DATA STORIES

The Community Well-Being & Public Safety Framework

Part 2: Operationalizing the Framework

Data Stories seeks to illuminate how data and analytics can advance criminal justice reform

Key Takeaways

- → **The Framework can be operationalized using existing public data.**
  Building a meaningful cross-sector dataset doesn't require new data infrastructure — just a guiding framework, an anchoring outcome, and disciplined indicator selection.

- → **Community well-being conditions are interconnected, not interdependent.**
  Indicators in every domain correlated with incarceration and with each other — evidence that calls for coordinated, cross-domain responses, not single program fixes

- → **A composite score shows where strain is concentrated; the indicators show why.**
  The Community Well-Being Score makes it possible to scan all of New York City's neighborhoods quickly and identify the ones that stand out — and the individual indicators explain what is driving the score in any given neighborhood — and should shape the response.

From Framework To Practice

Building on their established data partnership, the New York City Criminal Justice Agency (CJA) and Luminosity are now turning that same analytical rigor to a new question: what conditions, at the neighborhood level, are associated with the people who end up in New York City's jails? Through years of collaboration — integrating data feeds from the New York Office of Court Administration and the New York City Police Department with direct interviews of individuals moving through the city's courts — CJA and Luminosity have built a comprehensive analytical picture of New York City's pretrial population.

That same partnership is now turning its lens upstream.

This second paper in the Data Stories series takes the Community Well-Being & Public Safety Framework introduced in Part 1 and puts it to work. Working together, CJA and Luminosity integrated administrative data on more than 70,000 jail admissions with publicly available neighborhood-level indicators across health, education, economic security, housing, and civic infrastructure — building an analytical foundation that connects community conditions to incarceration patterns across all five boroughs of New York City.

The goal is not to predict who will be incarcerated. It is to identify where the conditions that produce incarceration are most concentrated — and where coordinated upstream investment could change those conditions before people ever enter the justice

system. The third and final paper in this series will bring those findings to life — mapping the Community Well-Being Score across New York City's neighborhoods and showing, in concrete terms, where that investment is needed most.

The Outcome Of Interest: Incarceration

Incarceration is an outcome directly tied to CJA's mission, and one of the data points where neighborhood conditions intersect most consequentially with the justice system. As of April 2026, 85% of the people held at Rikers were in pretrial detention — meaning the single largest pathway to responsibly reducing the jail population runs through reducing unnecessary pretrial incarceration.

Choosing incarceration as the outcome of interest also unlocks a specific kind of analysis. If neighborhood-level patterns of incarceration track with measurable conditions — housing strain, economic hardship, gaps in health or education infrastructure — then the data points toward where upstream investment could reduce downstream system involvement. The aim is not to predict who will be incarcerated, but to identify which neighborhoods are under the kind of strain that produces disproportionate rates of incarceration, and where coordinated supports could address those conditions before they escalate.

The same rigor that built a comprehensive picture of New York City's pretrial population is now being applied upstream — to the conditions that shape who enters the system in the first place

The analysis drew on over 70,000 jail admission records from 2023 through 2025. To translate those records into a neighborhood-level indicator, each was assigned to one of New York City's Neighborhood Tabulation Areas (NTAs) — geographies created by the Department of City Planning that approximate neighborhoods and serve as a common unit of analysis across city agencies. NTAs are large enough to support stable indicator construction and small enough to reveal meaningful variation in community conditions. To enable fair comparison across neighborhoods of different sizes, admissions were converted into rates per 10,000 residents.

Assigning records to NTAs required a hierarchical geocoding approach rather than reliance on a single address field. When a residential address was incomplete or unavailable, the analysis drew on other location information available in justice-system data — including arrest location, used as a proxy when no residential address was on file. This was a deliberate choice: people who are unhoused are less likely to have a stable address in administrative records, and excluding them would have systematically distorted the analysis by under-representing some of the neighborhoods most affected by incarceration. Using this approach, 93% of records were successfully mapped to an NTA. About 6% reflected addresses outside New York City and were excluded; 1% lacked any usable geographic information.

Selecting The Indicators

With incarceration mapped at the neighborhood level, the next step was to identify the community well-being indicators to test against those incarceration patterns. The Framework guided that search. Rather than starting with available datasets and working backward, the analysis started with the question the Framework asks: which neighborhood conditions, across which domains, most plausibly shape community well-being and public safety? The search for data followed from that question.

This analysis examined whether community conditions across NYC neighborhoods are associated with incarceration patterns — not whether any individual's characteristics or experiences explain who enters custody. Of the four ecological levels in the Framework, this phase drew on two: daily settings (home, school, work) and neighborhood conditions (the institutions

and infrastructure that shape daily life). Indicators capturing those settings were measured at the neighborhood level so they could be compared across the city.

Domains of Community Well-Being

Community well-being is broad by design — it covers everything that makes daily life stable or unstable. To make it analytically tractable, the analysis focused on five domains that together capture the conditions most directly tied to whether neighborhoods support or strain the people who live in them: **health, education, economic security, housing, and civic infrastructure.**

Each domain reflects a different way neighborhoods shape everyday life. Health captures access to care and the cumulative toll of neighborhood conditions on physical well-being. Education captures the opportunities available to children and the strength of the institutions that serve them. Economic security captures material stability — whether families can meet basic needs and weather setbacks. Housing captures whether people can stay in stable, safe homes. Civic infrastructure captures the condition and responsiveness of the shared neighborhood environment, from sanitation to street safety.

From Domains to Indicators

Within each domain, indicators were tested for three qualities: conceptual clarity, meaningful association with incarceration, and minimal redundancy with other indicators. The indicators examined within each domain are listed below. The next section describes how the analysis narrowed them to the final indicator set.

- **Health:** health insurance status and life expectancy at birth.
- **Education:** school enrollment and degree attainment.
- **Economic security:** poverty rate, public assistance rate, unemployment rate, median personal earnings, and median household income.
- **Housing:** eviction rate (residential instability), rent burden (housing affordability), and 311 housing and building complaints.
- **Civic infrastructure:** 311 calls related to sanitation, cleanliness, and noise (quality of life in the shared public realm).

These indicators capture present-day conditions, but those conditions are themselves products of longer histories — patterns of investment and disinvestment, redlining, and unequal access to opportunity that have shaped neighborhoods over generations. The data make today's conditions visible; the deeper forces that produced them are part of the context, even when not directly measured.

Three Data Sources

Three publicly available data sources supplied the community well-being indicators: the American Human Development Index, the American Community Survey, and NYC Open Data. New York City's open data laws made this cross-domain approach more feasible than it would be in many jurisdictions — by requiring agencies to publish administrative data through a single citywide portal, the city has built a common infrastructure for accessing data across domains. External sources like the Census and published neighborhood indexes extend that foundation, making it possible to assemble a meaningful cross-sector dataset without building new data infrastructure.

The American Human Development Index (AHDI), produced by Measure of America, was especially useful because it is available at the NTA level and its dimensions can be examined separately. This phase tested all three of AHDI's sub-indexes: the Health Index

The Framework set the question — which neighborhood conditions plausibly shape community well-being and public safety — and the search for data followed.

(life expectancy at birth), the Education Index (school enrollment and degree attainment), and the Income Index (median personal earnings).

The American Community Survey (ACS), conducted by the U.S. Census Bureau, provided neighborhood-level information on social, economic, and housing conditions. Because ACS data is available for small geographies and the Department of City Planning translates ACS estimates into NTAs, these data fit directly into the analysis.

supplied the 311 service-request data — a citywide record of non-emergency complaints routed to city agencies, with complaint type, responding agency, and geographic location attached. The 2025 311 dataset contained more than three million service requests citywide. From these, the thirty most common complaint types — together accounting for over 80% of all calls — were retained and tested against the neighborhood incarceration rate. Only complaint types with a meaningful association were kept.[1] The surviving complaints were grouped into three thematic categories: Housing & Buildings, Sanitation & Cleanliness, and Noise. Housing & Buildings was assigned to the Housing domain alongside rent burden and eviction rate. Sanitation & Cleanliness and Noise were assigned to the Civic Infrastructure domain.

Testing Indicators Against Incarceration

With indicators identified across all five domains, the next phase tested how those indicators related to incarceration patterns across New York City's neighborhoods. The goal was twofold: identify which indicators were most strongly associated with incarceration, and identify which ones added distinct information rather than reflecting the same underlying conditions as other indicators.

The analysis proceeded in two steps. The first examined the relationships between each individual indicator and incarceration rates, and tested how the indicators related to one another. The second combined the selected indicators into a single Community Well-Being Score that could be compared across neighborhoods.

Correlation Analysis

For each domain with multiple indicators, the analysis tested the strength and direction of the relationship between each indicator and incarceration rates. It also tested how the indicators related to one another — both within domains and across them. This second test mattered as much as the first: when two indicators move together too closely, they may be capturing the same underlying condition rather than adding distinct information to the analysis.

Across all five domains, multiple indicators showed meaningful correlations with incarceration rates, and many were also correlated with indicators in other domains — confirming the Framework's premise that community well-being conditions are interconnected rather than independent. The analysis then narrowed each domain to a single indicator: the one with the strongest relationship to incarceration and the least overlap with indicators already retained from other domains. This produced a final set of five indicators, one per domain:

- **Health:** AHDI Health Index [2]
- **Education:** AHDI Education Index [3]
- **Economic security:** Poverty Rate [4]
- **Housing:** 311 calls for Housing and Buildings [5]
- **Civic infrastructure:** 311 calls for Sanitation and Noise [6]

Building a Community Well-Being Score

A single comparable indicator makes it easier to see, at a glance, where neighborhoods stand. To build that, the five selected indicators were combined into a composite Community Well-Being Score for each NTA — one number per neighborhood that summarizes performance across health, education, economic security, housing, and civic infrastructure.

The five selected indicators were combined into a single Community Well-Being Score for each NTA. Each indicator was placed on a common scale and aligned in direction, so that a higher value consistently means stronger community well-being. The indicators were then averaged with equal weight to produce a composite score on a 1–100 scale. The composite was validated against neighborhood incarceration rates and was more closely associated with incarceration than any of the underlying indicators on their own[7]. To

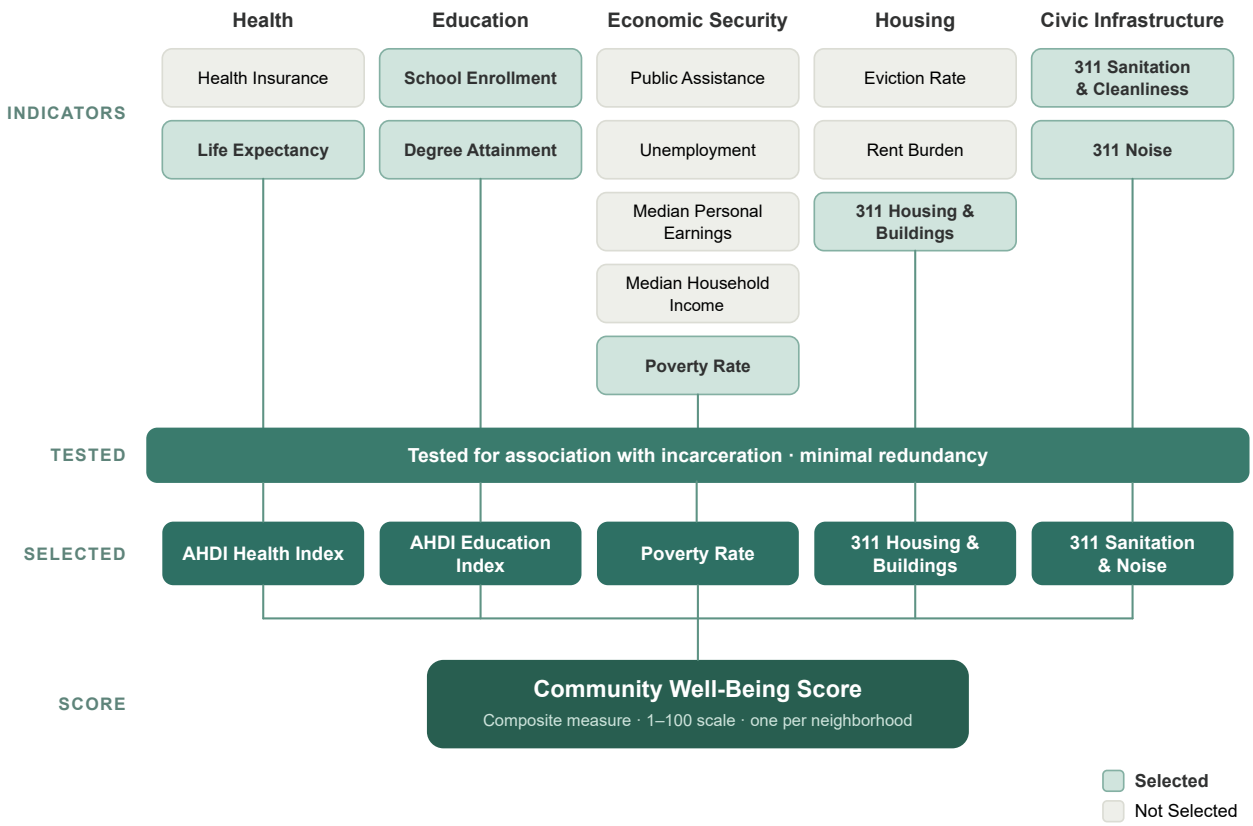

Figure 1: Selecting Indicators and Building the Community Well-Being Score

confirm the score is not an artifact of methodological choices, its construction was checked using different methodologies; the main pattern held across all of them[8].

NTAs were then grouped into above-average, near-the-average, and below-average categories using one standard deviation above and below the citywide mean[9]. Above-average community well-being scores signal more supportive neighborhood conditions, which the Framework posits should be associated with lower incarceration rates. Below-average scores signal neighborhoods where conditions are most likely to be straining the people who live there.

The value of the score lies in what it makes possible. With nearly 200 Residential NTAs across New York City, comparing five indicators across every neighborhood is unwieldy; comparing one composite score is not. The score makes it possible to scan the full city quickly, identify neighborhoods that stand out, and prioritize where deeper investigation is warranted. It also makes the relationship between community well-being and incarceration directly visible — a relationship that is difficult to see when the underlying indicators are examined separately. Still, the score doesn't replace the indicators behind it. It shows where strain is concentrated; the individual indicators show what is driving it — and different drivers call for different responses.

Importantly, the score is built entirely from publicly available data sources. Any jurisdiction with comparable open data infrastructure could construct a similar indicator, adapted to its own neighborhood geographies and locally relevant indicators. The approach is transferable; the score itself is a starting point for analysis, not a final verdict on any neighborhood.

Emerging Lessons

Several lessons emerged from operationalizing the Framework. They speak both to what is possible with publicly available cross-sector data and to the practical

realities of working with that data at a neighborhood level.

1. **The Framework Guided the Search, Not the Other Way Around**

Most cross-sector data projects start with available datasets and work backward to the questions those datasets can answer. This analysis took the opposite approach. The Framework set the question — which neighborhood conditions plausibly shape community well-being and public safety — and the search for data followed. Letting the Framework lead identified gaps where good indicators didn't exist, prevented inclusion of indicators just because they were available, and produced a dataset organized around the conditions the Framework identifies as consequential. The same data sources, assembled without a guiding framework, would yield a weaker analysis.

2. **Cross-System Data Integration Required Active Reconciliation**

Indicators came from sources with different geographies, reporting periods, and units of measurement. Census-tract data had to be aligned with NTAs. Reporting timeframes lagged by varying amounts. Raw counts had to be normalized to rates so neighborhoods of different sizes could be fairly compared. Each of these choices shaped what the analysis could see. Cross-sector analysis takes methodological work, not just access to data — and the choices made during integration should be documented as carefully as the findings themselves.

3. **Domains Are Interconnected, and Indicators Should Earn Their Spot**

Across all five domains, multiple indicators correlated with incarceration rates — and many also correlated with indicators in other domains. That second finding confirmed the Framework's premise that community well-being conditions are interconnected, not independent. It also made selection more demanding: when indicators overlap, retaining all of them inflates the appearance of evidence without adding distinct information. Testing each indicator against incarceration and against each other produced a leaner, more interpretable indicator set. Every indicator in the final analysis earned its spot.

The approach is transferable. Any jurisdiction with comparable open data infrastructure could construct a similar measure — adapted to its own neighborhoods and locally relevant indicators.

4. **This Phase Tested Associations, Not Early Warning Signals**

The first paper in this series framed community well-being indicators as potential early warning signals — indicators that shift before serious harm escalates. Demonstrating that role requires tracking these indicators over time. This phase tested something different and more foundational: the cross-sectional relationship between current neighborhood conditions and current incarceration patterns. It established that the associations the Framework predicts are visible in the data. Whether these indicators function as true early warning signals is a question for further analysis.

5. **Public Data Captures Strain More Readily Than Strengths**

The public data environment in New York City is unusually rich, and the analysis depended on it. But that environment has a built-in tilt. Public datasets regularly document poverty, eviction, service complaints, and other signs of distress. They rarely capture social cohesion, resident leadership, mutual aid, or the everyday problem-solving capacity that helps neighborhoods absorb stress. The indicators available here were stronger at identifying vulnerability than at representing resilience. Operationalizing the Framework more fully will require indicators designed to capture neighborhood assets directly.

Looking Ahead

This analysis showed that a meaningful cross-sector dataset can be built by pairing administrative justice records with publicly available community well-being data, without waiting for new data infrastructure or fully integrated systems. The Framework guided which conditions to examine. Incarceration anchored the analytic choices. Publicly available indicators across five domains, integrated at the neighborhood level, made the connection between community conditions and incarceration visible across all five boroughs.

Publication 3 takes the next step. It will map the Community Well-Being Score across New York City's neighborhoods, report the statistical relationships between community conditions and incarceration, and use brief neighborhood profiles to show what the patterns reveal in context. The goal is not just to describe where incarceration is concentrated, but to identify where coordinated upstream investment — across health, education, economic security, housing, and civic infrastructure — could ease the conditions that produce it.

Acknowledgments

CJA and Luminosity gratefully acknowledge the contributions of Dave Reddy and members of Luminosity's data analytics team — Abhijit Wayal, Ata Geldimuradov, and Naveen Veeramreddy — whose collaborative work on data integration and analysis made this publication possible. We also thank Branden Dupont whose editorial review sharpened both the analysis and the writing, and Cherise Fanno Burdeen, PhD, for the subject matter expertise she brought to the work.

References

Boxer, K. S., Hong, B., Kontokosta, C. E., & Neill, D. B. (2025). Estimating reporting bias in 311 complaint data. The Annals of Applied Statistics, 19(2), 1691–1713. https://doi.org/10.1214/24-AOAS2003

Kontokosta, C. E., & Hong, B. (2021). Bias in smart city governance: How socio-spatial disparities in 311 complaint behavior impact the fairness of data-driven decisions. Sustainable Cities and Society, 64, 102503.

OECD & European Commission Joint Research Centre. (2008). Handbook on Constructing Composite Indicators: Methodology and User Guide. OECD Publishing. https://doi.org/10.1787/9789264043466-en

Saisana, M., Saltelli, A., & Tarantola, S. (2005). Uncertainty and sensitivity analysis techniques as tools for the quality assessment of composite indicators. Journal of the Royal Statistical Society, Series A, 168(2), 307–323. https://doi.org/10.1111/j.1467-985X.2005.00350.x

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society, Series B, 57(1), 289–300.

Endnotes

1 A known concern with 311 data is that complaint volumes reflect not only neighborhood conditions but also residents' willingness or ability to report (Boxer, Hong, Kontokosta & Neill, 2025; Kontokosta & Hong, 2021). The filtering step addressed this directly: complaint types dropped from the analysis — including Damaged Tree, Missed Collection, Illegal Parking, Street Condition, and General Construction/Plumbing (e.g., No Building Permit, Inadequate Fence, Excess Debris) — showed negative association with incarceration rates. This variation across complaint types indicates the retained indicators reflect real conditions, not reporting behavior alone.

2 Measure of America (2022). American Human Development Index — Health Index sub-component (life expectancy at birth). Data2Go. NYC. Underlying data: NYC Department of Health and Mental Hygiene, 2011–2020. https://data2go.nyc

3 Measure of America (2022). American Human Development Index — Education Index sub-component (school enrollment ages 3–24, weighted one-third; degree attainment ages 25 and over, weighted two-thirds). Data2Go.NYC. Underlying data: U.S. Census Bureau, American Community Survey 1-year estimates, 2022. https://data2go.nyc

4 U.S. Census Bureau (2024). American Community Survey 5-year estimates, 2020–2024. Table B17001: Poverty Status in the Past 12 Months by Sex by Age. https://data.census.gov/table/ACSDT5Y2024.B17001

5 NYC Open Data (2025). 311 Service Requests from 2020 to Present. New York City Office of Technology and Innovation. https://data.cityofnewyork.us/Social-Services/311-Service-Requests-from-2020-to-Present/erm2-nwe9. The Housing & Buildings complaint subset was derived from this dataset.

6 NYC Open Data (2025). 311 Service Requests from 2020 to Present. New York City Office of Technology and Innovation. https://data.cityofnewyork.us/Social-Services/311-Service-Requests-from-2020-to-Present/erm2-nwe9. The Sanitation & Noise complaint subset was derived from this dataset.

7 As a check on the composite's construction, the Well-Being Score was compared against incarceration rate per 10,000 residents using Spearman rank correlation and found to align more closely than any single underlying indicator ($|\rho| = 0.80$ within boroughs; $|\rho| = 0.75$ citywide; $|\rho| = 0.69$ for the strongest single indicator). This confirmed that combining the indicators produced a more informative measure than any one alone — the rationale for using a composite.

8 The score's construction was tested against a range of alternative choices: combining the indicators by geometric rather than arithmetic mean, weighting the underlying clusters of indicators equally rather than each indicator equally, and removing each indicator in turn. A broader sensitivity run varied the normalization method, the combination method, and the set of indicators together across many specifications. Across these tests, the neighborhoods identified as above- and below-average remained largely stable, indicating the score reflects underlying conditions rather than any single methodological choice (Saisana, Saltelli & Tarantola, 2005; OECD, 2008).

9 Each indicator was rescaled to a common 1–100 range and aligned in direction so that higher values consistently indicate stronger well-being; the two 311 rate indicators were log-transformed before rescaling to reduce the influence of a small number of high-rate neighborhoods. The five aligned indicators were then averaged with equal weight (OECD, 2008). Tier cutpoints were set at one standard deviation above and below the citywide mean, which places roughly the top and bottom sixth of neighborhoods in the above- and below-average categories and the remaining two-thirds near the middle.