Mind the (Data) Gap: A Repositioning Engine for Spatiotemporal Mismatch

M. Shehryar Khan
Virginia Tech
Arlington, VA
shehryarkhan@vt.edu

Sanjna Kumari
Virginia Tech
Arlington, VA
sanjnak@vt.edu

Katelyn Crumpacker
Virginia Tech
Blacksburg, VA
katelync22@vt.edu

ABSTRACT

Bikeshare systems are widely deployed in U.S. cities, yet their spatial configuration often fails to align with actual mobility demand. In this study, we analyze pedestrian and bikeshare activity in New York City and Washington, DC, using sparse manual counts and continuous sensor data, respectively. This contrast is intentional and reflects real-world constraints in urban data availability. We address three questions: (1) how pedestrian activity correlates with bikeshare usage, (2) where high-pedestrian but low-bikeshare conditions reveal latent demand, and (3) which pedestrian-dense areas are poorly connected to bikeshare networks. To operationalize these insights, we introduce a pedestrian-informed rebalancing engine that reallocates bikeshare capacity using multimodal gap and accessibility metrics. Across both cities, our approach reduces supply-demand mismatch without increasing system-wide inequality, highlighting the value of pedestrian data for equitable bikeshare planning. We observe improvements in both under-supplied and over-supplied areas in NYC and DC, demonstrating that the repositioning engine remains robust under limited data conditions.

1 INTRODUCTION

Bikeshare systems have become a core component of multimodal transportation networks in large U.S. cities, offering a flexible and relatively low-cost alternative to automobile travel [4]. Despite their growth, substantial evidence shows that bikeshare networks often fail to provide equitable access. Station placement has been shown to disproportionately favor higher-income and centrally located neighborhoods, limiting access in historically underserved areas [2, 3, 8]. These patterns undermine the potential of bikeshare systems to support inclusive urban mobility.

Beyond equity concerns, bikeshare networks frequently exhibit structural misalignment with underlying human mobility patterns [5, 9]. Pedestrian activity, a direct signal of localized movement and potential short-trip demand, is rarely integrated into bikeshare planning decisions. As a result, stations are often deployed based on installation feasibility or incremental expansion rather than on where people already travel, leading to inefficiencies in coverage and utilization.

1.1 Problem Statement

The disconnect between pedestrian movement, bikeshare usage, and network coverage highlights the need for analytical frameworks that incorporate pedestrian mobility into bikeshare planning workflows. Areas with high pedestrian volumes but low bikeshare usage may reflect latent demand driven by accessibility constraints. However, planners lack systematic tools to identify these mismatches across neighborhoods or to operationalize them within bikeshare repositioning strategies, motivating work toward a Repositioning Engine Algorithm.

Data heterogeneity further complicates these challenges. Cities such as New York City have sparse manual pedestrian counts available publicly, while Washington, DC benefits from continuous automated sensing. Methods that fail to accommodate these differences risk reinforcing existing inequities and misallocations in bikeshare systems.

To address these gaps, we conduct a comparative analysis of New York City (NYC) and Washington, DC (DC), integrating pedestrian and bikeshare datasets to evaluate alignment between human activity and shared mobility infrastructure. We examined:

(1) how pedestrian traffic intensity correlates with bikeshare usage across time and space;
(2) which high-pedestrian traffic zones exhibit disproportionately low bikeshare activity, indicating latent demand; and
(3) which pedestrian-dense areas are weakly connected to the bikeshare network, forming accessibility gaps.

1.2 Contributions

To bridge empirical findings with actionable mobility planning, we introduce a *Repositioning Engine* algorithm, a data-adaptive reallocation technique that constructs multimodal mobility profiles, calculates supply demand gap scores, and proposes optimized bike redistribution strategies under operational constraints. By applying this framework across two cities with contrasting mobility ecologies, we demonstrate how pedestrian data can be operationalized to improve bikeshare coverage, efficiency, and equity. Furthermore, it is built with robustness in mind, to help tackle both sparse snapshot datasets (NYC), and healthier, continual datasets (DC).

Overall, this paper contributes a generalizable methodological framework for diagnosing misalignment in urban bikeshare systems and provides evidence-based strategies for more equitable and demand-responsive mobility planning.

2 RELATED WORK

In the previous literature, there are a lot of studies that look at the pedestrian movement or the bikeshare networks in big cities separately. However, the gap in the research is the use of pedestrian data to inform decisions like bikeshare network design, expansion, or rebalancing decisions.

2.1 Bikeshare system and Pedestrian Mobility: Prior Analytical Approaches

Several studies analyze and visualize city-scale bikeshare data in order to identify different patterns to create better bike sharing services. Oliveira et. al. [6] creates a timeline combined with a map for New York citi bike sharing networks. They suggest this data visualization tool can be used for dealing with or forecasting re-balancing after system outages. In the pedestrian mobility space, papers look into modeling and analyzing pedestrian data in order to understand people's movement. Angel et. Plaut [1] performs tempo-spatial analysis of pedestrian movement over the years. They use this analysis to comment on what influences people to walk and how Covid-19 decreased the volume and changed walking times.

2.2 Algorithmic Rebalancing Engines

In Schuijbroek et. al. [7], they created a algorithmic engine with the goal of rebalancing and routing bikes in bikeshare systems. They use inventory balancing, priority measures, and redistribution to work with mismatches in supply and demand without globally optimizing. Similarly, our algorithm does not globally optimize but looks at local imbalance metrics, while maintaining robustness against limited data availability. However, unlike Schuijbroek et al., we do not optimize routing but instead prioritize robustness to sparse sensing.

3 METHODOLOGY

To systematically evaluate the alignment between urban mobility patterns and bikeshare infrastructure, we developed a spatiotemporal analysis framework driven by three core research questions. The methodology was designed to be robust across varying data granularities, accommodating both continuous monitoring systems (Washington DC) and sparse manual count snapshots (New York City).

3.1 Research Questions & Analysis

3.1.1 RQ1: Correlation of Mobility Modalities. To answer *"How well does pedestrian traffic intensity correlate with bikeshare usage?"*, we first established a spatial coupling between heterogeneous data sources. Let $S_P = \{p_1, ..., p_n\}$ represent pedestrian sensor locations and $S_B = \{b_1, ..., b_m\}$ represent bikeshare stations.

For every pedestrian sensor p_i, we aggregated bikeshare activity A_{bike} from all stations within a distance threshold θ (set to 400m-800m depending on network density):

$$A_{bike}(p_i, t) = \sum_{j \in S_B} \mathbb{I}(dist(p_i, b_j) < \theta) \cdot (Starts_{j,t} + Ends_{j,t}) \tag{1}$$

where $Starts$ and $Ends$ represent trip volumes during time interval t.

To measure the strength of the relationship, we employed both parametric and non-parametric correlation coefficients:

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}, \quad \rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \tag{2}$$

where x and y represent normalized pedestrian and bikeshare counts, and d_i represents the difference in paired ranks.

Temporal Lag Analysis (Continuous Data Only): For the high-resolution DC dataset, we extended this analysis to test the "First Mile" hypothesis, specifically, whether pedestrian spikes at time t predict bikeshare spikes at time $t + k$. We calculated the time-lagged cross-correlation:

$$r_{lag}(k) = \text{Corr}(C_{ped}(t), A_{bike}(t + k)) \tag{3}$$

A peak correlation at $k > 0$ implies a causal direction where walking precedes riding.

3.1.2 RQ2: Identification of Latent Demand (The Gap Analysis). To answer *"Which high-traffic zones show low bikeshare activity?"*, we operationalized the concept of "Latent Demand." We defined an Accessibility Gap Score (S_{gap}) based on the divergence of standardized Z-scores for pedestrian density (Z_{ped}) and bikeshare activity (Z_{bike}):

$$S_{gap}(i) = Z_{ped}(i) - Z_{bike}(i) \tag{4}$$

Using this metric, we classified locations into two distinct structural categories:

(1) **Ghost Zones ($S_{gap} \gg 0$):** Areas where pedestrian demand significantly outstrips bikeshare supply, indicating potential for network expansion.
(2) **Workhorse Nodes ($S_{gap} \ll 0$):** Areas where bikeshare usage is disproportionately high relative to foot traffic, acting as critical network anchors.

To filter for statistically significant mismatches, we applied a **90/30 Threshold Rule**:

$$\text{Latent Demand} \iff (C_{ped} > P_{90}) \land (A_{bike} < P_{30}) \tag{5}$$

where P_{90} and P_{30} denote the 90th and 30th percentiles of the respective distributions.

Trip Duration Validation: To ensure low bikeshare usage was not due to walkable distances, we introduce **"The Cafe Defense"**, where we analyzed trip durations at target locations. Low usage was not correlated with ultra-short trips (< 5 mins), confirming accessibility friction rather than destination proximity as the primary barrier.

3.1.3 RQ3: Network Connectivity Assessment. To identify poorly connected pedestrian-dense areas, we computed the distance from each sensor p_i to the nearest bikeshare station:

$$d_{min}(p_i) = \min_j \|p_i - b_j\|_2 \tag{6}$$

Connectivity Deserts were defined as top-quartile pedestrian locations with $d_{min} > 800m$.

3.1.4 Handling Data Heterogeneity. The datasets required distinct preprocessing:

- **NYC (Sparse):** Bi-annual manual counts aggregated into broad windows (AM, Midday, PM).
- **DC (Continuous):** Hourly automated counts enabled 24-hour clustering and time-lagged causality tests, with optional aggregation to match NYC's temporal resolution.

3.2 Algorithmic Repositioning Engine

To translate the spatiotemporal insights from our analysis into actionable planning decisions, we developed the *Repositioning Engine*. This data-adaptive system is designed to ingest standardized mobility data (from either sparse snapshots or continuous streams) and output a spatially optimized "Reallocation Vector" that minimizes the network alignment gap.

The engine operates through a pipeline that combines *location profiling*, *priority scoring*, and *budget-constrained reallocation*.

3.2.1 Location Profiling and Scoring. The first stage constructs a "Stability Profile" for every location-period pair. While the raw S_{gap} (defined in Eq. 4) indicates the *direction* of misalignment, it does not capture the *magnitude* of the opportunity. A gap of +2.0 at a quiet intersection is less critical than a gap of +1.5 at a major transit hub.

To address this, we defined two priority scores that weigh the gap severity by the log-transformed volume of activity, U_i defining the under-utilization score, and O_i defining the over-utilization score:

$$U_i = \max(S_{gap}(i), 0) \cdot P(HighGap|i) \cdot \ln(1 + C_{ped}(i)) \tag{7}$$

$$O_i = \max(-S_{gap}(i), 0) \cdot P(LowGap|i) \cdot \ln(1 + A_{bike}(i)) \tag{8}$$

where $P(HighGap|i)$ represents the frequency with which location i is flagged as underserved across temporal windows. The logarithmic term $\ln(1 + C)$ ensures that high-volume hubs are prioritized without allowing outliers to completely dominate the optimization landscape.

3.2.2 Reallocation Logic. The core function of the engine is to compute a reallocation vector $\vec{\Delta}$ that shifts capacity from oversupplied nodes to undersupplied ones while maintaining the system's total resource constraint (conservation of bicycles).

We define a "Reallocation Pool" (R_{pool}) as a fraction α of the total system fleet (experimentally set to $\alpha = 0.10$ or 10%):

$$R_{pool} = \alpha \sum_i A_{bike}(i) \tag{9}$$

The algorithm distributes this pool probabilistically based on the normalized weights of the priority scores. The net change $\Delta_{bike}(i)$ for any location i is calculated as:

$$\Delta_{bike}(i) = R_{pool} \cdot \left(\frac{\text{Score}_{under}(i)}{\sum \text{Score}_{under}} - \frac{\text{Score}_{over}(i)}{\sum \text{Score}_{over}} \right) \tag{10}$$

This formula ensures that locations with high unmet pedestrian demand receive a net influx of resources, while the "Workhorse" locations contribute to the pool proportional to their excess capacity.

3.2.3 Evaluation Metrics (The "Before vs. After"). To quantify the engine's effectiveness, we introduced a "Misallocation Index" (M) and utilized the Gini Coefficient to measure equity. The goal is to minimize M, defined as the mean absolute deviation of the raw supply-demand gap:

$$M = \frac{1}{N} \sum_{i=1}^{N} |C_{ped}(i) - \beta \cdot A_{bike}(i)| \tag{11}$$

where β is a scaling factor normalizing bikeshare trips to pedestrian counts. A reduction in M post-reallocation indicates a system-wide improvement in alignment.

Algorithm 1 Greedy Reallocation Heuristic

Require: Enriched Data D, Pool Fraction α
Step 1: Profiling
for each location i **do**
  Calculate $S_{gap}(i)$ using Z-scores
  Compute U_i and O_i
end for
Step 2: Normalization
$W_{under} \leftarrow \sum U_i$
$W_{over} \leftarrow \sum O_i$
$Pool \leftarrow \alpha \cdot \sum BikeCount$
Step 3: Reallocation
for each location i **do**
  $\delta_{in} \leftarrow Pool \cdot (U_i / W_{under})$
  $\delta_{out} \leftarrow Pool \cdot (O_i / W_{over})$
  $NewBikeCount(i) \leftarrow BikeCount(i) + \delta_{in} - \delta_{out}$
end for
Step 4: Validation
return $NewBikeCount$, MisallocationIndex($NewBikeCount$)

4 DATASETS

To evaluate the robustness of our algorithmic engine across different urban sensing paradigms, we selected two distinct datasets representing opposite ends of the "data availability" spectrum: New York City (sparse, manual snapshots) and Washington, D.C. (continuous, automated monitoring). We restricted our analysis to data from 2022 to 2024 for both cities. This temporal window was selected due to data availability: earlier NYC pedestrian records contain observable gaps (see Figure 2).

4.1 New York City

For New York City, we utilized the *NYC Bi-Annual Pedestrian Counts* alongside *CitiBike System Data*. This pedestrian dataset is derived from manual counts conducted only twice per year (May and October) during specific peak windows. Figure 2 shows the true sparsity of NYC data, including a larger gap in 2020.

This dataset presents a significant "real-world" challenge common in urban planning: **Temporal Sparsity**. Unlike automated systems, these snapshots capture only a fraction of annual activity, introducing noise from weather or daily anomalies. We treat this limitation as a key feature of our experimental design. By testing our engine on this sparse data, we validate its ability to function in resource-constrained environments where continuous sensor networks are unavailable.

Although the spread of pedestrian counters can still be improved (as observed in Figure 1), placement of these counters is not the chief problem; the snapshot, bi-annual data readings continue to be the main challenge here.

To address this, we harmonized the continuous CitiBike data to match the coarse resolution of the pedestrian surveys, aggregating trip starts/ends strictly within the defined AM (07:00-10:00), Midday (11:00-14:00), and PM (16:00-20:00) observation windows.

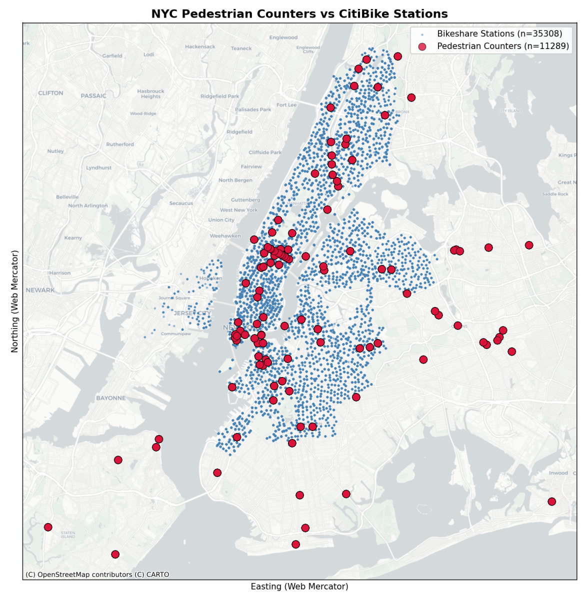

Figure 1: Spatial coverage of NYC Pedestrian Counters (dots) relative to CitiBike Stations.

4.2 Washington D.C.

For Washington D.C., we utilized the *DC Automated Bicycle and Pedestrian Counts* alongside *Capital Bikeshare* data. In contrast to NYC, this network consists of automated sensors providing continuous hourly resolution.

This dataset serves as our "Ground Truth" benchmark. The high temporal fidelity allows us to validate the engine's core assumptions, specifically the correlation lag analysis (RQ1), without the noise introduced by manual sampling.

We also observe the spatial coverage of the DC counters relative to the Capital Bikeshare Stations in Figure 3.

4.2.1 Dataset Comparison and Harmonization. The stark difference in data volume between the two cities is visualized in Figure 4. While DC provides a consistent stream of mobility signals, NYC provides high-intensity "bursts" of data at high-traffic intersections.

Our preprocessing pipeline standardizes these inputs into a unified *Location-Period-Intensity* vector, enabling the *Repositioning Engine* to process both the sparse NYC snapshots

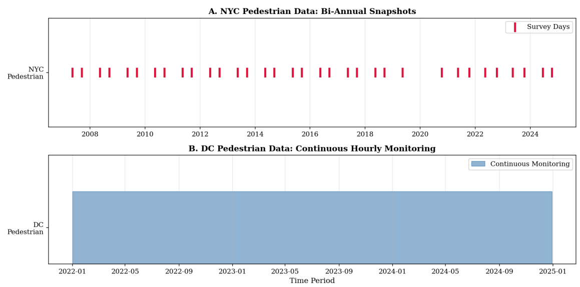

Figure 2: Temporal Resolution Comparison: The density of DC's continuous monitoring vs. NYC's sparse snapshot approach .

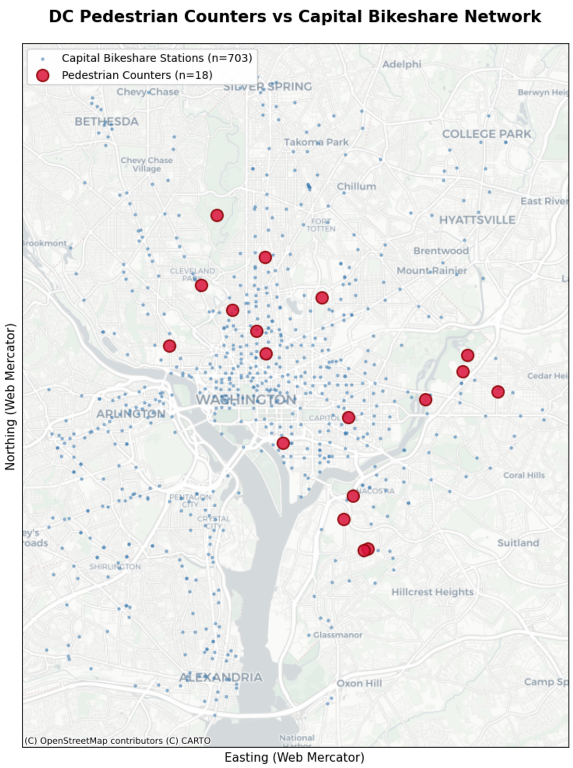

Figure 3: Spatial coverage of DC Pedestrian Counters (dots) relative to Capital BikeShare Stations.

and the dense DC streams without modification. This dual-city approach ensures our findings are not artifacts of a specific data collection method.

5 RESULTS

This section evaluates the three research questions (RQ1–RQ3) and the performance of the proposed repositioning engine. Results are reported separately for the two cities—New York City (NYC) and Washington, DC (DC), reflecting their contrasting data regimes (sparse peak-period counts versus continuous hourly sensing).

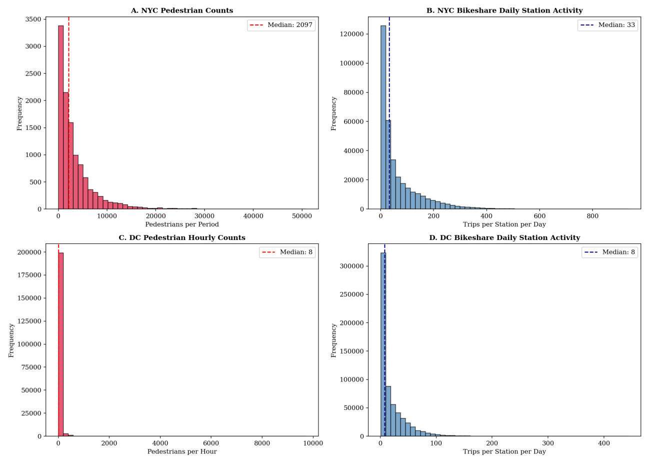

Figure 4: Scale of Observations: Comparison of total synchronized sensor-hours processed for each city.

5.1 RQ1: Relationship Between Pedestrian Traffic and Bikeshare Usage

NYC. NYC provides a sparse but high-quality snapshot of pedestrian activity. After processing 114 pedestrian counter locations, the final dataset contains **11,289 matched ped-bike observations**. Pedestrian volumes are highly skewed (**skewness = 2.79**), with a **median of 2,097** and extreme locations exceeding **50,000**, confirming a small number of pedestrian *super-hubs* dominate traffic.

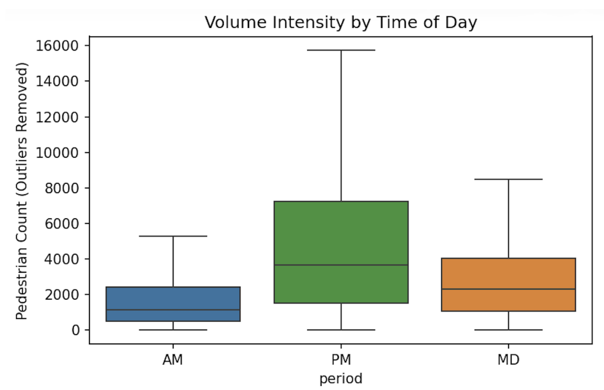

Figure 5: NYC pedestrian counts by time of day (AM/MD/PM).

Correlation patterns are positive but heterogeneous. Period-specific relationships demonstrate that PM activity is most predictive of bikeshare turnover, followed by MD and AM (Fig. 5). Structural mismatches identify locations with persistent supply-demand imbalance, such as *Fifth Ave.* and *W 34th St.* locations (see NYC entries in Table 1).

DC. The DC dataset provides a fully continuous 24-hour sensor view, comprising **202,963 pedestrian-hours across 18 automated counters** and **6,092,547 station-hours** of

Table 1: Representative Ghost Zones Identified via Z-Score Structural Mismatch

Location	Period	Ped	Bike	Gap
NYC Ghost Zones				
Fifth Ave	MD	13,620	22,092	3.26
Fifth Ave	PM	15,073	40,436	1.72
Eighth Ave	MD	12,072	29,039	2.13
W 34th St	AM	7,886	36,496	1.89
W 34th St	PM	20,727	71,751	1.72
DC Ghost Zones				
Anacostia River Tr., 11th St	AM	523	3.1	6.11
Anacostia River Tr., 11th St	MD	513	3.3	4.63
Oxon Run Park E. Bank	MD	355	1.3	3.29
Met Branch Tr.	AM	185	219	1.51

bikeshare activity. Spatial coverage is strong: every pedestrian counter lies within **800 m** of at least one bikeshare station.

After spatial and temporal matching, the dataset contains **137,746 hourly ped-bike observations**. Accessibility analysis shows that most counters are located close to the bikeshare network (**mean distance = 209 m; minimum = 18 m**), though a small number of trail and park locations exhibit substantially weaker access, with distances approaching **1 km** (e.g., Kenilworth Park and Oxon Run).

Segmented correlations reveal a striking contrast with NYC: all three periods (AM/MD/PM) show **weakly negative** Pearson correlations ($r = -0.07$ to -0.15; all $p < 0.001$). This reflects DC's trail-heavy network: high pedestrian flows on greenways correlate poorly with bikeshare demand, which is more commuter-oriented.

Overall, RQ1 shows that: (1) NYC's compact grid produces strong colocated pedestrian-bike usage, (2) DC's mixed commuting-recreational character weakens global coupling, highlighting the need for city-specific engine calibration.

5.2 RQ2: Accessibility Gaps and Latent Demand

NYC. NYC's dense bikeshare network and sparse pedestrian sampling result in few high-pedestrian/low-bikeshare anomalies. Applying the 90th/30th percentile rule identifies only **two** latent-demand observations. However, visual inspection (Fig. 6) reveals clear spatial clusters, particularly areas where pedestrian counters sit at the margins of the CitiBike footprint.

DC. The DC system exhibits substantial latent demand. Using the same 90th/30th protocol, **2,995 hourly observations** are flagged as latent demand. As shown in Table 2, these cluster heavily around three trail corridors: Rock Creek Trail @ Peirce Mill dominates with **1,182 flagged hours**,

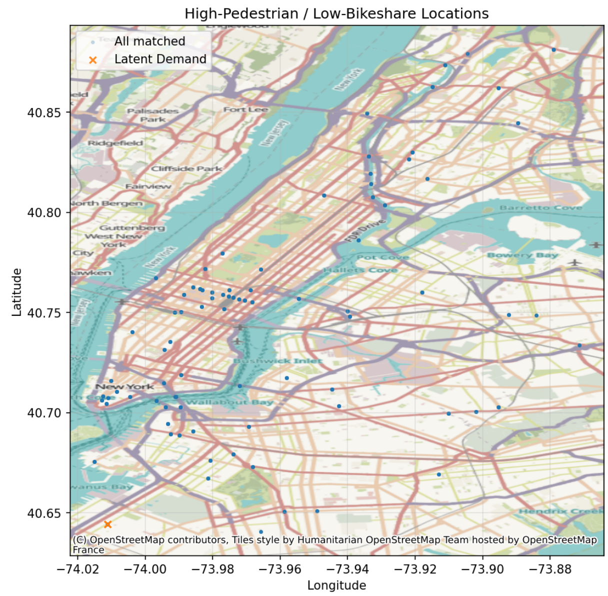

Figure 6: NYC latent-demand points relative to all matched observations.

followed by multiple segments of the Anacostia River Trail system.

These locations illustrate cases where pedestrian volume is high but bikeshare accessibility is limited, either due to station distance or insufficient station capacity.

Table 2: Top DC Latent-Demand Locations

Location	Latent-Demand Hrs.
Rock Creek Trail @ Peirce Mill	1,182
Anacostia River Trail, 11th St	475
Rose Park Trail @ P Street NW	454
Anacostia River Trail, Deane Ave	439
Anacostia River Trail, Benning	179

We also use the DC dataset to showcase *The Cafe Defense*, showing that only **19.3% of DC bikeshare trips are under 5 minutes**, implying that lack of usage is driven not by "too-short-to-bike" trips, but instead by accessibility and station placement.

5.3 RQ3: Network Connectivity and Accessibility

DC. DC exhibits significant heterogeneity in its connectivity profile. Accessibility analysis shows counters on trail systems (e.g., Kenilworth Park, Oxon Run) located hundreds of meters from the nearest station, and with few reachable stations within a 10-minute ride. These locations dominate both the latent demand counts and the top "ghost zones" in structural mismatch ranking (see DC entries in Table 1).

These results make clear that spatial connectivity, not raw demand, is the primary determinant of DC bikeshare usage at pedestrian-dense locations.

NYC. NYC accessibility gaps are less structural and more tied to land-use asymmetry. We notice some locations repeatedly appear as ghost zones, where exceptionally high pedestrian flows occur near stations that are either overburdened or underutilized.

The gap analysis (Table 1) confirms that NYC contains both high-ped/low-bike gaps and hyper-efficient "workhorse" stations.

5.4 Engine Performance: Repositioning Algorithm

We consolidate all results of our proposed repositioning engine in Table 3. The algorithm consistently improves alignment between pedestrian demand and bikeshare supply across both cities. Severe undersupply is reduced by **41.8% in NYC** and **36.0% in DC**, indicating successful reallocation away from critical shortage zones without expanding the overall system footprint.

The Misallocation Index decreases in both cities (**-193 in NYC, -7.2 in DC**), demonstrating improved demand alignment. Importantly, these gains are achieved without increasing inequality: the Bike Gini coefficient improves slightly in NYC (**0.543 → 0.530**) and remains stable in DC, showing that pedestrian-informed reallocation can reduce supply-demand mismatch while preserving system-wide equity.

6 DISCUSSION

6.1 Implications for Urban Planning

This study demonstrates that pedestrian activity is a critical but underutilized signal for evaluating and improving bikeshare networks. Results from RQ1 show that pedestrian-bikeshare coupling is highly context dependent: NYC exhibits moderate alignment due to its dense, mixed-use grid, while DC shows weak or negative correlations driven by trail-based pedestrian traffic that is poorly served by existing bikeshare infrastructure. These findings caution against assuming that pedestrian volume alone predicts bikeshare usage without accounting for spatial context and network structure.

The identification of latent-demand zones and ghost zones (RQ2) provides actionable insights for targeted intervention. In DC, persistent high-pedestrian/low-bikeshare clusters along trail corridors indicate clear opportunities for station infill or capacity expansion. In NYC, fewer latent-demand observations suggest that inefficiencies arise primarily from congestion and misallocation rather than absence of infrastructure. Connectivity analysis (RQ3) further highlights that accessibility gaps, even on the order of a few hundred meters, can significantly suppress bikeshare usage in pedestrian-dense areas.

The repositioning engine operationalizes these insights by reallocating capacity based on structural mismatch rather than historical usage alone. Across both cities, the engine reduces misallocation and severe undersupply, demonstrating the practical value of pedestrian-informed planning as a complement to physical network expansion.

6.2 Limitations and Ethical Considerations

NYC's temporally sparse pedestrian data restricts short-term inference, while DC's automated counters have limited spatial coverage. Additionally, distance-based accessibility measures do not fully capture real-world walking constraints such as barriers or perceived safety.

From an equity perspective, improving bikeshare access in pedestrian-dense areas may yield mobility benefits but also risks reinforcing gentrification pressures if not paired with affordability and community safeguards. Pedestrian sensor placement itself may reflect existing planning biases. Consequently, pedestrian-informed bikeshare planning should be integrated within broader equity-focused transportation policies rather than applied in isolation.

Finally, the repositioning engine we have built is meant to be robust against sparse data, but the results it gives will not account for multiple other factors such as city planning laws, costs, and usage during the months for which data is missing.

7 CONCLUSION

This study presents a unified spatiotemporal framework for evaluating the alignment between pedestrian mobility patterns and bikeshare system performance in New York City and Washington, DC. Our results show that pedestrian activity serves as a meaningful signal for identifying structural mismatches in bikeshare supply, though the strength and nature of these relationships vary by urban context and data availability. While NYC exhibits heterogeneous but generally positive pedestrian–bikeshare coupling, DC reveals weaker or negative associations driven by its mixed commuter–recreational network.

We further introduce a Repositioning Engine that converts multimodal mobility signals into data-adaptive reallocation strategies. Despite being designed for high-resolution inputs, the engine performs robustly under NYC's sparse biannual pedestrian observations, achieving measurable reductions in undersupply and misallocation.

Overall, this work demonstrates that integrating pedestrian mobility data into bikeshare planning workflows enables more effective detection of latent demand, improved

Table 3: Repositioning Engine Performance: System-Level Improvements

	NYC			DC			
System-Wide Metric	**Before**	**After**	**Change**	**Before**	**After**	**Change**	**Interpretation**
Severe Undersupply Share	0.1005	0.0585	**−41.8%**	0.0966	0.0619	**−36.0%**	Critical shortage reduction (*lower better*)
Severe Oversupply Share	0.1005	0.1032	+2.7%	0.0999	0.0836	**−16.3%**	Improved balance (*lower better*)
Misallocation Index	15,366	15,173	**−193**	243.0	235.8	**−7.2**	Better demand alignment (*lower better*)
Bike Gini Coefficient	0.5430	0.5299	**−0.0131**	0.6045	0.6056	+0.0011	Equity maintained (*lower better*)

accessibility, and more equitable micromobility system design, even under constrained sensing regimes.

8 ACKNOWLEDGMENT

We acknowledge the open data platforms that enabled this research, including the Capital Bikeshare, CitiBike NYC, DC Automated Bicycle & Pedestrian Counts, the NYC Bi-Annual Pedestrian Counts. These resources were essential in conducting a multi-city analysis of pedestrian mobility and bikeshare network alignment. We also thank Dr. Naren Ramakrishnan, whose instruction in our course provided the foundational concepts and analytical perspectives that informed the development of this work.

9 AUTHOR CONTRIBUTIONS

All authors contributed equally to this project. M. Shehryar developed and implemented the repositioning engine and produced the associated results and visualizations. Sanjna Kumari and Katelyn Crumpacker generated and consolidated results for the New York City and Washington, DC datasets, respectively. All authors jointly wrote and revised the manuscript.

DATA AND CODE AVAILABILITY

The datasets used in this study are publicly available, along with our code:
Capital Bikeshare (DC):
https://capitalbikeshare.com/system-data
DC Pedestrian Counts: District Department of Transportation.
https://catalog.data.gov/dataset/bicycle-and-pedestrian-automated-counts
Citi Bike (NYC):
https://citibikenyc.com/system-data
NYC Pedestrian Counts: NYC Department of Transportation.
https://catalog.data.gov/dataset/bi-annual-pedestrian-counts-6cc0a
All datasets accessed between 2022–2024.
Analysis and Repositioning Framework:
https://github.com/MKatelyn109/Urban-Final-Project

REFERENCES

[1] Avital Angel and Pnina Plaut. 2024. Tempo-spatial analysis of pedestrian movement in the built environment based on crowdsourced big data. *Cities* 149 (2024), 104917. https://doi.org/10.1016/j.cities.2024.104917

[2] Mohammad Sadegh Bahadori, Alexandre B. Gonçalves, and Filipe Moura. 2021. A Systematic Review of Station Location Techniques for Bicycle-Sharing Systems Planning and Operation. *ISPRS International Journal of Geo-Information* 10, 8 (2021). https://doi.org/10.3390/ijgi10080554

[3] Lindsay Braun, Daniel Rodriguez, and Penny Gordon-Larsen. 2018. 2397 - Social (in)Equity in Access to Cycling Infrastructure: Examining the Distribution of Bike Lanes with Respect to Area-Level Sociodemographic Characteristics in 23 Large U.S. Cities. *Journal of Transport Health* 9 (2018), S28. https://doi.org/10.1016/j.jth.2018.05.086

[4] Paul DeMaio. 2009. Bike-sharing: History, Impacts, Models of Provision, and Future. *Journal of Public Transportation* 12, 4 (2009), 41–56. https://doi.org/10.5038/2375-0901.12.4.3

[5] Ashley Dhanani, Lusine Tarkhanyan, and Laura Vaughan. 2017. Estimating pedestrian demand for active transport evaluation and planning. *Transportation Research Part A: Policy and Practice* 103 (2017), 54–69. https://doi.org/10.1016/j.tra.2017.05.020

[6] Guilherme N. Oliveira, Jose L. Sotomayor, Rafael P. Torchelsen, Cláudio T. Silva, and João L.D. Comba. 2016. Visual analysis of bike-sharing systems. *Computers Graphics* 60 (2016), 119–129. https://doi.org/10.1016/j.cag.2016.08.005

[7] J. Schuijbroek, R.C. Hampshire, and W.-J. van Hoeve. 2017. Inventory rebalancing and vehicle routing in bike sharing systems. *European Journal of Operational Research* 257, 3 (2017), 992–1004. https://doi.org/10.1016/j.ejor.2016.08.029

[8] Andrea Smith, Irina Olaru, and Menno Yap. 2015. Equity implications of bikeshare: A framework and initial analysis. *Transport Reviews* 35, 6 (2015), 1–20.

[9] Jinjun Tang, Fang Liu, Yinhai Wang, and Hua Wang. 2015. Uncovering urban human mobility from large scale taxi GPS data. *Physica A: Statistical Mechanics and its Applications* 438 (2015), 140–153. https://doi.org/10.1016/j.physa.2015.06.032