Working Paper: Explaining Public Park Reviews in California using Natural Language Processing

The Massive Data Institute: Greenspace Challenge

Himangshu Kumar[a,1], Ayush Lahiri[a,c]
[a] *Georgetown University;* [c] *Ashoka University*

Abstract

Abstract

Traditional metrics of urban infrastructure equity rely predominantly on spatial coverage—measuring proximity to amenities like parks—under the assumption that "access" equates to utility. We challenge this assumption by employing Natural Language Processing (NLP) to measure the *lived experience* of public goods. Utilizing a dataset of approximately 1.5 million Google Reviews for green spaces in California, we implement a two-stage machine learning pipeline: Latent Dirichlet Allocation (LDA) for topic discovery, followed by a DeBERTa-based Aspect-Based Sentiment Analysis (ABSA) model. This allows us to decompose user feedback into specific dimensions of infrastructure quality—specifically safety, maintenance (cleanliness), and social environment (e.g., homelessness). We merge these sentiment scores with tract-level demographic data from the American Community Survey (ACS). Our central finding reveals a divergence between spatial access and perceived quality: while administrative data suggests equitable coverage, granular sentiment analysis uncovers statistically distinct deficits in safety and maintenance perceptions in lower-income and minority-heavy tracts. This suggests that free-form text can serve as a scalable, high-frequency sensor for infrastructure decay that rigid survey instruments fail to capture.

1 Introduction

The equitable distribution of public goods is a central tenet of urban planning and development economics [1, 2]. Historically, the efficacy of this distribution has been measured through "stocks" and "flows"—budgetary allocations for park maintenance or geospatial analyses of catchment areas. Under these traditional metrics, a neighborhood is considered served if a green space exists within a specific radius of its residents [3].

However, physical proximity does not guarantee functional utility. A park that is unsafe, ill-maintained, or structurally degraded offers the same statistical "coverage" as a pristine facility, yet provides vastly different welfare outcomes [4]. This creates a measurement gap: administrative data captures the *existence* of infrastructure, but rarely its *quality* or the *subjective experience* of its users. Standard methods to bridge this gap, such as municipal surveys or systematic audits, are costly, low-frequency, and suffer from rigid questionnaire designs that may miss organic community concerns [5].

In this paper, we explore whether unstructured, free-form text data can fill this void. We treat online reviews not merely as consumer feedback, but as "passive sensors" of civic infrastructure quality [6, 7]. Focusing on California—a state with high socioeconomic diversity and dense urban green spaces—we employ a dual-track identification strategy. We first measure **"administrative access"** using the full universe of park locations to test for spatial equity. We then contrast this with **"perceived quality,"** derived from a stratified, balanced sample of user narratives. This allows us to isolate the divergence between coverage on paper and the reality on the ground.

Our methodological contribution lies in the application of Aspect-Based Sentiment Analysis (ABSA) to this stratified sample [8]. While standard sentiment analysis might classify a review as "negative," it fails to distinguish between a negative review caused by *safety concerns* versus one caused by *weather*. By utilizing a pre-trained DeBERTa transformer model [9], we isolate specific infrastructural dimensions—Safety, Hygiene, and Homelessness—and regress these specific sentiment scores against tract-level socioeconomic indicators, controlling for the baseline density of parks.

2 Data

Our analysis integrates two primary data sources: unstructured user-generated text from Google Maps and administrative demographic records from the US Census Bureau. To address our dual research questions—regarding both spatial access and qualitative experience—we construct two distinct analytical datasets derived from the same source corpus.

2.1 The Review Corpus and Analytical Samples

Review data was sourced from the City78 dataset made available to us by the Massive Data Institute through the Environmental Impact Data Collaborative (EIDC) portal [10] and the Department of Housing and Urban Development as part of the Greenspace Research Challenge [11].Approximately 30,000 parks and open spaces across the United States, along with other public spaces outside the perview of our research question. Each observation includes geolocation attributes (latitude/longitude) and the full text of user reviews. We restrict our analysis to the State of California (2015–2019) to ensure temporal relevance with our demographic covariates. From this raw corpus, we derive two datasets:

1. The Access Universe ($\mathcal{D}_{access}$): To measure physical infrastructure distribution, we retain the full universe of unique park locations in California present in the dataset. This results in a sample of $N = 2,761$ unique green spaces distributed across 2,107 Census Tracts. This dataset is used exclusively to calculate park density per capita.

2. The Sentiment Sample ($\mathcal{D}_{sentiment}$): To measure the lived experience of infrastructure, we draw a subsample of textual reviews. Due to the computational constraints of transformer-based inference, analyzing the full corpus of $\sim$1.5 million reviews is infeasible. Furthermore, online ratings typically exhibit a "J-shaped" distribution where extreme 1-star and 5-star ratings are overrepresented. To correct for this and capture the nuance of average user experiences, we constructed a balanced, stratified sample of 5,000 reviews, drawing 1,000 reviews randomly from each star rating category (1–5).

2.2 Demographic Covariates

We link both datasets to the 2019 American Community Survey (ACS) 5-Year Estimates. Using the coordinate data for each park, we performed a spatial join to assign each observation to a Census Tract. For each tract t, we extracted a vector of socioeconomic covariates (X_t) designed to capture economic stability, racial composition, and population intensity:

- **Economic Indicators:** $\ln(\text{Median Household Income})$ and Median Home Value. These variables proxy for the local tax base and private capital stock.
- **Racial Composition:** To identify specific disparities rather than generic minority effects, we control for the percentage of residents identifying as Black (Non-Hispanic), Hispanic/Latino, and Asian. White (Non-Hispanic) serves as the reference category.
- **Urban Intensity:** We control for Population Density (people per sq. mile). This is a critical control; lower-income areas in California are often denser, potentially leading to higher rates of infrastructure depreciation due to usage intensity rather than administrative neglect.
- **Age Structure:** We control for the age composition of the census tract to adjust for potential rating heterogeneity, as survey literature suggests that younger demographics tend to exhibit a baseline negativity bias in online feedback.

3 Methodology: From Text to Signal

Our empirical strategy relies on transforming unstructured text into structured indices of infrastructure quality. This process involves a two-stage machine learning pipeline: unsupervised topic discovery to validate salient dimensions, followed by supervised aspect-based sentiment quantification.

3.1 Text Preprocessing

Prior to feature extraction, the raw review corpus underwent a standardized cleaning pipeline to reduce noise and enforce schema consistency. First, we applied Unicode normalization to resolve encoding errors common in web-scraped data. Second, we utilized regular expressions to strip non-textual artifacts, including URL links, HTML tags, and emojis, which constitute a significant portion of user-generated content on mobile platforms.

It is important to note that we employed a bifurcated preprocessing strategy. **For the topic modeling phase (Phase I)**, we applied aggressive filtration—including lower-casing, punctuation removal, lemmatization, and the exclusion of standard English stop-words—to maximize the coherence of the bag-of-words representation. **Conversely, for the Sentiment Analysis phase (Phase II)**, we retained the original capitalization and sentence delimiters continuing to remove non-textual artifacts, as the DeBERTa transformer architecture relies on syntactic structure to resolve context and coreference.

3.2 Phase I: Latent Topic Discovery (LDA)

Before deploying computationally expensive deep learning models, we first validated which "aspects" of park quality were strictly endogenous to the user experience. We employed Latent Dirichlet Allocation (LDA), a generative probabilistic model, to identify latent topic clusters within the corpus.

We utilized LDA for *feature selection* rather than final classification. The model identified co-occurring high-frequency terms such as ["homeless", "needles", "tents"] grouping into a distinct topic, and ["play", "swings", "kids"] into another. This unsupervised validation justified our selection of four specific aspects for the subsequent ABSA analysis:

1. **Safety/Homelessness:** Perceptions of security, vagrancy, and illicit activity.
2. **Cleanliness:** References to trash, maintenance, and facility hygiene.
3. **Children/Amenities:** Suitability for families and equipment quality.
4. **General Aesthetics:** (Used as a control aspect).

3.3 Phase II: Aspect-Based Sentiment Analysis

Standard dictionary-based methods or simple keyword matching (e.g., REGEX) are insufficient for policy analysis due to semantic ambiguity. For instance, the phrases *"I love that they help the homeless here"* and *"Too many homeless people, felt unsafe"* would both trigger a keyword match but represent diametrically opposite welfare outcomes.

To resolve this, we employ the `DeBERTa-v3-base-ABSA` architecture. DeBERTa (Decoding-enhanced BERT with Disentangled Attention) improves upon standard BERT models by utilizing a disentangled attention mechanism, where each word is represented using two vectors: one for its content and another for its relative position. This is particularly critical for aspect-based sentiment, where the descriptive adjective (e.g., "dirty") may be syntactically distant from the target aspect (e.g., "bathrooms") within a complex sentence structure.

For every review i in the stratified sample, we construct input pairs consisting of the raw text and a target aspect k. The model estimates the conditional probability distribution of sentiment S:

$$P(S|\text{Review}_i, \text{Aspect}_k) \quad \text{where } S \in \{\text{Pos, Neu, Neg}\}$$

The DeBERTa architecture returns a probability distribution over three sentiment labels: $\mathcal{L} = \{\text{Positive}, \text{Neutral}, \text{Negative}\}$. While using raw probabilities captures uncertainty, for the purpose of regression analysis, we require a single continuous scalar that represents both the *direction* and *intensity* of the sentiment.

To construct this metric, we employ a *High-Confidence Thresholding* strategy to filter out ambiguous or weak signals. Let p_c denote the probability assigned to class $c \in \mathcal{L}$. We define the *Composite Polarity Score* (CPS_{ik}) for aspect k in review i as follows:

$$CPS_{ik} = \begin{cases} p_{pos} & \text{if } \arg\max(\mathcal{L}) = \text{Pos} \wedge p_{pos} > \tau \\ -p_{neg} & \text{if } \arg\max(\mathcal{L}) = \text{Neg} \wedge p_{neg} > \tau \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

Where $\tau = 0.6$ is our predefined confidence threshold.

This transformation yields a bounded variable $CPS_{ik} \in [-1, 1]$, where values approaching -1 indicate high-confidence structural deficits, values approaching $+1$ indicate strong satisfaction, and 0 represents either neutral statements or ambiguous text.

3.4 Empirical Specification

Our identification strategy relies on a "divergence framework." We test whether the correlation between socioeconomic status and public goods provisioning differs depending on whether "provisioning" is measured by administrative proximity or subjective quality. We estimate two distinct specifications:

Model 1: The Administrative Access Test
First, to test for the existence of "park deserts," we regress the park density of a census tract on its demographic characteristics using the full universe dataset ($\mathcal{D}_{access}$):

$$\text{Density}_t = \alpha + \beta_1 \ln(\text{Inc}_t) + \Gamma Race_t + \phi \text{PopDens}_t + \lambda \text{Age}_t + \delta_c + \epsilon_t \quad (2)$$

Where Density_t is the number of parks per 1,000 residents in tract t. If administrative planning is equitable, we expect β_1 and Γ to be statistically indistinguishable from zero.

Model 2: The Lived Experience Test
Second, to test for qualitative disparities, we regress the DeBERTa-derived sentiment scores on the same vector of covariates using the stratified review sample ($\mathcal{D}_{sentiment}$):

$$\text{CPS}_{ikt} = \alpha + \beta_2 \ln(\text{Inc}_t) + \Gamma Race_t + \phi \text{PopDens}_t + \lambda \text{Age}_t + \delta_c + \epsilon_{it} \quad (3)$$

Where:

- CPS_{ikt} is the Composite Polarity Score for aspect k (e.g., Safety) in review i located in tract t.
- $Race_t$ is a vector of racial composition percentages (Black, Hispanic, Asian).
- PopDens_t controls for usage intensity, and Age_t controls for demographic rating bias.
- δ_c represents County Fixed Effects, restricting comparisons to within the same administrative jurisdiction.

The Divergence Hypothesis: We hypothesize that $\beta_1 \approx 0$ in Eq. (2) (indicating equal access), while $\beta_2 < 0$ and $\Gamma > 0$ in Eq. (3) (indicating that lower income and higher minority shares predict significantly worse perceived quality).

3.5 Interpretation of the Stratified Estimator

It is important to qualify the interpretation of β_2 in Equation (3). Our analytical sample ($\mathcal{D}_{sentiment}$) is restricted to 5,000 reviews primarily due to the significant computational costs associated with the DeBERTa inference pipeline. Consequently, this sample is spatially sparse; not every census tract in California is represented equally, and the limited sample size prevents us from calculating precise tract-level sentiment estimates.

Therefore, our results should not be interpreted as a definitive, statewide census of park quality. Instead, we present these findings as **preliminary insights** into the structural relationship between demographics and user experience. Therefore, our results should not be interpreted as a definitive representation of population, but rather as a validation of the ABSA methodology. While we acknowledge that the specific magnitude of our coefficients may be influenced by the spatial sparsity resulting from computational constraints, our primary objective is to demonstrate the capacity of ABSA to extract granular infrastructure insights from unstructured text and to illustrate how these latent signals can be operationalized within standard inferential frameworks to yield actionable policy intelligence. We posit that these preliminary findings serve to motivate further research in settings where sufficient computational resources allow for the scaling of this approach to the full universe of civic feedback, where such sampling constraints would not exist.

4 Results

4.1 Model Validation: Disentangling Sentiment

Before analyzing aggregate trends, we first validate the ABSA model's ability to disentangle complex, multi-aspect user narratives. A common challenge in standard NLP is the "bleeding" of sentiment across topics. To demonstrate our model's specificity, we examine a representative review from the corpus that conflates homelessness with child safety.

Before analyzing aggregate trends, we validate the ABSA model's ability to disentangle complex, multi-valent user narratives. A common challenge in standard NLP is the "bleeding" of sentiment, where a negative clause overshadows positive nuance. To demonstrate our model's specificity, we examine a representative review that contrasts high-quality amenities with environmental concerns.

Example 1: Child Friendly vs. Safety

Sample Text: "My kids actually enjoy playing on the structures here, they are quite new. But I feel uncomfortable staying long because there are usually a few homeless people around in an inebriated state." *(Text edited for sensitivity)*

When processed for the aspect **"Children/Amenities"**, the model detects the specific praise for the infrastructure, assigning a Composite Polarity Score (CPS) of **+0.98**.

However, when the same sentence is processed for the aspect **"Safety"**, the model isolates the user's specific discomfort regarding the environment, assigning a CPS of **-0.92**. This polarity shift demonstrates the model's capacity to distinguish between the *provision* of goods (the playground exists and is good) and the *consumption* of goods (the environment makes it hard to use), a distinction lost in a simple 3-star rating.

This distinction is critical: while the presence of homelessness is noted as negative, the model correctly identifies that the *implication* for the "Children" aspect is far more severe (effectively rendering the park unusable). This granular separation prevents the noise of general complaints from obscuring specific, actionable safety signals.

Example 2: Aesthetics vs. Maintenance

Sample Text: "Honestly a beautiful spot with great shading from the big trees, perfect for a walk. It's just a shame that the trash cans are always overflowing and there is litter all over the grass." *(Text edited for clarity)*

Here, the divergence is strictly operational. For the aspect **"General Aesthetics"**, the model focuses on the "beautiful spot" and "trees," returning a Positive score of **+0.94**.

Conversely, for the aspect **"Cleanliness"**, the model latches onto "overflowing" and "litter," flipping the sentiment to a strong Negative score of **-0.91**. This validates our ability to separate the natural endowment of a park from the administrative effort (or lack thereof) put into maintaining it.

4.2 Manual Validation and Consistency Check

Given the "black box" nature of deep learning inference, relying solely on aggregate metrics poses risks of hallucination. Ideally, we would manually audit the entire corpus; however, manual annotation is prohibitively resource-intensive. To approximate a ground-truth validation, we extracted a stratified random sub-sample of $N = 125$ reviews (25 randomly selected from each star-rating bucket).

Two independent annotators coded these reviews for the target aspects. We observed a high degree of concordance between human judgment and model output ($> 85\%$ agreement on polarity direction). Where disagreements occurred, they followed a consistent pattern: the model tended to classify subtly negative phrasing as "Neutral" (assigning a score of 0 due to our $\tau < 0.6$ threshold), whereas human annotators perceived a specific, albeit weak, negative sentiment. This suggests that our sentiment scores likely exhibit a *conservative bias*, effectively serving as a lower-bound estimate of user dissatisfaction.

Table 1 presents the mean Composite Polarity Scores (CPS) for this subsample. The results track intuitively with the ground-truth star ratings, validating that the DeBERTa model is correctly aligning aspect sentiment with overall user satisfaction, while capturing the nuance that even low-rated parks may still have positive aesthetic qualities.

Table 1: **Mean Composite Polarity Score by Star Rating (Validation Sample)**

	Mean Aspect Score (CPS)			
Star Rating	**Safety**	**Cleanliness**	**Children**	**Aesthetics**
1 Star	-0.82	-0.91	-0.65	-0.15
2 Star	-0.45	-0.58	-0.22	0.12
3 Star	-0.05	-0.14	0.08	0.65
4 Star	0.42	0.35	0.55	0.73
5 Star	0.68	0.71	0.82	0.71

Note: Sample $N = 125$. Scores range from -1 (Neg) to +1 (Pos).

4.3 The Divergence: Access vs. Experience

It is important to note the distinct units of analysis across these specifications: Column (1) is estimated at the census tract level ($N = 2,107$ unique tracts containing parks for which we had data), whereas Columns (2–4) are estimated at the individual review level ($N = 5,000$ stratified reviews)."

Table 2 presents our primary inferential results. We estimate the same specification across four distinct dependent variables to test the "divergence hypothesis." Column (1) models physical access (Park Density) using the full universe of locations. Columns (2) through (4) model the *lived experience* (Composite Polarity Score) for specific aspects using the stratified review sample. **As discussed earlier the results should be interpreted with caution (we discuss the weaknesses in further detail in later sections)**.

4.3.1 Administrative Equity vs. Operational Neglect

The comparison between Column (1) and Columns (2–4) reveals a stark disconnect between the *provision* of infrastructure and its *maintenance*.

1. The Null Result on Access: Consistent with recent findings in urban epidemiology, we find little evidence of *park deserts* driven by demographics. In Column (1), the coefficient for Median Household Income is positive but only weakly significant ($p < 0.1$), while the coefficients for Home Value and all racial subgroups remain statistically indistinguishable from zero. This confirms that, administratively, green space is allocated largely equitably; while there is a marginal correlation with income, lower-income and minority-majority tracts are not systematically deprived of physical park locations [12, 13].

While wealthier tracts possess a slightly higher density of parks per capita—likely a function of lower population density in affluent suburbs—the magnitude is economically negligible. This aligns with recent audits suggesting that while traditional access metrics (like the 10-minute walk standard) have improved, they increasingly mask deep disparities in the *usability* and *condition* of those spaces [14, 15].

2. The Racial Stratification of Quality: However, when we analyze the *quality* of these spaces, the equitable picture collapses. Crucially, the "minority penalty" is not uniform; it is concentrated in specific communities.

- **Black Communities:** We observe the most severe divergence in tracts with higher Black populations. The coefficient is strongly negative and significant across both *Cleanliness* ($\beta = -0.124$) and *Children/Amenities* ($\beta = -0.145$). This implies that parks in these neighborhoods are systematically more likely to be flagged for safety hazards and disrepair compared to white-majority baselines.

- **Hispanic Communities:** The divergence here is aspect-specific. While *Children/Amenities* shows no significant deficit (suggesting playgrounds exist and are functional), there is a significant negative coefficient for *Cleanliness* ($\beta = -0.082, p < 0.05$). This points to a deficit in "operational" maintenance (sanitation, trash removal) rather than "capital" failure.

- **Asian Communities:** Notably, the coefficient for *% Asian* is statistically insignificant across all specifications. Asian-majority tracts exhibit sentiment scores indistinguishable from the White baseline, suggesting the infrastructure gap is strictly segmented by specific racial lines.

3. Income vs. Wealth Effects: We include both $\ln(\text{Median Income})$ and $\ln(\text{Home Value})$ to disentangle liquid resources from asset wealth. We find that sentiment is driven primarily by income ($\beta \approx 0.11^{**}$), while home value remains largely insignificant. This suggests that park quality is also responsive to the immediate socioeconomic status of residents (and potentially their capacity for civic advocacy).

4. Income vs. Wealth (Multicollinearity): We include both $\ln(\text{Median Income})$ and $\ln(\text{Home Value})$ to disentangle liquid resources from asset wealth. As expected, these variables are highly correlated ($r > 0.7$). The results show that Income absorbs the explanatory power ($\beta \approx 0.11^{**}$), rendering Home Value insignificant. This suggests that *Income* serves as the primary proxy for socioeconomic status in this context, capturing the community's capacity for civic advocacy and tax-base stability more effectively than passive real estate values.

Table 2: **The Divergence: Determinants of Access vs. Quality**

Dimension	(1) Access	(2) Cleanliness	(3) Children	(4) Homeless
$\ln$(Med. Income)	0.012*	0.114**	0.125**	0.041*
	(0.007)	(0.042)	(0.045)	(0.022)
$\ln$(Home Value)	0.001	0.021	0.015	0.008
	(0.004)	(0.035)	(0.032)	(0.018)
Reference: % White				
% Black	-0.002	-0.124***	-0.145***	-0.051**
	(0.004)	(0.042)	(0.048)	(0.025)
% Hispanic	-0.001	-0.082**	-0.025	-0.015
	(0.003)	(0.038)	(0.031)	(0.019)
% Asian	0.001	-0.012	0.005	-0.004
	(0.003)	(0.031)	(0.029)	(0.015)
Pop. Density	0.004	-0.015	-0.012	-0.008
	(0.003)	(0.011)	(0.009)	(0.007)
Median Age	0.001	0.003	0.007	-0.002
	(0.001)	(0.005)	(0.006)	(0.004)
County FE	Yes	Yes	Yes	Yes
Observations	2,107	5000	5000	5000
R^2	0.04	0.21	0.19	0.11

Std errors clustered at Tract. *$p<0.1$, **$p<0.05$, ***$p<0.01$
Dep Var Col 1: Density (per 1k). Dep Var Cols 2-4: CPS [-1, 1].

4.4 Implications and Policy Relevance

These results underscore a critical limitation in current urban auditing methods: the "existence" of a park does not imply the "provision" of a public good. The statistical significance of the Black and Hispanic coefficients suggests that inequality is not driven by where parks are built (CapEx), but by how they are maintained (OpEx).

Limitations and Scope of Inference It is imperative to interpret these coefficients within the strict bounds of our data generating process. We identify three primary sources of selection bias that constrain the generalizability of these findings.

1. The "Ground Truth" Assumption (Measurement Error): Our measurement of administrative access (Column 1) relies exclusively on the City78 dataset provided by HUD as a proxy for the universe of green spaces. We lack an external set of "publicly accessible green spaces", as we include them in this study (as opposed to just parks or forests) to validate this coverage. If "digital invisibility" is correlated with socioeconomic status—i.e., if parks in lower-income tracts are systematically less likely to be indexed on Google Maps—our estimates of park density will be biased downward in those areas. Consequently, the "null result" on access refers strictly to *digitally visible* infrastructure; the physical reality may differ.

2. Spatial Sparsity and Sampling Bias: Our sentiment analysis (Columns 2–4) utilizes a stratified random sample of 5,000 reviews drawn from a corpus of 1.5 million, due to multiple resource constraints at both a computational and personnel level. While stratification corrects for the "J-shaped" rating distribution, it alters the natural probability weights of the population. Furthermore, the sample is spatially sparse; not every census tract in California is represented. As a result, our coefficients (β) should be interpreted as estimating the *gradient of user sentiment* among the population of *reviewed* parks, rather than an average of *all* parks in California. We are effectively modeling the "conditional expectation of a digitally engaged user," which may differ from the experience of the average resident, particularly in areas with lower digital literacy.

3. Orthogonality of Homelessness: Finally, the lower explanatory power (R^2) and marginal significance in the *Homelessness* specification suggest that this specific phenomenon is driven by factors orthogonal to tract-level income, such as urban centrality, shelter density, or transit accessibility. Linear demographic modeling is likely insufficient to capture the complex spatial dynamics of the housing crisis.

Nevertheless, these findings lend evidence to the promise of NLP as a high-frequency policy tool. Even with a limited sample, the model successfully recovered verifiable patterns of structural neglect that aggregate administrative data masked.

5 Conclusion

This study demonstrates that Natural Language Processing should be a serious method of consideration to "audit" the quality of public goods at a scale impossible for traditional survey methods. By moving beyond simple star ratings to aspect-based sentiment, we revealed that the *experience* of public space is stratified by income and race. While the state may provide the land, the "shadows" of neglected maintenance and safety concerns are unevenly cast, disproportionately affecting lower-income communities. Future work should focus on scaling this methodology to a much larger dataset with the appropriate computational resources and real-time municipal dashboards, allowing city planners to react to accurately identified, infrastructure decay before it becomes entrenched.

References

[1] Talen, E. (1998). Visualizing fairness: Equity maps for planners. *Journal of the American Planning Association*, 64(1), 22–38.

[2] Boone, C. G., Buckley, G. L., Grove, J. M., & Sister, C. (2009). Parks and people: An environmental justice inquiry in Baltimore, Maryland. *Annals of the Association of American Geographers*, 99(4), 767–787.

[3] Dai, D. (2011). Racial/ethnic and socioeconomic disparities in urban green space accessibility: Where is the environmental justice? *Landscape and Urban Planning*, 102(4), 234–244.

[4] Rigolon, A. (2016). A complex landscape of inequity in access to urban parks: A literature review. *Landscape and Urban Planning*, 153, 160–169.

[5] Mao, L., Wu, C., & Wang, L. (2022). A comparison of perceived and geographic access to predict urban park use. *Cities*, 124, 103597.

[6] Liu, Y., Liu, X., Gao, S., Gong, L., Kang, C., Zhi, Y., Chi, G., & Shi, L. (2015). Social sensing: A new approach to understanding our socioeconomic environments. *Annals of the Association of American Geographers*, 105(3), 512–530.

[7] Glaeser, E. L., Kim, H., & Luca, M. (2022). Nowcasting the local economy: Using Yelp data to measure economic activity. In *Big Data for Twenty-First-Century Economic Statistics* (pp. 369–392). University of Chicago Press.

[8] Pontiki, M., et al. (2016). SemEval-2016 task 5: Aspect based sentiment analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)* (pp. 19–30).

[9] He, P., Liu, X., Gao, J., & Chen, W. (2021). DeBERTa: Decoding-enhanced BERT with disentangled attention. *International Conference on Learning Representations (ICLR)*.

[10] Massive Data Institute. (n.d.). *Environmental Impact Data Collaborative (EIDC)*. Georgetown University. Retrieved from `https://mdi.georgetown.edu/eidc/`

[11] Massive Data Institute. (n.d.). *Greenspace Research Challenge*. Georgetown University. Retrieved from `https://mdi.georgetown.edu/pbi/greenspace/`

[12] Wen, M., Zhang, X., & Harris, C. D. (2013). Spatial disparities in the distribution of parks and green spaces in the USA. *Annals of Behavioral Medicine*, 45(suppl_1), S18–S27.

[13] Rigolon, A., et al. (2021). Green space and health equity: A systematic review on the potential of green space to reduce health disparities. *International Journal of Environmental Research and Public Health*, 18(5), 2563.

[14] Vaughan, K. B., et al. (2025). Hiding access to quality parks: A case study examining ParkServe and objective park assessment in an under-resourced community. *Cities & Health*, 1–12.

[15] Trust for Public Land. (2024). *The ParkScore Index: 2024 Report*. Retrieved from `https://www.tpl.org/parkscore`