Watts on the Roof?

An Analysis of Residential Rooftop Solar in the Greater Boston Area Using Machine Learning Classification and Linear Regression

By Aidan Banerjee

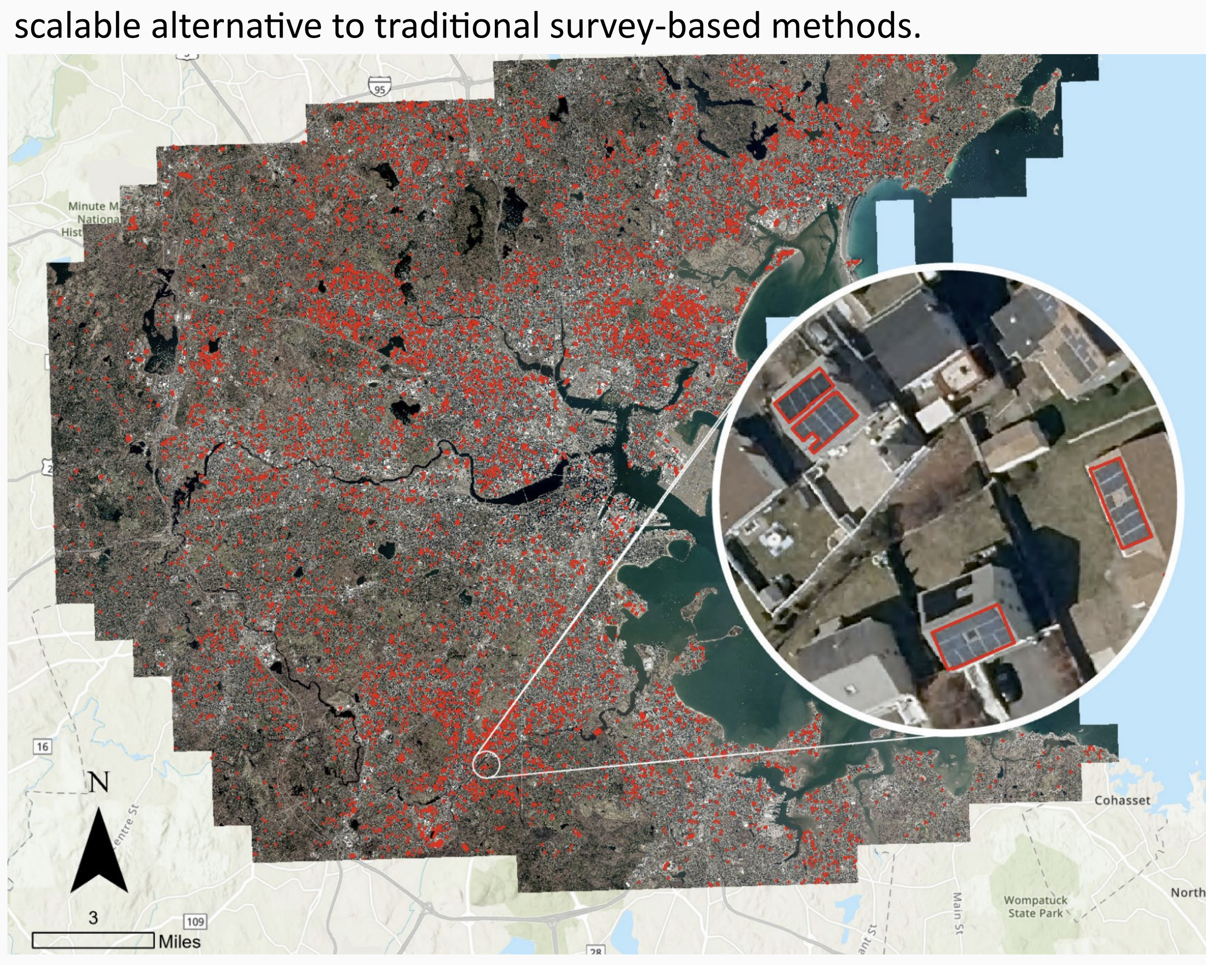
Introduction

This project examines the spatial distribution of residential rooftop solar adoption across towns within the I-495 Beltway of Massachusetts, using 2023 image data as a snapshot year. Leveraging a high-resolution solar panel identification model developed by ESRI, rooftop solar installations were mapped from satellite imagery and analyzed for geographic clustering using Local Moran's I — a spatial autocorrelation method that identifies statistically significant hot and cold spots. Solar patters were also compared to median household income at the census tract level, using data from the American Community Survey (ACS). By integrating solar infrastructure data with localized income levels, this study aims to assess whether wealthier communities are more likely to participate in the transition to clean energy, and to identify potential disparities in solar access across the region.

Model Description

The ESRI Solar Panel Detection model is a deep learning-based tool designed to automate the identification of rooftop solar panels in high-resolution aerial or satellite imagery. Built on the Mask R-CNN architecture and implemented through the ArcGIS API for Python, the model is capable of detecting individual solar panels by generating bounding boxes around them (as shown in the image below).

It accepts inputs such as raster images, mosaic datasets, or image services with a spatial resolution of 5–15 centimeters, and outputs a feature class delineating detected panel locations. Designed for U.S. geographies, the model achieves an average precision of 0.76, making it a reliable solution for large-scale solar infrastructure mapping. Its GPU-accelerated inference capabilities allow for efficient processing of large image sets, offering a scalable alternative to traditional survey-based methods.



Project Application

To run the model, I used 2023 aerial imagery from MassGIS, which provides statewide satellite photographs at a 15 cm spatial resolution. These image tiles were first downloaded using a custom Python script that scraped and unpacked .zip files directly from the MassGIS server. After download, the tiles were mosaiced together in ArcGIS Pro to create a unified raster suitable for model input. Processing the full dataset through the ESRI model required approximately 12 hours of compute time, and yielded over 12,500 detected solar panels across the study area.

The ESRI model provides a powerful and scalable approach to solar panel detection, however, several limitations should be noted. The model identifies individual solar panels, not entire systems, which can complicate comparisons with installer-reported or utility-scale data. False positives—such as skylights or reflective surfaces—may inflate counts, while false negatives can occur when panels are obscured by shadows or roof angles.

Despite these challenges, this method remains highly effective. Traditional sources such as permit records, utility interconnection data, or surveys are often incomplete, inconsistent, or difficult to access. By automating detection through high-resolution imagery, this approach enables consistent, replicable measurement of solar adoption across large geographic areas.

Aidan Banerjee **GIS-102-01 Advanced GIS | Spring 2025**

Data Sources: MassGIS Data: 2023 Aerial Imagery| Socio-economic data: American Community Survey 1-year 2023 estimates, census Census tracts | ESRI

Projection System: NAD 1983 StatePlane Massachusetts FIPS 2001 (Meters)

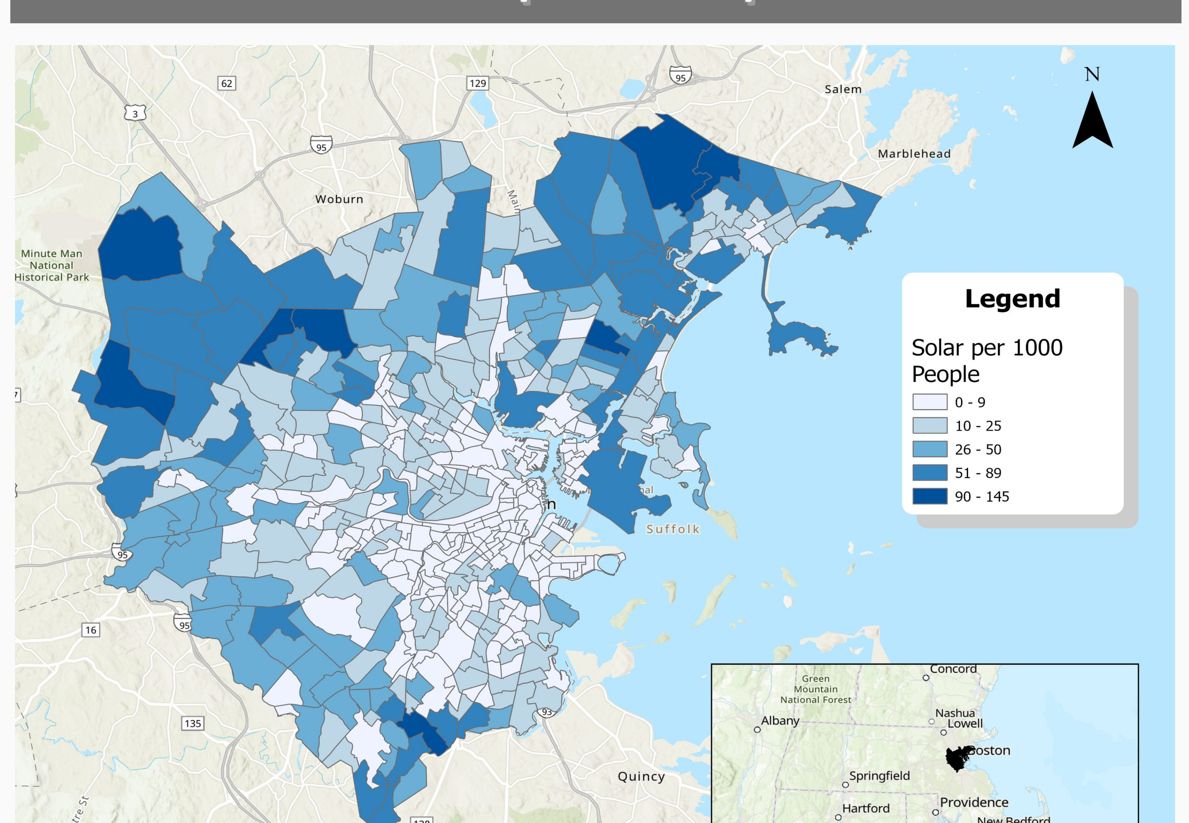
Methodology

- Using a custom python script, satellite imagery data was scraped off of the MassGIS Aeriel 2023 Image database.
- The images were then mosaiced together into a singular raster on ArcGIS Pro
- Running ESRI's solar classification model, bounding boxes around positively identified solar panels are outputted and formatted into point data for analysis
- Using Census population data, the number of solar panels was normalized and aggregated by census tracts
- To test for statistically significant results, local moran's I was used to identify census tracts with relative high or low solar adaptation and compared to income data.
- Shapefiles for ACS medium household income were compiled and joined to the current solar count tract data and a linear regression was conducted.

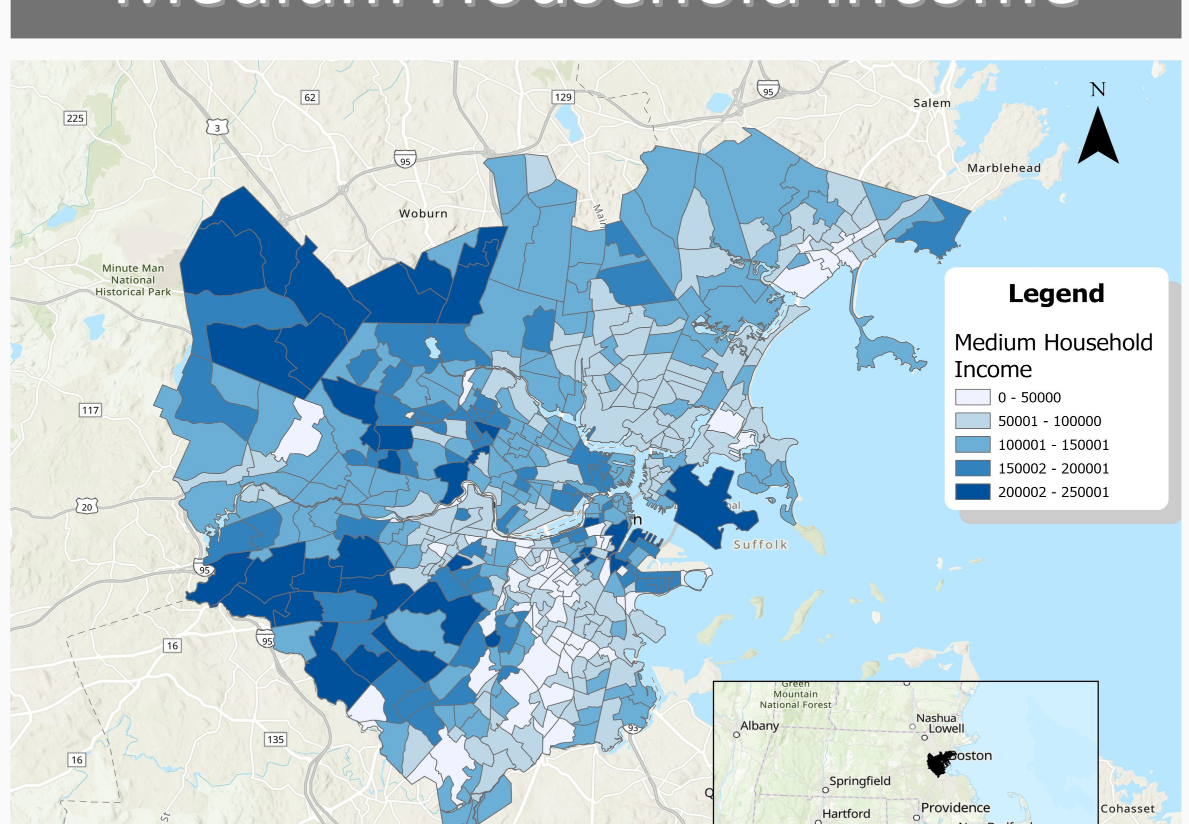
Results

The aggregation of solar panel counts by census tract—normalized using population data—revealed a clear concentration of per-capita rooftop solar adoption in the suburban areas surrounding Boston. Even after normalization, solar installations were more prominent on the outskirts of the city, particularly in wealthier suburban towns along the 495 Beltway. This spatial trend closely mirrored the distribution of median household income, with higher-income census tracts located outside of downtown Boston also exhibiting higher per-capita solar adoption.

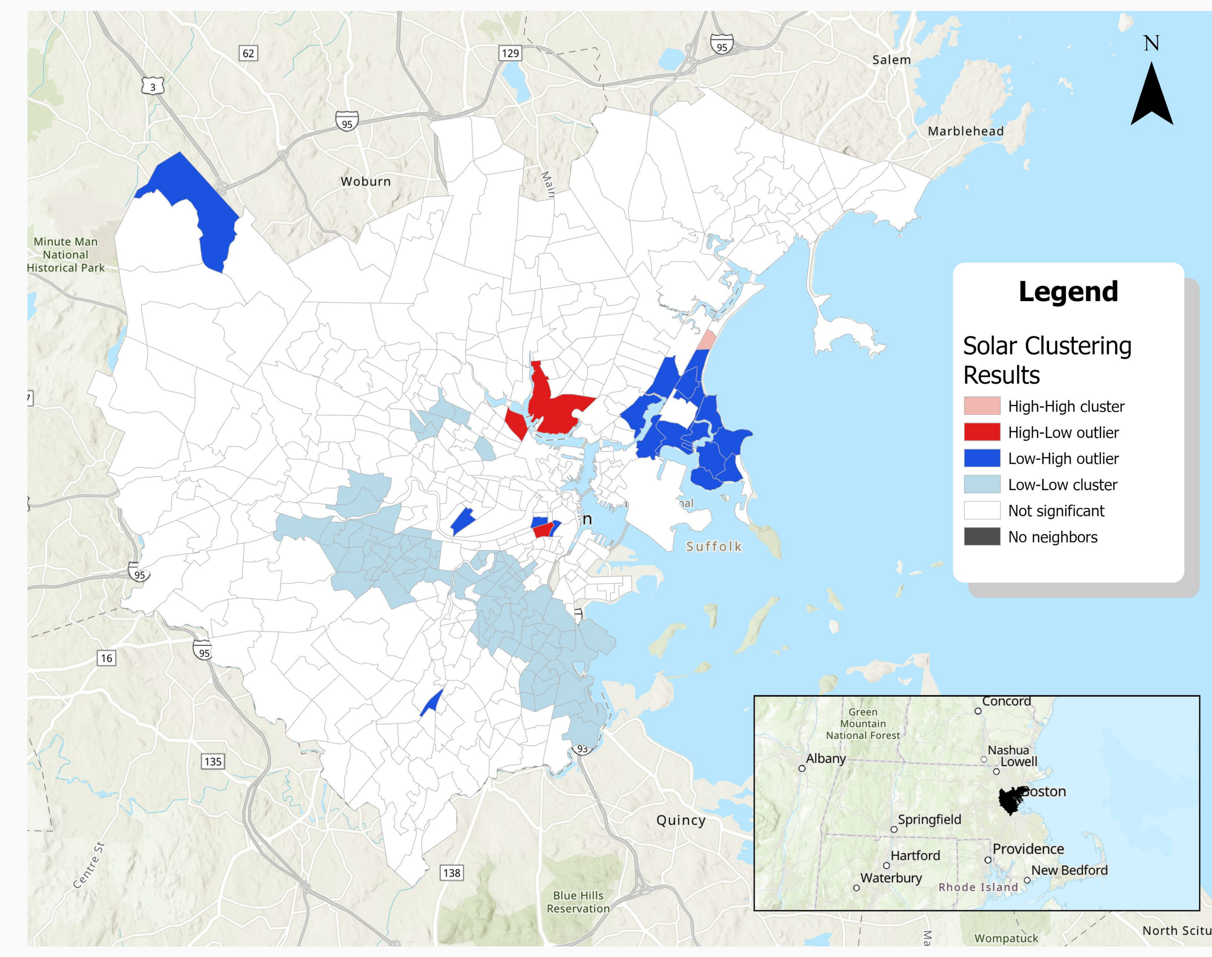
Solar per Capita



Medium Household Income



Local Moran's I Results



Despite the correlative trend in medium income and solar adaptation, the Local Moran's I analysis of solar panel distribution showed fewer statistically significant high-high clusters in these outer tracts than expected. In contrast, a distinct low-low cluster appeared in downtown Boston, indicating a statistically significant grouping of census tracts with low solar adoption surrounded by similarly low-adopting neighbors. This may reflect urban barriers to solar deployment such as denser building stock, limited roof access, and higher proportions of renters versus homeowners.

	Coefficient	Std Deviation	t statistic	Probability
Medium Household Income	0.000086	0.000024	3.651849	0.00029
Intercept	13.895298	2.930407	4.741763	0.000003

A linear regression analysis confirmed a statistically significant relationship between median household income and solar adoption. With a coefficient of 0.000086 and a p-value of 0.00029, the model indicates that as income increases, so does the number of rooftop solar installations—underscoring the socioeconomic disparity in access to clean energy technologies

Discussion & Next Steps

This project identified a clear and statistically significant linear relationship between median household income and the number of rooftop solar panels across census tracts in the 495 Beltway of Massachusetts These findings are consistent with existing research on energy justice, which often highlights economic privilege as a key driver of access to clean energy infrastructure. By combining machine learning-based detection methods with spatial statistical analysis, this project was able to provide a replicable, data-driven approach to mapping solar adoption and exploring its socioeconomic dimensions.

It is important to acknowledge several limitations that may have influenced the analysis. The ESRI solar detection model likely introduced both false positives and false negatives, skewing results. Additionally, although a strong correlation between income and solar adoption was observed, causation cannot be assumed. However, the spatial consistency of the trend and the strength of the model results lend support to the explanatory value of income as a key indicator of solar access in the region.

To extend this research, future studies could incorporate a multivariate linear regression framework to account for the interaction of multiple factors that influence solar adoption. Methodological improvements could include integrating zoning laws, permit data, and utility interconnection records to better explain solar adoption patterns. Additionally, refining panel detection for accuracy, and expanding the analysis across Massachusetts or to other cities for broader comparison would proved more effective results.

Ultimately, this project establishes a scalable, data-driven framework for analyzing rooftop solar adoption and its relationship to income, offering a powerful tool for assessing clean energy equity. By identifying spatial patterns and disparities, the study helps highlight communities that may be underserved in the renewable energy transition and provides a foundation for more inclusive policy efforts.