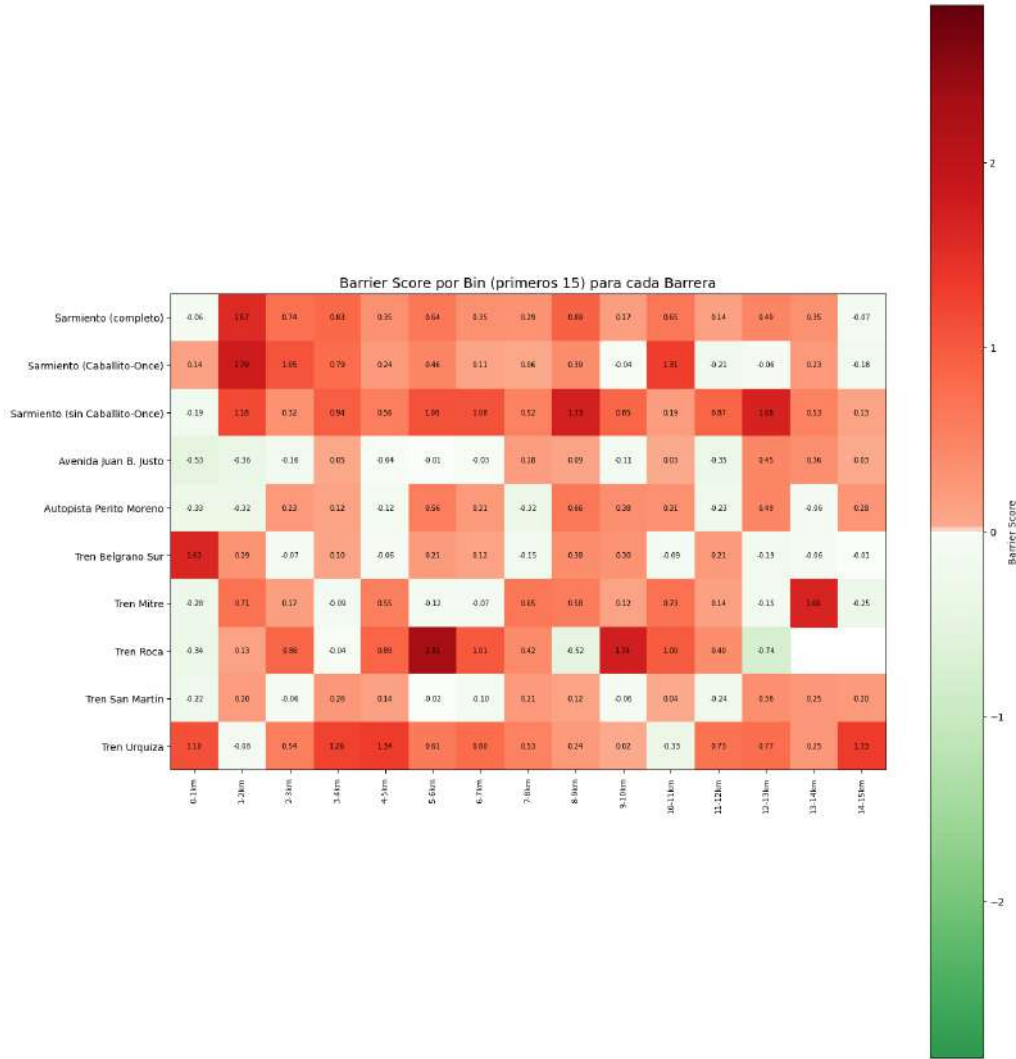


Fig. 6.4: Heatmap del Barrier Score por bin de distancia ($B(d)$) para las principales infraestructuras analizadas. Cada recuadro muestra el valor de $B(d)$ para una barrera (fila) en un bin de distancia de 1 km (columna). Los colores rojos indican un efecto barrera positivo, mientras que los colores verdes/azules indicarían un efecto de atracción (BS negativo).

1-2 km, el efecto barrera emerge con una fuerza notable ($B(d) = 1,57$), siendo este el pico de su impacto. A partir de allí, el efecto, si bien se mantiene consistentemente positivo en la mayoría de los bins, comienza a disminuir gradualmente a medida que aumenta la distancia de los viajes. Esto sugiere que el Sarmiento es un obstáculo especialmente significativo para los viajes de corta y media distancia (entre 1 y 8 km), mientras que su influencia relativa se diluye en trayectos más largos.

2. **Sarmiento (Caballito-Once):** Este tramo, que es el más céntrico, muestra un comportamiento similar al de la línea completa, pero con un pico de efecto barrera aún más pronunciado en el bin de 1-2 km ($B(d) = 1,79$). Esto lo consolida como el segmento con el mayor efecto disuasorio para viajes de corta distancia. La disminución posterior también es gradual, manteniendo valores positivos en casi todos los

bins analizados.

3. **Sarmiento (sin Caballito-Once):** Al analizar los tramos restantes de la línea (excluyendo la parte más céntrica), se observa un patrón diferente. El efecto barrera positivo aparece de forma significativa a partir del bin de 1-2 km, pero a diferencia de los otros tramos, su máximo valor se registra en un bin de distancia considerablemente mayor (9-10 km, con un $B(d)$ de **1,73**). Esto podría indicar que en estas zonas menos centrales, la barrera se percibe de manera distinta, y su máximo impacto se da en viajes de media-larga distancia. El valor también muy elevado en el bin 12-13 km ($B(d) = 1.68$) es notable y podría deberse a una combinación de pocos cruces observados y una mayor expectativa de cruces en el modelo nulo para ese rango de distancia específico.

En síntesis, el análisis por bin de distancia confirma que el tren Sarmiento podría estar actuando como una barrera robusta en casi todas las escalas de viaje. Su impacto es más drástico en los viajes de corta y media distancia, especialmente en su tramo céntrico. La dinámica del efecto varía a lo largo de su traza, sugiriendo que la interacción entre la infraestructura, el contexto urbano y las necesidades de movilidad de los estudiantes no es uniforme, sino que depende fuertemente tanto de la ubicación como de la longitud del desplazamiento.

6.5. Resultados preliminares: validación estadística del efecto barrera

El trabajo de Aiello et al. [19], que inspira nuestra metodología, propone realizar una prueba de significancia estadística (como un **test t**) para confirmar que la diferencia entre el número de cruces observados en los datos reales (O) y el número de cruces esperados por el modelo nulo (E) es estadísticamente significativa. Para llevar a cabo esta validación, es necesario generar no una, sino múltiples realizaciones del modelo nulo y comparar la distribución de los cruces esperados con el único valor de los cruces observados.

Para este fin, se generaron **100 realizaciones** independientes del modelo nulo. Sin embargo, este enfoque presenta un desafío computacional considerable. Reconstruir la ruta completa de cada una de las miles de trayectorias nuevas de cada modelo nulo buscando el camino más corto sobre la red de calles de CABA es un proceso muy costoso en términos de tiempo de cómputo.

Para superar esta limitación y poder realizar la validación estadística de manera eficiente, se propuso un criterio simplificado para determinar el cruce de una barrera. En lugar de calcular la intersección geométrica de cada trayectoria con la línea exacta del tren, se definieron dos grandes polígonos geográficos: uno que abarca la zona "Norte" de la vía del tren Sarmiento y otro que abarca la zona "Sur", como se muestra en la Figura 6.5. Con esta aproximación, se considera que una trayectoria casa-escuela cruza la barrera si su punto de origen se encuentra en un polígono y su punto de destino se encuentra en el otro (origen en Sur y destino en Norte, o viceversa).

Es importante notar que este método basado en polígonos puede ser más conservador que el de intersección directa, ya que podría no contar algunas trayectorias que cruzan la vía pero su origen o destino no caen sobre estos polígonos. Sin embargo, al capturar correctamente la gran mayoría de los cruces directos (un subconjunto representativo del fenómeno), permite realizar un análisis estadístico válido y computacionalmente factible sobre la diferencia entre los cruces observados y esperados.

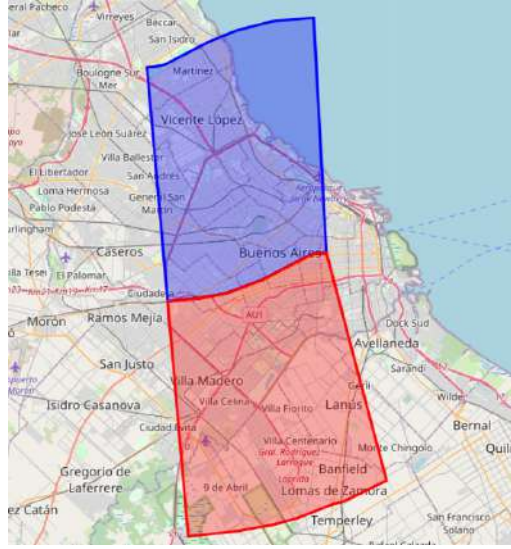


Fig. 6.5: Definición de polígonos Norte y Sur para la validación estadística del cruce del tren Sarmiento. Las trayectorias con origen en un polígono y destino en el otro se consideran como cruces.

Con esta metodología, se contabilizó el número de cruces para los datos reales y para cada una de las 100 realizaciones del modelo nulo. La Figura 6.6 presenta el histograma de la distribución del número de cruces obtenidos en los 100 modelos nulos, y lo compara con el número de cruces observados en el modelo real.

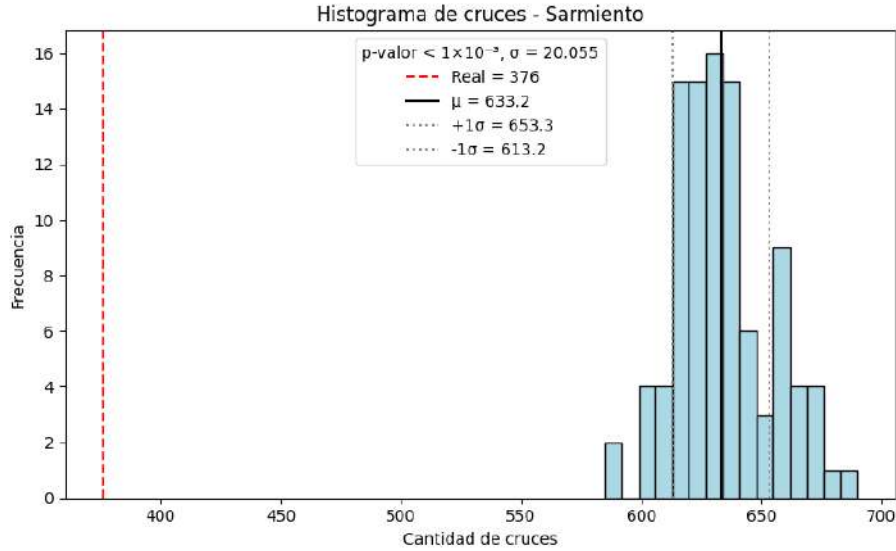


Fig. 6.6: Histograma del número de cruces del tren Sarmiento en 100 realizaciones del modelo nulo. La línea vertical roja indica el número de cruces observados en los datos reales. También se muestran la media y la desviación estándar de la distribución de los modelos nulos.

El análisis de esta distribución arroja los siguientes resultados:

- Número de Cruces Observados (Reales): $O = [376]$.

- Número de Cruces Esperados (Media de los modelos nulos): $\bar{E} = [633,2]$.
- Desviación Estándar de los Cruces Esperados: $\sigma_E = [20,055]$.

Como se observa visualmente en el histograma, el número de cruces reales (línea roja) se encuentra muy alejado de la distribución de los cruces esperados por los modelos nulos, ubicándose en la cola izquierda de la misma. Para formalizar esta observación, se realizó una **prueba de hipótesis** estadística.

La lógica de esta prueba es la siguiente: partimos de una **hipótesis nula (H_0)**, que es la suposición que queremos poner a prueba. En nuestro caso, la hipótesis nula establece que *"no hay un efecto barrera significativo y, por lo tanto, el número de cruces observados en la realidad (O) es simplemente un valor que podría haber sido generado por nuestro modelo nulo aleatorio"*. Es decir, H_0 postula que O proviene de la misma distribución de probabilidad que los valores de cruces esperados (E) que hemos generado.

Para evaluar esta hipótesis, calculamos el **p-valor** que representa la probabilidad de obtener un resultado tan extremo (o más extremo) que el observado en nuestros datos reales, asumiendo que la hipótesis nula es verdadera. En nuestro contexto, dado que observamos menos cruces de los esperados, calculamos la probabilidad de que, por puro azar, un modelo nulo generara un número de cruces igual o menor al que observamos en la realidad.

El resultado de esta prueba arrojó un p-valor extremadamente bajo ($p < 0,001$). Este valor tan pequeño significa que es extraordinariamente improbable (menos de 1 en 1000) haber observado tan pocos cruces si la hipótesis nula fuera cierta. Dado que esta probabilidad es inferior al umbral de significancia estadística estándar (usualmente $\alpha = 0,05$ o $\alpha = 0,01$), permitiría rechazar la hipótesis nula.

Al rechazar la H_0 , concluimos que el número de cruces observados no es una simple fluctuación aleatoria dentro de lo que el modelo nulo esperaría. En otras palabras, la diferencia entre el número de cruces reales y los esperados es estadísticamente significativa y no producto del azar. Este resultado preliminar podría validar la existencia de un efecto barrera medible y robusto asociado al tren Sarmiento, y da sustento a los análisis detallados que se presentan a continuación. Esperamos extender este test para los cálculos del BSD en ambas direcciones y para el $B(d)$ para cada bin de distancia y para todas las barreras; permitiendo descartar casos en los que no podemos rechazar la H_0 para esos cruces observados.

6.6. Síntesis de los hallazgos sobre el efecto barrera

El análisis de la movilidad escolar en transporte público en CABA y la aplicación de la métrica del Barrier Score han arrojado una serie de hallazgos significativos sobre el impacto de las infraestructuras lineales, y en particular del tren Sarmiento, como barreras urbanas.

En primer lugar, el análisis descriptivo de los cruces observados sobre el tren Sarmiento reveló que solo un porcentaje relativamente bajo (aproximadamente el 11,2 %) del total de las 5.425 trayectorias casa-escuela reconstruidas implican cruzar esta vía. Además, estos cruces no se distribuyen uniformemente a lo largo de la traza, sino que se concentran marcadamente en el tramo entre Caballito y Once (67 % de los cruces). Quizás el hallazgo descriptivo más llamativo fue la fuerte asimetría direccional de estos cruces observados: el 77,6 % de los estudiantes que cruzan el Sarmiento lo hacen desde la zona Sur hacia la zona

Norte, sugiriendo ya una diferencia fundamental en cómo se utiliza esta infraestructura según el origen del viaje.

La cuantificación del efecto barrera mediante el **Barrier Score (BS) global** situó al tren Sarmiento ($BS = 0,527$) como la infraestructura con el mayor impacto disuasorio para la movilidad escolar entre todas las analizadas, superando notablemente a avenidas principales como Juan B. Justo ($BS = 0,008$) y autopistas como la Perito Moreno ($BS = 0,141$). De manera destacada, se encontró que al excluir el tramo más permeable del Sarmiento (Caballito-Once), el BS para los segmentos restantes de esta línea se incrementaba a un valor aún mayor ($BS = 0,806$), indicando que gran parte de su extensión actúa como una barrera muy significativa.

El análisis del **Barrier Score Direccional (BSD)** profundizó la comprensión de estas dinámicas, revelando asimetrías cruciales. Para el tren Sarmiento, se confirmó que su efecto barrera es marcadamente más intenso para las trayectorias que se originan al Norte de la vía y se dirigen hacia el Sur. En su tramo principal (Caballito-Liniers), el BSD Norte→Sur alcanzó un valor de 1,046, indicando que se esperarían más del doble de cruces en esta dirección si no existiera el efecto restrictivo del tren. Este valor fue sustancialmente mayor que el BSD Sur→Norte (0,261) para el mismo tramo. Esta pronunciada asimetría sugiere que el tren Sarmiento limita de forma más severa la movilidad de los estudiantes que residen al norte de sus vías y que potencialmente buscarían acceder a establecimientos educativos en el sur. Se observó una dinámica similar para otras infraestructuras como el Tren Belgrano Sur aunque se observaron bastante menos datos por lo que el resultado es más sensible.

Finalmente, el análisis desagregado del **Barrier Score por bin de distancia (B(d))** añadió una capa de profundidad a estos hallazgos, mostrando que el efecto barrera no es constante, sino que depende de la longitud del viaje. Para el Ferrocarril Sarmiento, se observó que el impacto es más pronunciado para los viajes de corta y media distancia (entre 1 y 8 km), alcanzando su pico máximo en este rango, para luego disminuir gradualmente a medida que los viajes se hacen más largos. Esta dinámica confirma que la infraestructura afecta principalmente a la movilidad de proximidad, mientras que su influencia relativa se diluye en desplazamientos a mayor escala.

7. CONCLUSIONES Y TRABAJO A FUTURO

El presente trabajo de tesis se propuso investigar la existencia de segregación educativa en la Ciudad Autónoma de Buenos Aires y evaluar el rol del tren Sarmiento como una posible frontera urbana que condiciona la movilidad escolar. A través de un análisis multifacético que combinó datos de establecimientos educativos, registros masivos de movilidad en transporte público y la aplicación de la métrica del Barrier Score, se han alcanzado conclusiones significativas y se han abierto nuevas aristas para futuras investigaciones.

7.1. Conclusiones principales

Es fundamental iniciar esta sección de conclusiones reconociendo el alcance y las limitaciones del presente estudio. Los hallazgos que se discuten a continuación se basan en un conjunto específico de datos y en una serie de supuestos metodológicos que deben ser tenidos en cuenta. En primer lugar, el análisis se centra exclusivamente en la movilidad de estudiantes de nivel primario que utilizan el transporte público en CABA, y en base a datos de un día de noviembre de 2019. Las 5.425 trayectorias reconstruidas, si bien constituyen un conjunto de datos masivo para este tipo de análisis, representan una fracción del universo total de estudiantes que asistieron a la escuela en el día estudiado, aproximadamente un 3,64 % de los 148.747 alumnos matriculados a escuelas primarias en 2019. No contamos con información sobre los patrones de movilidad de aquellos que se desplazan caminando, en vehículo particular o en transporte escolar, quienes podrían tener una propensión diferente a cruzar las barreras analizadas.

En segundo lugar, la reconstrucción de las trayectorias casa-escuela a partir de los datos de la tarjeta SUBE es en sí misma una inferencia basada en una serie de supuestos sobre los patrones de viaje típicos, dada la granularidad de los datos disponibles. Aunque se ha buscado que estos criterios sean lo más realistas posible, no representan una observación directa del viaje completo.

Con estas importantes salvedades en mente, los resultados de la investigación sugieren que el tren Sarmiento se correlaciona con patrones de movilidad escolar que refuerzan la división territorial de la ciudad, actuando, para este subconjunto de la población, como una frontera efectiva. El análisis de las trayectorias reconstruidas muestra una tendencia marcada: los estudiantes que residen al sur del Sarmiento tienden a asistir a escuelas ubicadas en esa misma zona, y análogamente, aquellos que viven al norte suelen concurrir a establecimientos situados al norte de la vía. En consecuencia, la movilidad escolar de este grupo muestra una clara tendencia a evitar el cruce de la vía férrea, desarrollándose predominantemente dentro de los límites espaciales que esta impone.

Más allá de este efecto general, el estudio revela una asimetría crucial en el impacto del Sarmiento como barrera. Los datos del Barrier Score Direccional indican que el efecto disuasorio del tren es considerablemente mayor para las trayectorias que se originarían en el Norte y se dirigirían hacia el Sur, en comparación con la dirección opuesta. Esto sugiere que el tren Sarmiento interrumpe en mayor medida los potenciales viajes desde el norte hacia el sur. Las razones subyacentes a esta asimetría pueden ser complejas y multifactoriales. Por un lado, la mayor matrícula en las escuelas del sur podría indicar una menor disponibilidad de vacantes, obligando a estudiantes de esa zona a buscar alter-

nativas al norte. Por otro lado, también es plausible que existan factores socioeconómicos o percepciones que disuadan a las familias residentes en el norte de considerar escuelas al sur. Independientemente de la combinación precisa de causas, el resultado neto observado es que la barrera parece ser funcionalmente "más alta" en la dirección Norte-Sur para la movilidad escolar en transporte público.

Estos patrones de movilidad restringida, aunque limitados al grupo y modo de transporte analizados, cobran una relevancia particular cuando se superponen con la geografía educativa de la ciudad. Como se evidenció en el análisis exploratorio, existe una brecha norte-sur en cuanto a los indicadores de desempeño escolar. Al correlacionarse con una movilidad que tiende a contener a los estudiantes dentro de sus respectivas zonas, la presencia del tren Sarmiento podría ser un factor que contribuye a que los alumnos de distintas partes de la ciudad accedan a conjuntos de oportunidades educativas diferentes.

En síntesis, este trabajo aporta evidencia cuantitativa inicial sobre cómo una infraestructura de transporte puede estar alineada con patrones de segregación educativa en el espacio urbano y quizás profundizarlos. Si bien no se puede afirmar causalidad, los resultados describen un mecanismo plausible de fragmentación y sientan una base sólida para futuras investigaciones que puedan explorar estas dinámicas con otros modos de transporte y en otros contextos.

7.2. Trabajo a futuro y perspectivas

Los hallazgos de esta tesis abren múltiples líneas de investigación futura que podrían profundizar y expandir la comprensión de las barreras urbanas y sus consecuencias.

En primer lugar, el presente estudio se centró en la movilidad escolar realizada a través del transporte público. Sería sumamente valioso complementar este análisis investigando los patrones de cruce (o no cruce) del tren Sarmiento para aquellos estudiantes que se desplazan caminando o en vehículo particular (auto, transporte escolar). Para ello, se podría diseñar e implementar un plan de encuestas dirigidas a padres, madres o tutores en las entradas y salidas de las escuelas, buscando recabar información sobre los modos de transporte habituales y las percepciones sobre las barreras. Estos datos permitirían construir una imagen más completa de la movilidad escolar en su totalidad.

Paralelamente, para enriquecer la comprensión de por qué el Sarmiento (u otras infraestructuras) funcionan como barreras, sería beneficioso realizar observaciones de campo detalladas. Esto implicaría visitar puntos clave a lo largo de la traza ferroviaria para analizar la calidad y disponibilidad de los cruces peatonales y vehiculares (pasos a nivel, puentes, túneles), la frecuencia de interrupciones del servicio, la percepción de seguridad en el entorno de los cruces, y observar directamente los flujos de estudiantes en horarios de entrada y salida escolar. Este enfoque cualitativo podría revelar factores contextuales que los datos masivos no capturan.

Quizás la línea de trabajo futuro más prometedora, y la que reviste un mayor interés propio, es la extensión de esta metodología a otros contextos y escalas.

- **Otros grupos sociales y propósitos de viaje:** El marco del Barrier Score podría aplicarse para estudiar el impacto de las barreras en la movilidad de otros grupos poblacionales, como personas mayores, estudiantes de secundaria o universitarios, o incluso la población general. Se podrían analizar viajes con otros propósitos, como el acceso al trabajo, a servicios de salud, a espacios de ocio o culturales, identificando así barreras que afectan diversas dimensiones de la vida urbana cotidiana.

- **Otras ciudades y contextos urbanos:** Si se dispone de datos de movilidad y de la geografía de las infraestructuras, este enfoque es replicable en otras ciudades, tanto en Argentina como en otros países, permitiendo análisis comparativos y la identificación de patrones comunes o particularidades locales en la forma en que las barreras configuran la segregación.
- **Análisis temporal y evaluación de intervenciones:** La disponibilidad de datos de movilidad a lo largo del tiempo (meses o años) abriría la posibilidad de realizar análisis longitudinales del efecto barrera. Esto sería invaluable para evaluar el impacto de intervenciones urbanas, como la construcción de nuevos puentes, la mejora de pasos bajo nivel, o incluso la remoción de ciertas barreras. Se podría medir objetivamente si estas obras públicas lograron reducir el efecto barrera y mejorar la conectividad.

En última instancia, el objetivo fundamental que subyace a este tipo de investigación es el de identificar las fronteras físicas que fragmentan nuestras ciudades con el propósito de informar y promover intervenciones urbanas que las mitiguen o eliminen. Al comprender mejor cómo las infraestructuras modelan la movilidad y el acceso a oportunidades, podemos aspirar a generar herramientas útiles para quienes diseñan políticas y obras que fomenten una mayor integración socio-espacial, contribuyendo a la construcción de ciudades menos segregadas, más justas e inclusivas para todos sus habitantes.

BIBLIOGRAFÍA

- [1] Douglas S. Massey y Nancy A. Denton. «The Dimensions of Residential Segregation». En: *Social Forces* 67.2 (1988), págs. 281-315. ISSN: 0370-1573. DOI: 10.1093/sf/67.2.281. URL: <https://academic.oup.com/sf/article-pdf/67/2/281/6514769/67-2-281.pdf>.
- [2] David M. Cutler y Edward L. Glaeser. «Are Ghettos Good or Bad?» En: *The Quarterly Journal of Economics* 112.3 (ago. de 1997), págs. 827-872. ISSN: 0033-5533.
- [3] Francisco Sabatini. *La segregación social del espacio en las ciudades de América Latina*. No disponible en línea. Washington, DC: Banco Interamericano de Desarrollo, 2003.
- [4] Jane Jacobs. *The Death and Life of Great American Cities*. New York: Random House, 1961.
- [5] Markus Schläpfer et al. «The Scaling of Human Interactions with City Size». En: *Journal of The Royal Society Interface* 11.98 (2014), pág. 20130789. DOI: 10.1098/rsif.2013.0789.
- [6] Rick Grannis. «The Importance of Trivial Streets: Residential Streets and Residential Segregation». En: *American Journal of Sociology* 103.6 (1998), págs. 1530-1564. DOI: 10.1086/231400. URL: <https://www.journals.uchicago.edu/doi/pdf/10.1086/231400>.
- [7] Paulo R. Anciaes et al. «Urban Transport and Community Severance: Linking Research and Policy to Link People and Places». En: *Journal of Transport & Health* 3.3 (2016). Artículo open access bajo CC BY en ResearchGate, págs. 268-277. DOI: 10.1016/j.jth.2016.06.006. URL: https://www.researchgate.net/publication/305988095_Urban_transport_and_community_severance_Linking_research_and_policy_to_link_people_and_places.
- [8] Elizabeth Oltmans Ananat. «The Wrong Side(s) of the Tracks: The Causal Effects of Racial Segregation on Urban Poverty and Inequality». En: *American Economic Journal: Applied Economics* 3.2 (2011), págs. 34-66. DOI: 10.1257/app.3.2.34. URL: <https://www.aeaweb.org/articles?id=10.1257/app.3.2.34>.
- [9] Gergő Tóth et al. «Inequality is rising where social network segregation interacts with urban topology». En: *Nature Communications* 12.1 (2021). Acceso vía repositorio institucional (PDF open access), pág. 1143. DOI: 10.1038/s41467-021-21460-0. URL: <https://kti.krtk.hu/wp-content/uploads/2021/03/CERSIEWP202116-1.pdf>.
- [10] Elizabeth Roberto y Elizabeth Korver-Glenn. «The Spatial Structure and Local Experience of Residential Segregation». En: *Spatial Demography* 9.3 (2021), págs. 277-307. DOI: 10.1007/s40980-021-00090-x. URL: <https://link.springer.com/article/10.1007/s40980-021-00090-x>.
- [11] Douglas S. Massey y Nancy A. Denton. *American Apartheid: Segregation and the Making of the Underclass*. Cambridge, MA: Harvard University Press, 1993. ISBN: 9780674018211.

- [12] Sonia Arbaci. «(Re)Viewing Ethnic Residential Segregation in Southern European Cities: Housing and Urban Regimes as Mechanisms of Marginalisation». En: *Housing Studies* 23.4 (2008), págs. 589-613. DOI: 10.1080/02673030802117050. URL: <https://www.tandfonline.com/doi/full/10.1080/02673030802117050>.
- [13] Rémi Louf y Marc Barthelemy. «Patterns of Residential Segregation». En: *PLoS ONE* 11.6 (2016), e0157476. DOI: 10.1371/journal.pone.0157476. URL: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0157476>.
- [14] Pierre Bourdieu y Jean-Claude Passeron. *Los herederos: los estudiantes y la cultura*. París, Francia: Les Éditions de Minuit, 1964.
- [15] Juan Carlos Tedesco. «Crisis económica, educación y futuro en América Latina». En: *Nueva Sociedad* 84 (1986), págs. 80-90.
- [16] Argentinos por la Educación. *Indicadores educativos – Ciudad de Buenos Aires*. <https://argentinosporlaeducacion.org/indicadores/ciudad-de-buenos-aires/>. Último acceso: febrero de 2025. 2025.
- [17] Victoria Arcón e Inés Caridi. «Homogenization Scales in Residential Segregation Through Random Walkers». En: *Chaos, Solitons Fractals* 197 (ago. de 2025). Metrics based on SSRN preprint (DOI original: 10.2139/ssrn.4993130), pág. 116433. DOI: 10.1016/j.chaos.2025.116433. URL: <https://linkinghub.elsevier.com/retrieve/pii/S0960077925004461>.
- [18] Mateo Neira et al. «Urban segregation on multilayered transport networks: a random walk approach». En: *Scientific Reports* 14 (2024), pág. 8370. DOI: 10.1038/s41598-024-58932-9. URL: <https://doi.org/10.1038/s41598-024-58932-9>.
- [19] Luca Maria Aiello et al. «Urban highways are barriers to social ties». En: *Proceedings of the National Academy of Sciences* 122.10 (2025), e2408937122. DOI: 10.1073/pnas.2408937122. URL: <https://www.pnas.org/doi/10.1073/pnas.2408937122>.
- [20] Colin D. Buchanan. *Traffic in Towns: A Study of the Long Term Problems of Traffic in Urban Areas*. London: H. M. Stationery Office, 1963.
- [21] Richard Rothstein. *The Color of Law: A Forgotten History of How Our Government Segregated America*. Liveright Publishing, 2017.
- [22] Emily Badger. *How railroads, highways and other man-made lines racially divide America's cities*. Accessed: 2025-03-18. 2015. URL: <https://www.washingtonpost.com/news/wonk/wp/2015/07/16/how-railroads-highways-and-other-man-made-lines-racially-divide-americas-cities/>.
- [23] James Kanter. *Roads to nowhere: how infrastructure built on American inequality*. The Guardian. Accessed: 2025-03-18. 2018. URL: <https://www.theguardian.com/cities/2018/feb/21/roads-nowhere-infrastructure-american-inequality>.
- [24] Donald Appleyard. *Livable Streets*. Berkeley: University of California Press, 1981.
- [25] Salvador Rueda. «Superblocks for the Design of New Cities and Renovation of Existing Ones: Barcelona's Case». En: *Integrating Human Health into Urban and Transport Planning*. Ed. por Mark Nieuwenhuijsen y Haneen Khreis. Cham: Springer International Publishing, 2018, págs. 135-153.

-
- [26] Carlos Moreno et al. «Introducing the 15-Minute City: Sustainability, Resilience and Place Identity in Future Post-Pandemic Cities». En: *Smart Cities* 4.1 (2021), págs. 93-111. DOI: 10.3390/smartcities4010006. URL: <https://www.mdpi.com/2624-6511/4/1/6>.
- [27] Ministerio de Transporte de Argentina. *Líneas de trenes del Área Metropolitana de Buenos Aires (AMBA)*. Último acceso: febrero de 2025. 2024. URL: <https://datos.transporte.gob.ar/dataset/f87b93d4-ade2-44fc-a409-d3736ba9f3ba/resource/87fb1c41-89b5-42db-b45a-992672931b2d/download/rmba-ferrocarril-lineas.zip>.
- [28] B. H. Kemp. «Social Impact of a Highway on an Urban Community». En: *Highway Research Record* 75 (1965), págs. 92-102.
- [29] Donald Appleyard y Mark Lintell. «The Environmental Quality of City Streets: The Residents' Viewpoint». En: *Journal of the American Institute of Planners* 38.2 (1972), págs. 84-101. DOI: 10.1080/01944367208977410. URL: <https://doi.org/10.1080/01944367208977410>.
- [30] Yong-Li Xu et al. «Quantifying segregation in an integrated urban physical-social space». En: *Journal of The Royal Society Interface* 16.160 (2019), pág. 20190536. DOI: 10.1098/rsif.2019.0536. URL: <https://royalsocietypublishing.org/doi/10.1098/rsif.2019.0536>.
- [31] Esteban Moro et al. «Mobility patterns are associated with experienced income segregation in large US cities». En: *Nature Communications* 12.1 (2021), pág. 4633. DOI: 10.1038/s41467-021-24899-8. URL: <https://www.nature.com/articles/s41467-021-24899-8>.
- [32] Zhaoyang Fan et al. «Diversity beyond density: Experienced social mixing of urban streets». En: *PNAS Nexus* 2.4 (2023). PDF completo disponible en ResearchGate (CC BY-NC-ND 4.0), pgad077. DOI: 10.1093/pnasnexus/pgad077. URL: https://www.researchgate.net/publication/369148229_Diversity_beyond_density_experienced_social_mixing_of_urban_streets.
- [33] Z. Kallus et al. «Spatial fingerprints of community structure in human interaction network for an extensive set of large-scale regions». En: *PLoS One* 10.8 (2015), e0135404. DOI: 10.1371/journal.pone.0135404. URL: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0135404>.
- [34] Gergő Pintér y Balázs Lengyel. «Neighborhoods and Boundaries of Urban Mobility». En: *arXiv preprint* (2023). DOI: 10.48550/arXiv.2312.11343. URL: <https://arxiv.org/abs/2312.11343>.
- [35] Banco Interamericano de Desarrollo. *Matriz Origen-Destino Transporte Público*. GitHub repository. Accessed: 2025-03-18. 2025. URL: <https://github.com/EL-BID/Matriz-Origen-Destino-Transporte-Publico>.
- [36] Felipe Gonzalez y Sebastián Anapolsky. *Identificando la desigualdad en los patrones de movilidad en transporte público*. Ed. por Julieta Abad y Cristian Moleres. Editorial y lugar de publicación no especificados. 2025.
- [37] Fengli Xu et al. «Using Human Mobility Data to Quantify Experienced Urban Inequalities». En: *Nature Human Behaviour* (2025). DOI: 10.1038/s41562-024-02079-0. URL: <https://www.nature.com/articles/s41562-024-02079-0>.

- [38] Sandro Sousa y Vincenzo Nicosia. «Quantifying ethnic segregation in cities through random walks». En: *Nature Communications* 13.1 (2022), pág. 5809. DOI: 10.1038/s41467-022-33344-3. URL: <https://doi.org/10.1038/s41467-022-33344-3>.
- [39] Luciana Bottinelli. «Educación y desigualdad. Un repaso por algunos aportes de la sociología de la educación en la Argentina». En: *Revista Sociedad* 37.1 (2017), págs. 95-111.
- [40] William Julius Wilson. *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy*. Chicago, IL: University of Chicago Press, 1987. ISBN: 978-0226041212.
- [41] Sean F. Reardon y Ann Owens. «60 Years After Brown: Trends and Consequences of School Segregation». En: *Annual Review of Sociology* 40 (2014). Accedido: 2025-06-29, págs. 199-218. DOI: 10.1146/annurev-soc-071913-043152. URL: <https://www.annualreviews.org/doi/10.1146/annurev-soc-071913-043152>.
- [42] Pierre Bourdieu y Jean-Claude Passeron. *Reproduction in Education, Society and Culture*. London: Sage Publications, 1977. ISBN: 978-0803984550.
- [43] Elizabeth Roberto y Jackelyn Hwang. «Barriers to Integration: Physical Boundaries and the Spatial Structure of Residential Segregation». En: *arXiv preprint* (2017). Preprint publicado originalmente en 2015; este trabajo destaca cómo “physical barriers—features of the urban environment that disconnect locations—such as highways, railroad tracks, and dead-end streets divide urban space in ways that reinforce or exacerbate segregation.” DOI: 10.48550/arXiv.1509.02574. URL: <https://arxiv.org/abs/1509.02574>.
- [44] Hugo Barbosa et al. «Human mobility: Models and applications». En: *Physics Reports* 734 (2018), págs. 1-74. DOI: 10.1016/j.physrep.2018.01.001. URL: <https://doi.org/10.1016/j.physrep.2018.01.001>.
- [45] Zhaomiao Zhang et al. «A review of destination estimation for entry-only automated fare collection (AFC) data». En: *Transport Reviews* 43.2 (2023). Solicitar PDF mediante 'Request full-text' en ResearchGate, págs. 157-182. DOI: 10.1080/01441647.2022.2133965. URL: https://www.researchgate.net/publication/373194332_Zhang_et_al_2023.
- [46] Christopher Kuner y Massimo Marelli, eds. *Handbook on Data Protection in Humanitarian Action*. Accessed: 2025-06-29; PDF disponible directamente desde ICRC. International Committee of the Red Cross (ICRC), 2017. URL: <https://www.icrc.org/en/download/file/18145/data-prot-humanitarian-action.pdf>.
- [47] European Union Agency for Cybersecurity (ENISA). *Handbook on Security of Personal Data Processing*. Luxembourg: Publications Office of the European Union, 2018. URL: <https://www.enisa.europa.eu/publications/handbook-on-security-of-personal-data-processing>.
- [48] Gobierno de la Ciudad de Buenos Aires. *Escuelas públicas y privadas de CABA*. Último acceso: febrero de 2025. 2024. URL: https://cdn.buenosaires.gob.ar/datosabiertos/datasets/ministerio-de-educacion/establecimientos-educativos/establecimientos_educativos_WGS84.geojson.
- [49] Leo Breiman. «Random Forests». En: *Machine Learning* 45.1 (2001), págs. 5-32. DOI: 10.1023/A:1010933404324.

-
- [50] Stuart Lloyd. «Least Squares Quantization in PCM». En: *IEEE Transactions on Information Theory* 28.2 (1982), págs. 129-137. DOI: 10.1109/TIT.1982.1056489.
- [51] Gustavo Mera. «¿Cómo afecta la localización a las condiciones habitacionales? Un análisis espacial para la Región Metropolitana de Buenos Aires». En: *Revista Iberoamericana de Urbanismo* 17 (2022). Accessed: 2025-06-29, págs. 121-145. DOI: 10.37230/riur.2022.n17.8. URL: https://www.scielo.org.ar/scielo.php?script=sci_arttext&pid=S1853-36552022000100008.
- [52] Geoff Boeing. «OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks». En: *Computers, Environment and Urban Systems* 65 (2017). PDF completo disponible en ResearchGate (enlace abierto), págs. 126-139. DOI: 10.1016/j.compenvurbsys.2017.05.004. URL: https://www.researchgate.net/publication/316444466_OSMnx_A_Python_package_to_work_with_graph-theoretic_OpenStreetMap_street_networks.
- [53] Gina Ross. «Active Transport, Physical Activity, and Distance Between Home and School». En: *Journal of Physical Activity and Health* 13.4 (2016), págs. 447-456. DOI: 10.1123/jpah.2015-0307.