

Data Collection Report: Mapbox Satellite Imagery Retrieval for Solar Assessment

This report provides a brief overview of the satellite imagery acquisition workflow using the Mapbox Static Images API for the EcoInnovators Hackathon. The goal was to create a high-resolution, coordinate-based dataset suitable for rooftop solar classification and later machine-learning analysis.

Data Source Overview

The imagery for this dataset was sourced from Mapbox Satellite (satellite-v9), a globally maintained, high-resolution imagery layer. Mapbox provides a stable Static Images API that delivers consistent rooftop-level clarity, supports 1200×1200 pixel Retina-quality outputs, and requires secure, token-based authentication. These factors necessitated a retrieval workflow designed with rate limiting, secure token handling, and robust error management.

Data Collection Methodology

A structured, repeatable workflow was followed to maintain accuracy, consistency, and scalability:

1. Dataset Preparation

A CSV file containing sample ID, latitude, longitude, and solar presence label was validated for formatting and geospatial correctness. Since Mapbox requires longitude-first coordinate order, preprocessing ensured compatibility.

2. Resolution & Style Selection

To obtain rooftop-level clarity needed for solar classification:

- Style: mapbox/satellite-v9
- Zoom level: 19.5
- Output size: 600×600@2x (Retina), generating 1200×1200 pixel tiles

3. URL Construction

Each imagery request was generated dynamically using the Mapbox Static Images endpoint. The script automatically inserted:

- tokenized authentication
- coordinates (longitude, latitude)
- zoom, style, and resolution parameters

4. Automated Retrieval with Smart Resume

A Python automation pipeline handled the retrieval process:

- skipped previously downloaded tiles
- introduced delays to prevent API throttling
- retried failed requests
- standardized filenames using `{sampleid}_{has_solar}.png`

5. Logging & Progress Tracking

Each retrieval attempt was recorded in a log file documenting:

- coordinates
- sample ID
- request status
- filename
- cumulative progress and ETA

Retrieval Summary

- Total images downloaded: 3,000
- Success rate: >99%
- Average speed: 0.5–0.6 images/second
- Image resolution: 1200×1200 pixels
- File size: ~100 KB per tile

Most locations yielded clear, high-resolution tiles suitable for rooftop solar classification tasks.

India-Specific Imagery Observations

Mapbox Satellite imagery for Indian regions generally exhibits strong sharpness in metropolitan areas. Semi-urban regions may reflect slightly older tiles depending on source coverage, but overall consistency was higher than most free imagery sources.

Conclusion

The Mapbox Static Images API provided a reliable, high-quality foundation for generating the solar detection dataset. Its rooftop-level clarity, consistent delivery, and scalable API make it well-suited for research, classification, and modeling workflows.

Future work may incorporate multi-zoom datasets or enhanced commercial imagery sources for improved model training accuracy.