

Investigating the Correlation Between Urban Greenspace and Mental Health Through Geospatial and NLP Analysis



Saanvi Kakde, Computer Science
Mentor: Adwith Malpe, Assistant Teaching Professor
Ira A. Fulton Schools of Engineering

Introduction

Prior research suggests exposure to nature can prevent adverse mental health effects; however, most studies rely on surveys that are difficult to scale and prone to self-reporting bias.

Can pretrained NLP models combined with NDVI-based vegetation mapping model the relationship between urban greenspace and mental well-being?

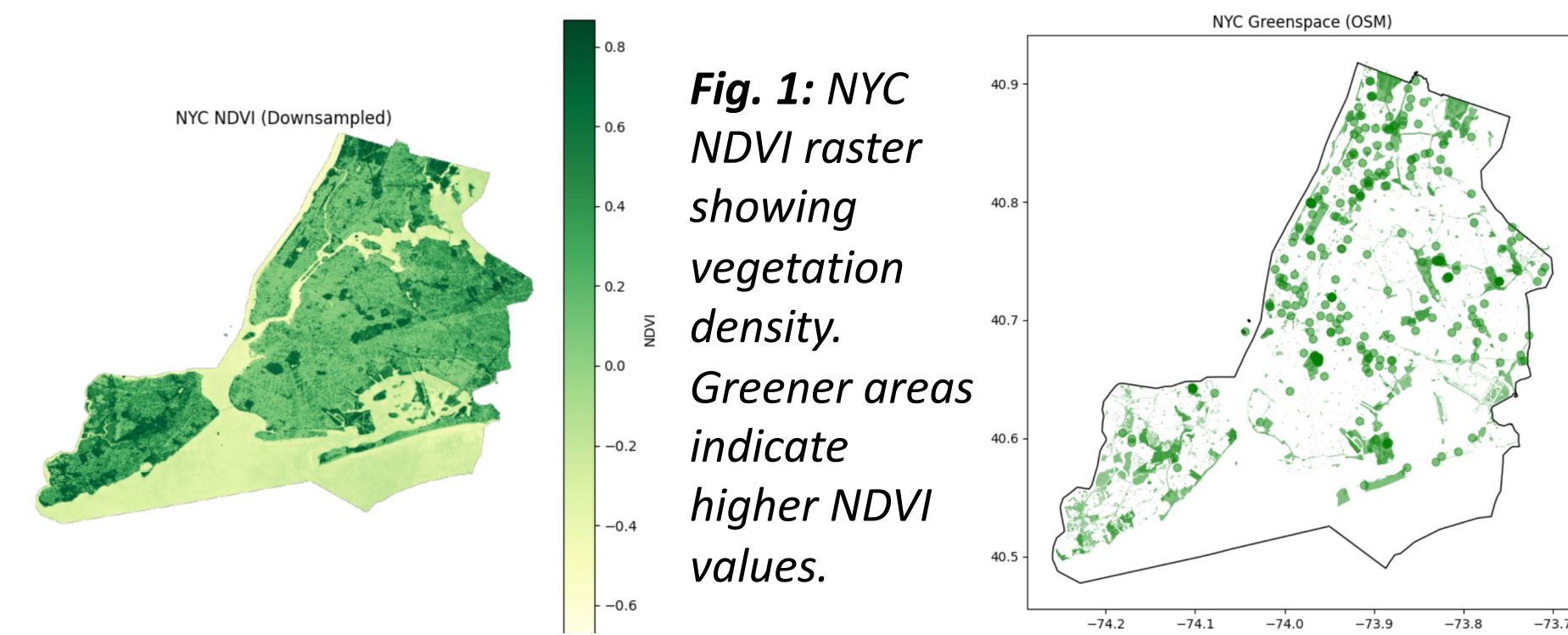


Fig. 1: NYC NDVI raster showing vegetation density. Greener areas indicate higher NDVI values.

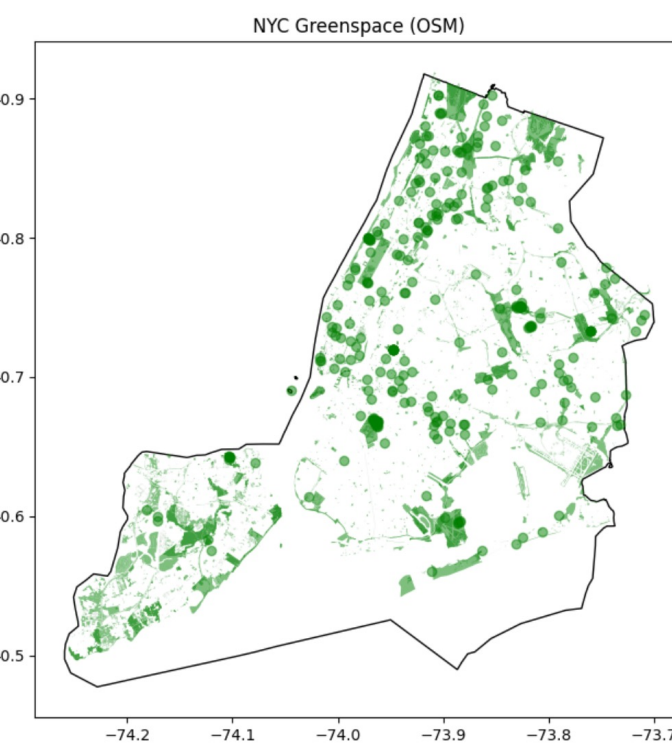


Fig. 2: NYC greenspace polygons extracted from OpenStreet Map

Methods

1. Data Collection

- Sentinel-2 imagery (Google Earth Engine) to obtain NDVI values
- Greenspace data from OpenStreetMap (OSMnx)
- 20,000 geo-tagged NYC tweets
- Google Maps API reviews

2. Sentiment Analysis

- Cardiff NLP's Twitter-RoBERTa-base model classifies tweets
- BERT classifiers applied to extract sentiment scores

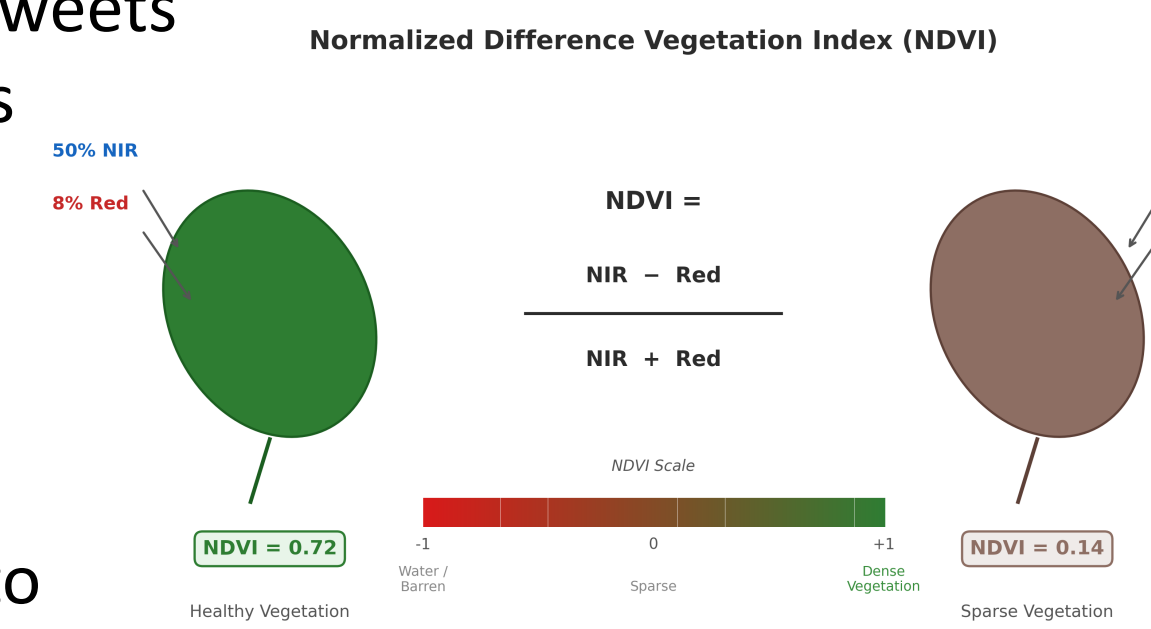


Fig. 3: NDVI measurement

3. Spatial Integration

- Datasets spatially joined to NDVI coordinates using GeoPandas
- Regression modeling applied to evaluate whether vegetation density predicts sentiment patterns across cities

Results

- NDVI vs. tweet sentiment: $r = 0.004$, $p = 0.594$
- Median NDVI nearly identical across sentiment classes: **(0.123)**, **(0.117)**, **(0.125)**
- Tweet density concentrated in NDVI areas < 0.2
- NYC tweet mean sentiment: **+0.064**
- Reviews positivity score: NYC **91%**, Chicago **97%**, Maricopa **88%**;
- Average greenspace Google Maps star-rating exceeded **4.5/5**

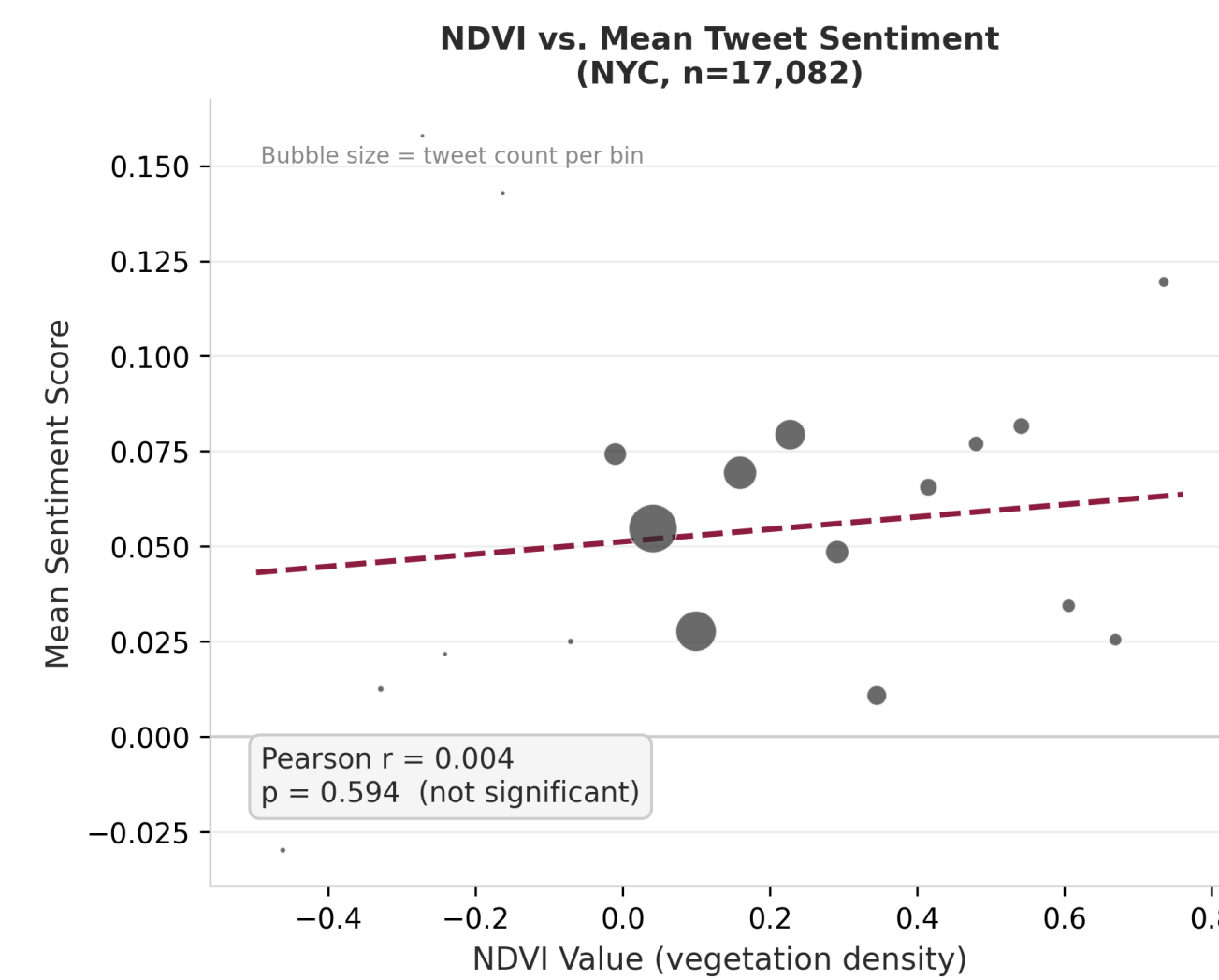


Fig. 4: NDVI vs. mean tweet sentiment binned across 20 NDVI intervals. Bubble size reflects tweet count per bin.

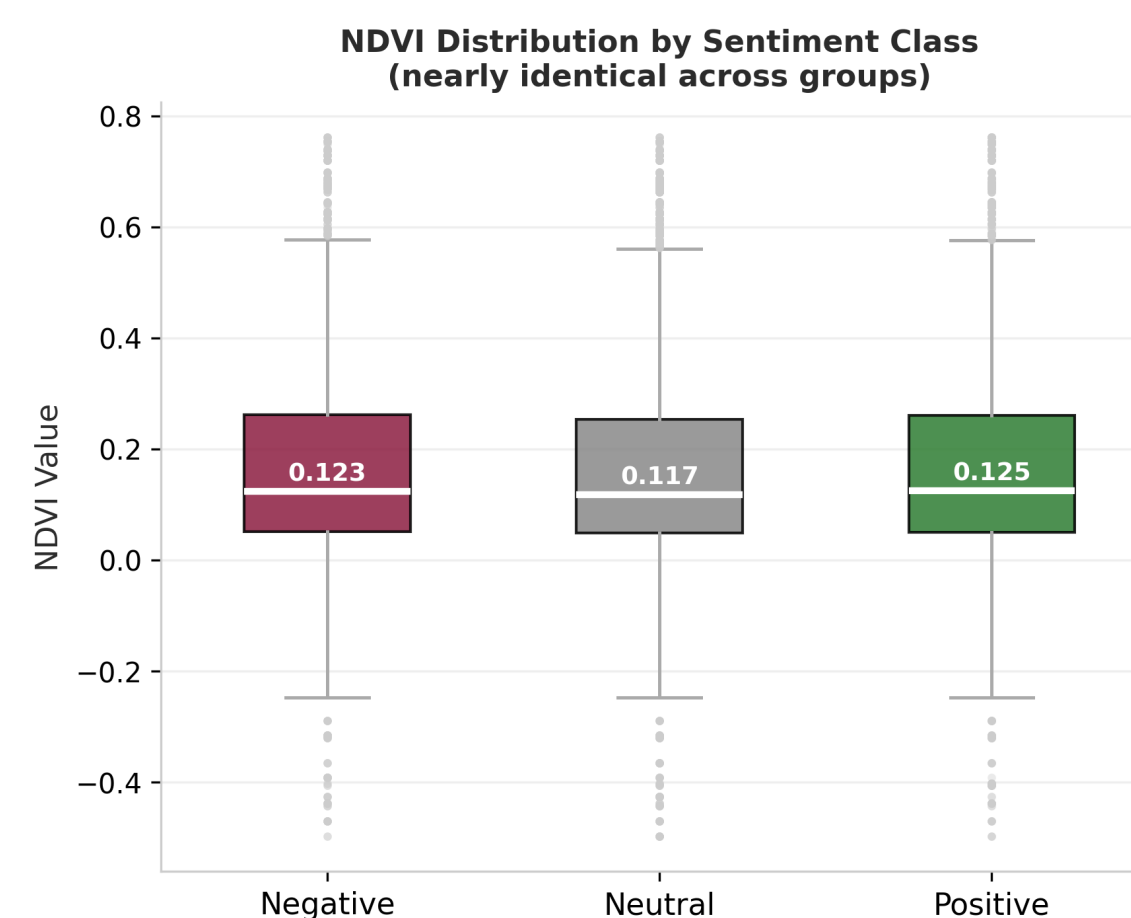


Fig. 5: NDVI distribution by sentiment class.

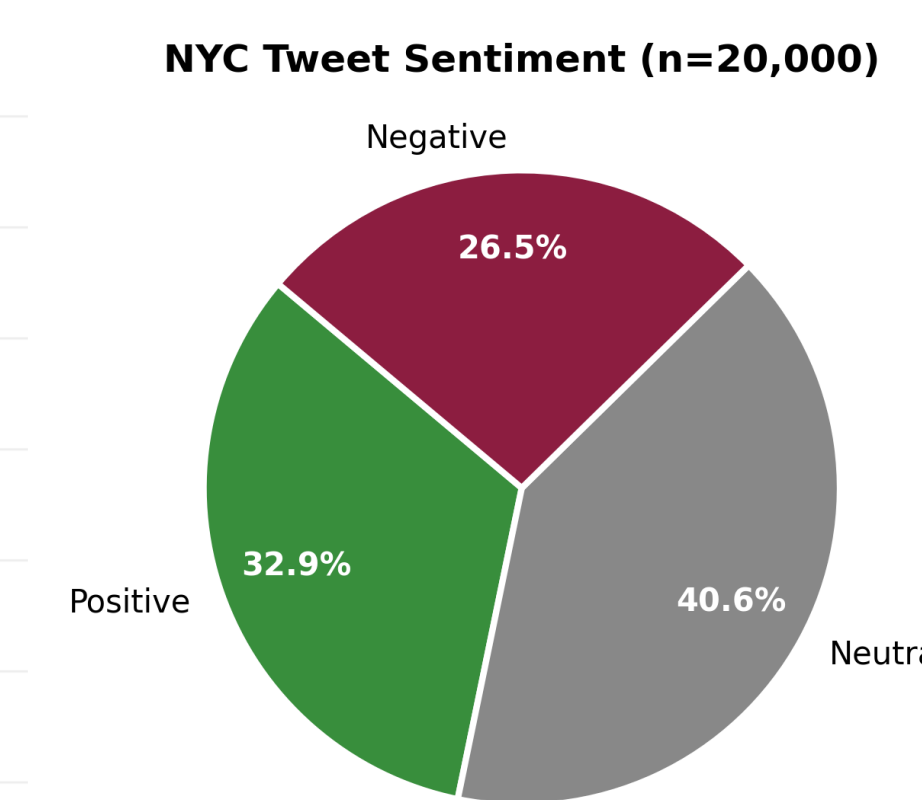


Fig. 6: NYC tweet sentiment distribution (n = 20,000) classified using Cardiff NLP's Twitter-RoBERTa model.

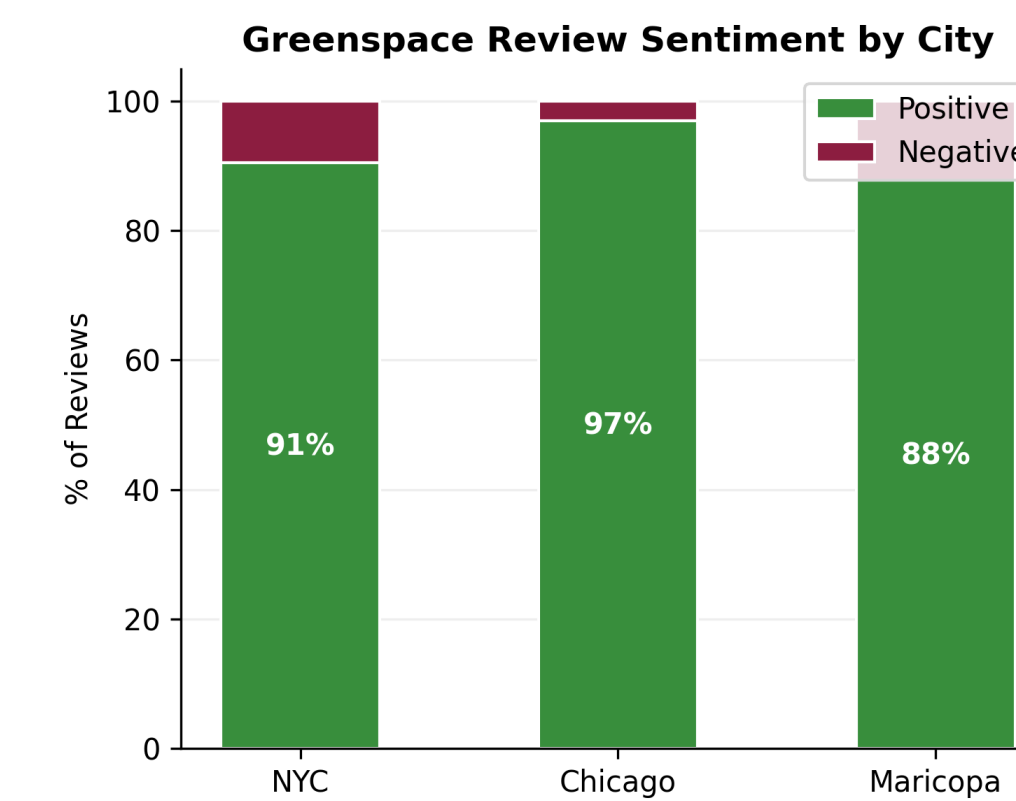


Fig. 7: Google Reviews sentiment by city for greenspace locations. Positive sentiment dominated across all three cities.

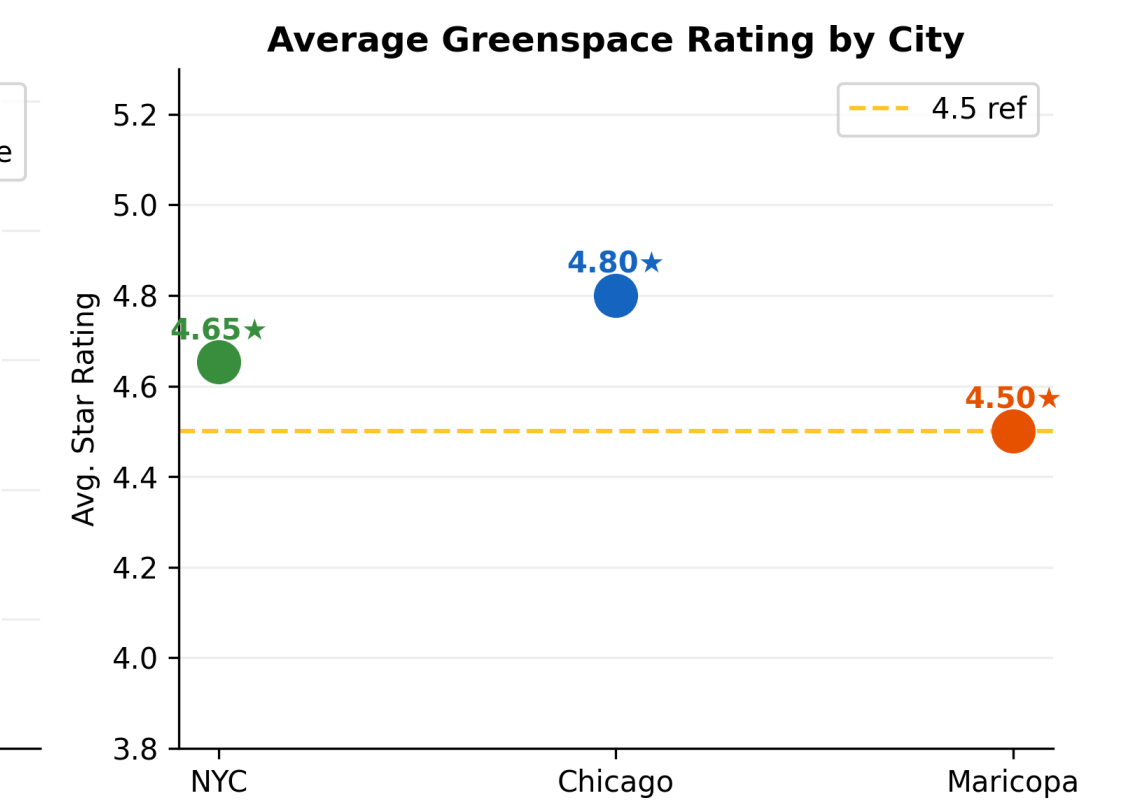


Fig. 8: Average star rating by city. All cities exceeded 4.5/5, indicating strong positive response to greenspace when actively visited.

Conclusions

- NDVI and tweet sentiment showed **no significant correlation**, confirming the null hypothesis
- Near **identical NDVI medians** suggest greenspace density does not predict emotional tone in passive social media
- Passive geo-tagged text may not reliably capture well-being: a key methodological finding
- Tweet concentration in **low-NDVI areas** limits statistical power across the greenspace spectrum given area
- High greenspace ratings across all cities suggests **overall positive response to greenspace** when actively visiting it
- Findings inform future research design and support **evidence based** urban planning and public health policy

Future Work

- Incorporate **higher resolution NDVI raster data** at census-level to improve spatial granularity
- Expand to additional cities to improve quality of data
- Apply **spatial autocorrelation** and fine-tune sentiment models on mental health-specific language
- Compare sentiment across **high- vs. low-income** areas to inform public health equity research