Urban Heat Island Effect Analysis

Analyzing factors influencing the Urban Heat Island (UHI) effect. This presentation outlines our methodology and findings.

Team: Icebreakers 48

Introduction

UHI effect

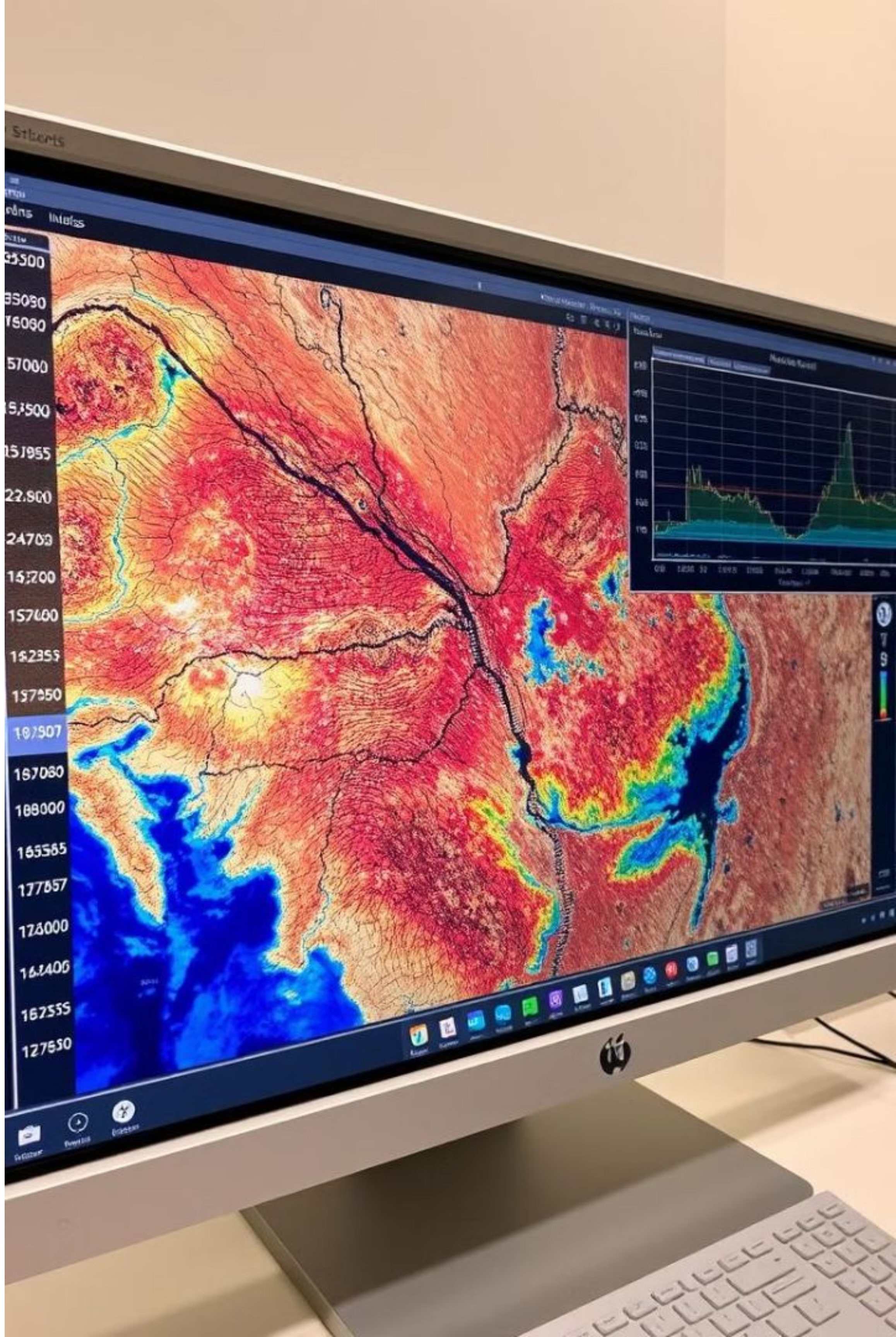
The Urban Heat Island (UHI) Effect is when cities are warmer than nearby rural areas due to heat-absorbing surfaces, reduced vegetation, and human activities. This increases energy use, air pollution, and health risks, but can be mitigated with green spaces, reflective materials, and sustainable urban planning.

Our Goal

We focus on the development of a predictive model for Urban Heat Islands (UHIs) and outlining a practical business plan for its implementation. We will explore the environmental and public health challenges posed by UHIs and demonstrate the potential of our model for mitigating these issues.

Data

Sentinel-2 satellite imagery provides high-resolution multispectral data across a wide range of wavelengths, enabling comprehensive analysis of urban environments. It is based on the Geographical Coordinate System (GCS).



Data : Satellite and Calculated Data

Satellite Data

Spectral Bands

B01 - B12

Obtain all spectral bands from satellite images.

Target

UHI Index

GCS data

Longitude / Latitude

Calculated Data

To discover some hidden information within features and potential factors to improve our model

1. NDVI (Normalized Difference Vegetation Index)

   NDVI = (B08 - B04) / (B08 + B04)

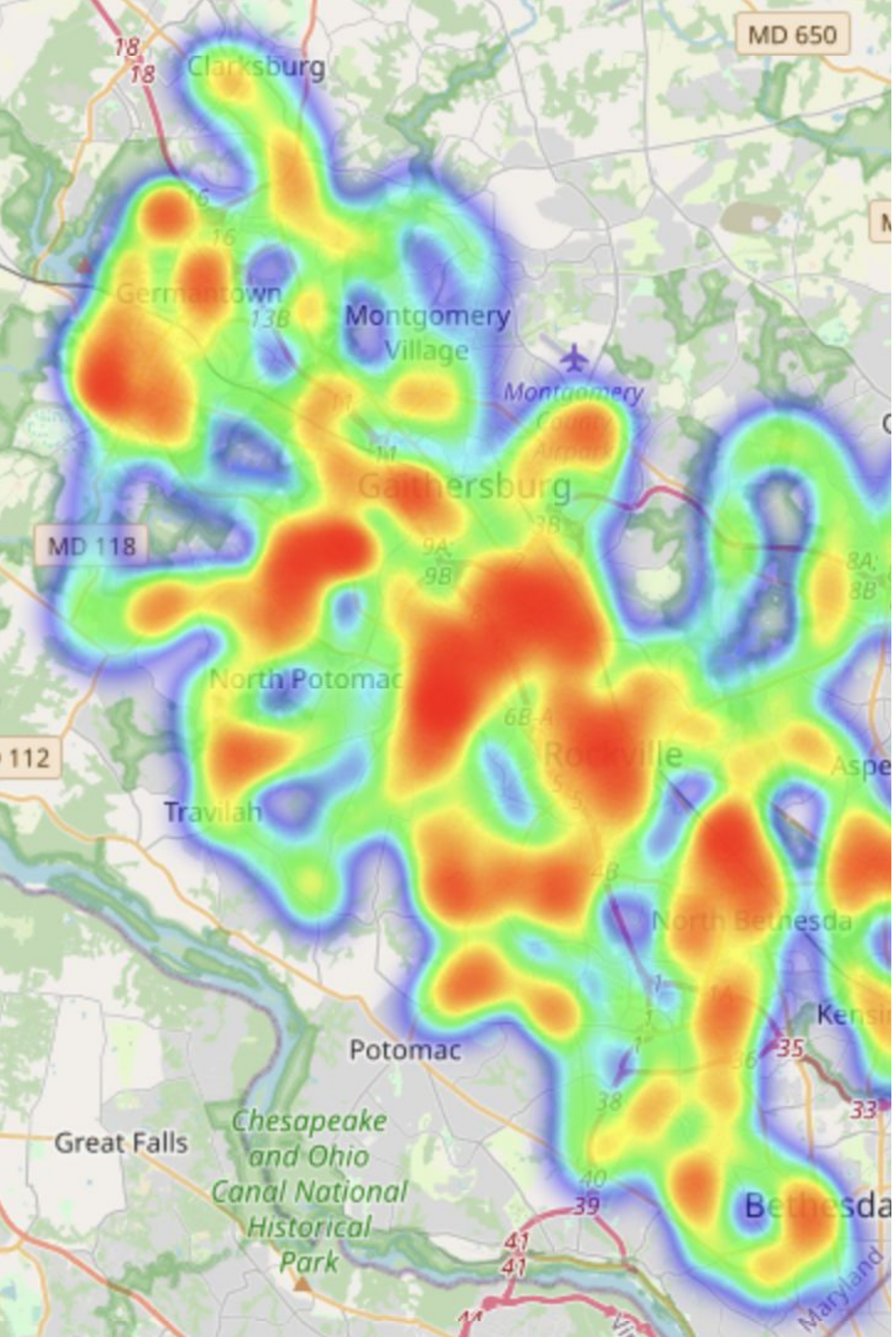
2. NDBI (Normalized Difference Built-up Index))

   NDBI = (B11 - B08) / (B11 + B08)

3. Moisture Content : B11 / B12

4. Soil Ratio (Bare Soil Coverage) B06 / B07

5. Vegetation Health B08 / B02



Exploratory Data Analysis

Seeking Insights

We explored the data using various techniques to intuitively gain insights. This section is designed to help people understand the structure and key features through visualization.

Exploratory Data Analysis

- Total number of data : 30382
- The number of total features : 23
- The spatial range of data :

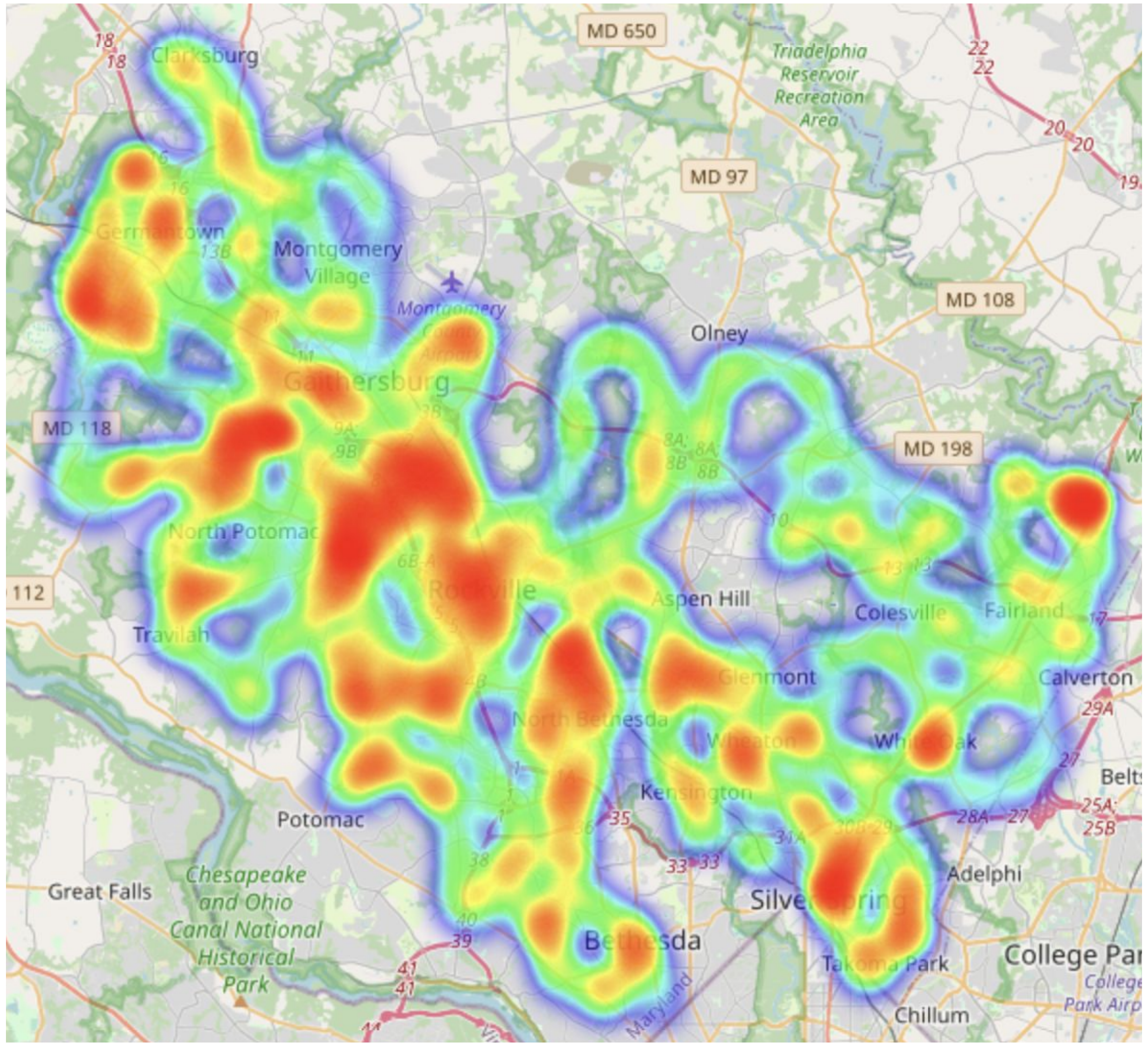
  * Longitude from -77.3027 to -76.929971

  * Latitude from 38.9639 to 39.2316

  A width of 20 miles (32.1826 km) and
  a length of 18.47 miles (29.72 km)
  about **10 m** in a one data pixel

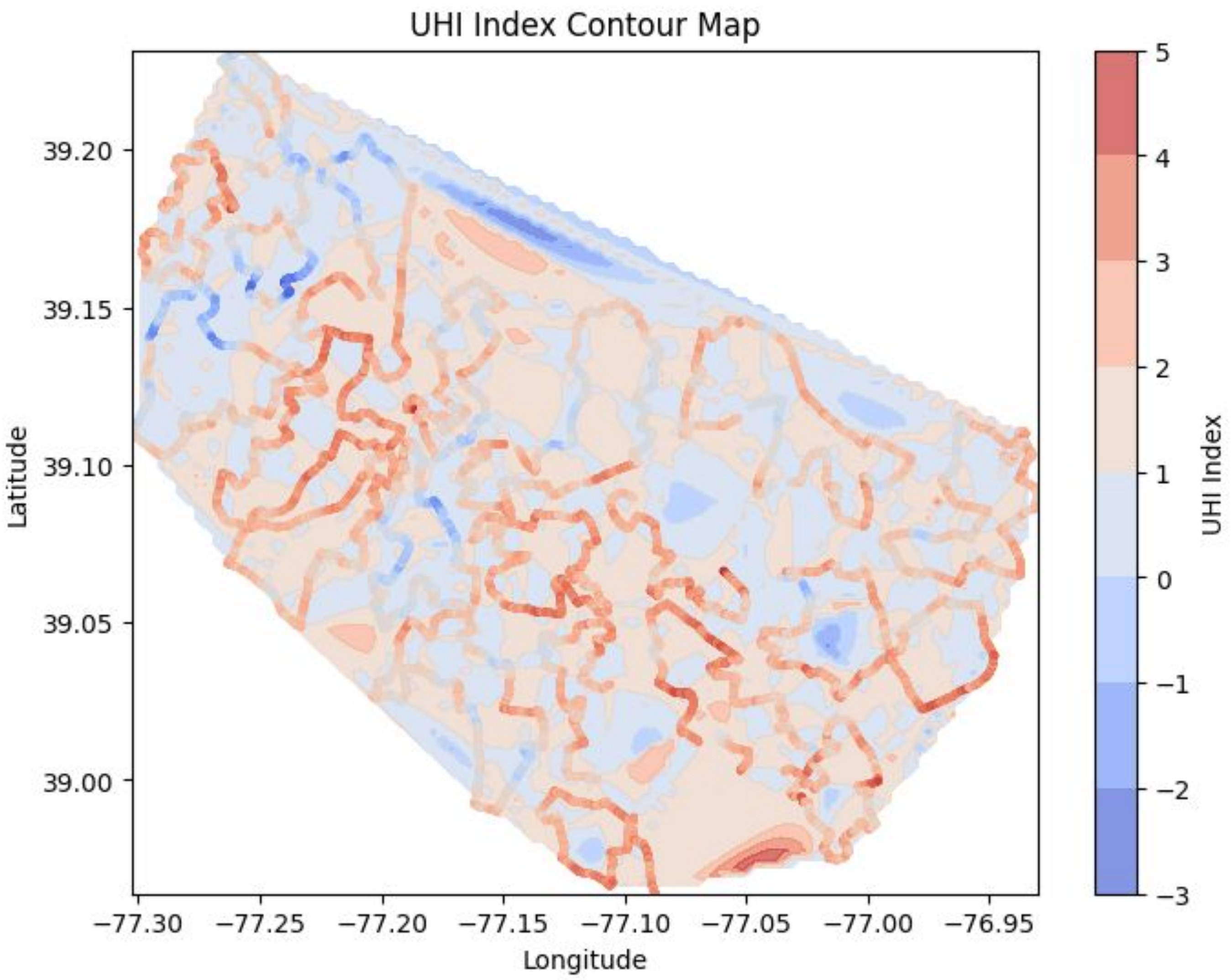
  Too detailed individual pixel-level data

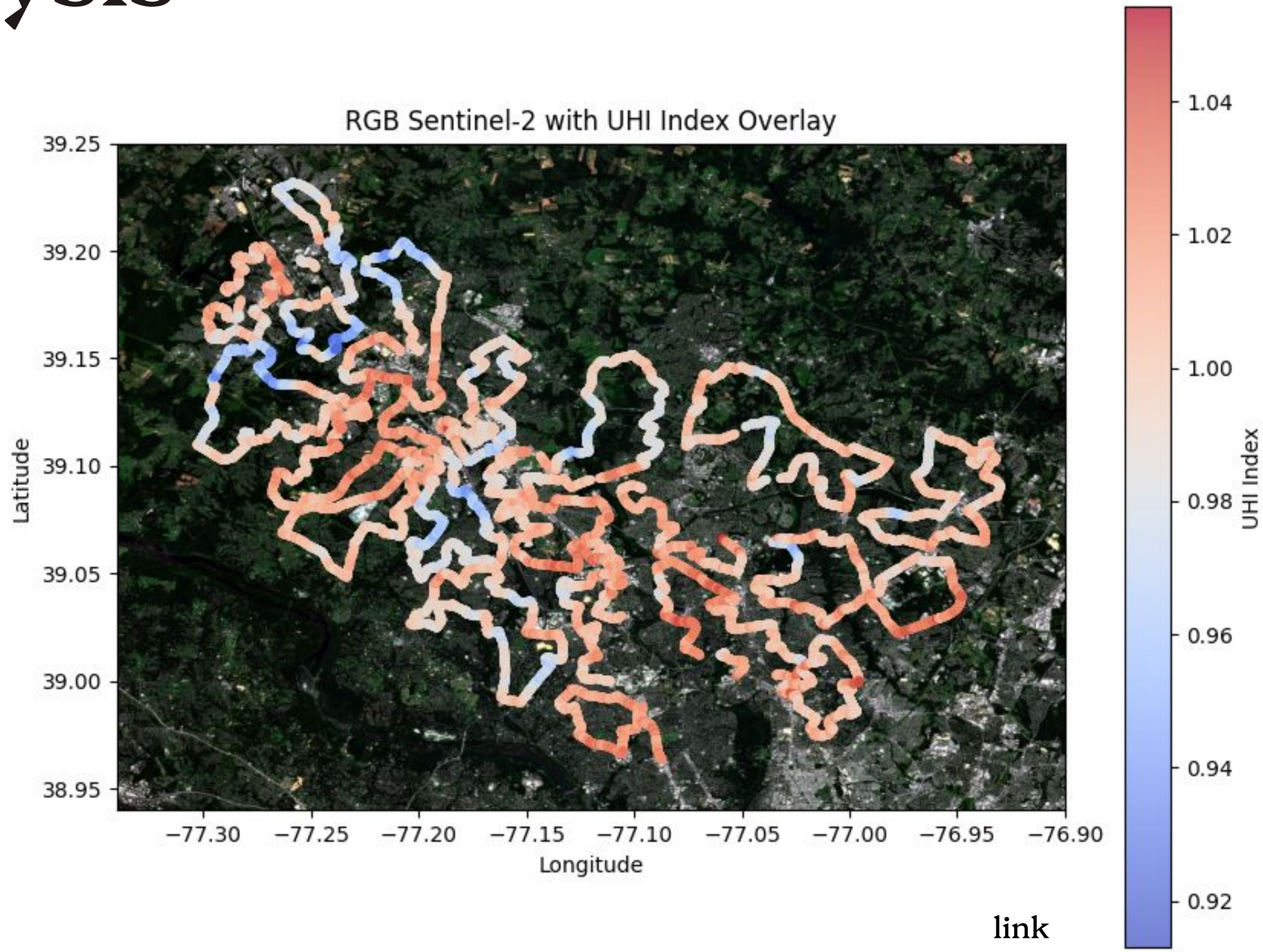
Noise



[Link]

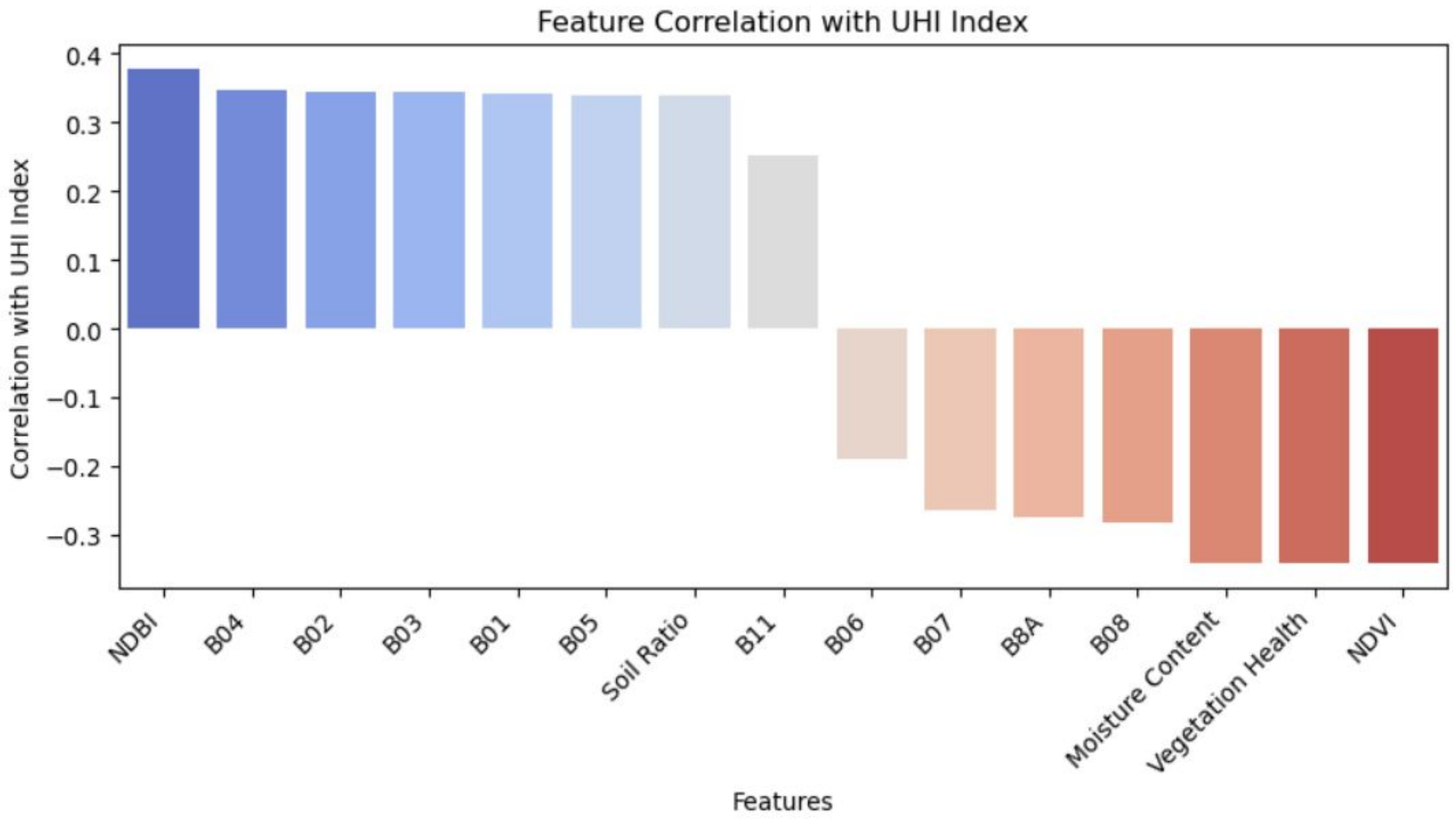
Exploratory Data Analysis





link

Exploratory Data Analysis



Building(NDBI)

Positive correlation. Absorbs and stores heat.

Soil Ratio(Bare soil coverage)

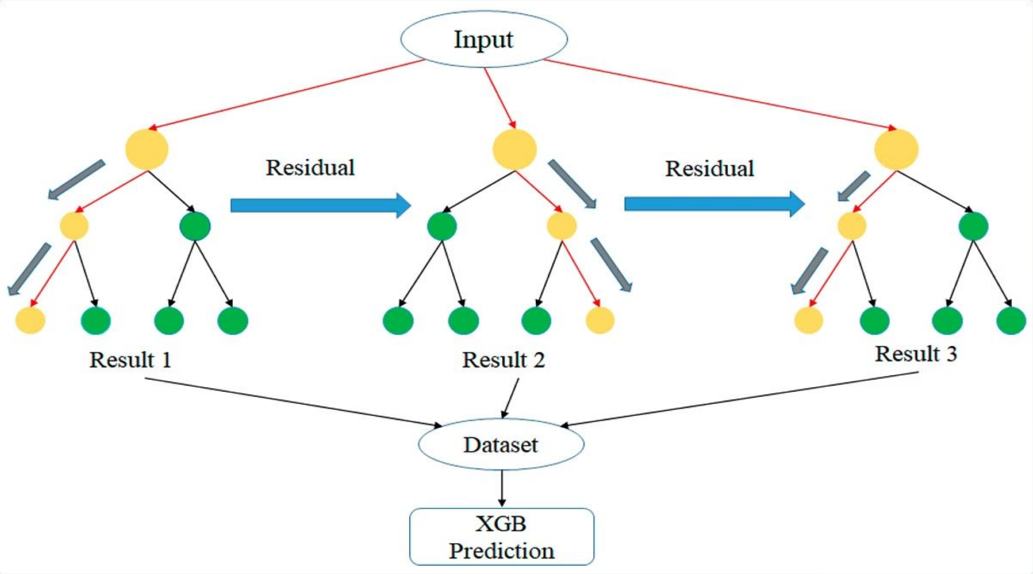
Positive Correlation. Soil stores and releases heat efficiently, which absorbs heat during the day and releases it at night

Moisture Content

Negative correlation. Water cools the surrounding air through evaporation.

Vegetation (NDVI, Vegetation Health)

Negative correlation. Releases moisture and reflects sunlight.



Model Development & Performance

Model Approach

We primarily utilize the **XGBoost** (Extreme Gradient Boosting) algorithm, an advanced machine learning technique known for its high accuracy and ability to handle complex data relationships.

Model Development & Performance

Model Performance

The **XGBoost** model demonstrates improved performance with the highest R^2 value **0.4,** comparing to the base random forest model. This significant improvement indicates the model's ability to accurately predict urban heat island intensity based on the extracted features.

Model Comparison

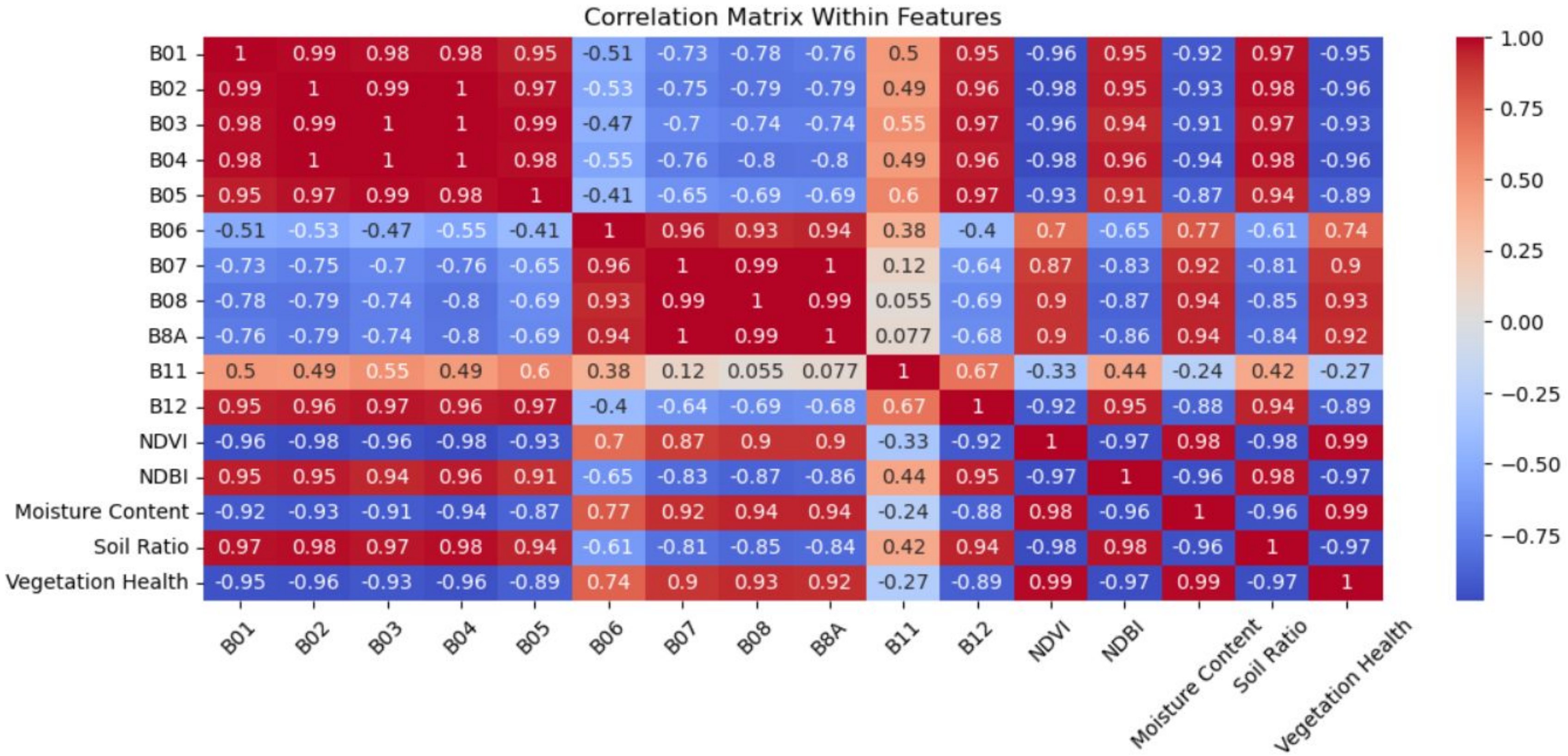
We evaluate other machine learning models (random forest, LightGBM, Tab Transformer, MLP, LinearRegression) against traditional ML approaches. While deep learning models offer potential for improved accuracy, traditional methods like XGBoost prove more resilient in handling the complexity and heterogeneity of urban environments.

Models	Random Forest	LightGBM	Tab Transformer	**XGBoost**	MLP	LinearRegression
R-square score	0.39	0.36	0.08	**0.4**	0.2	0.2

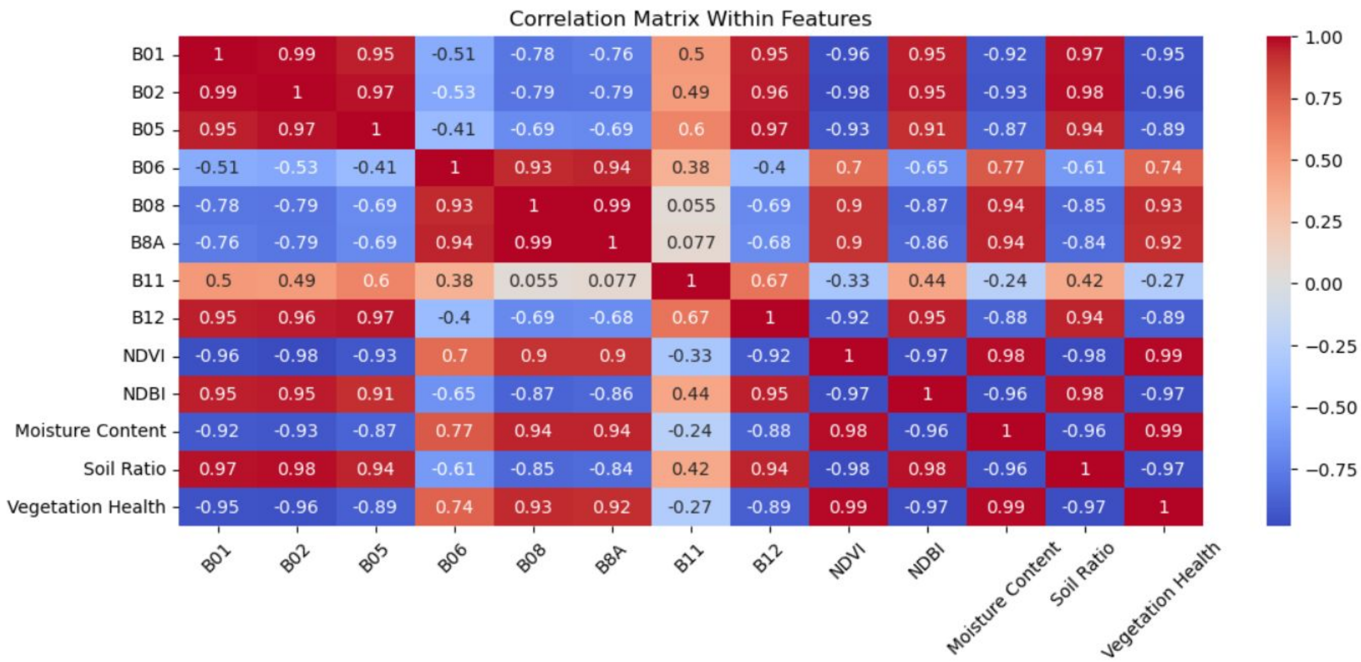
Feature Selection

Heatmap Analysis: Identified highly correlated features.

Before:



After:



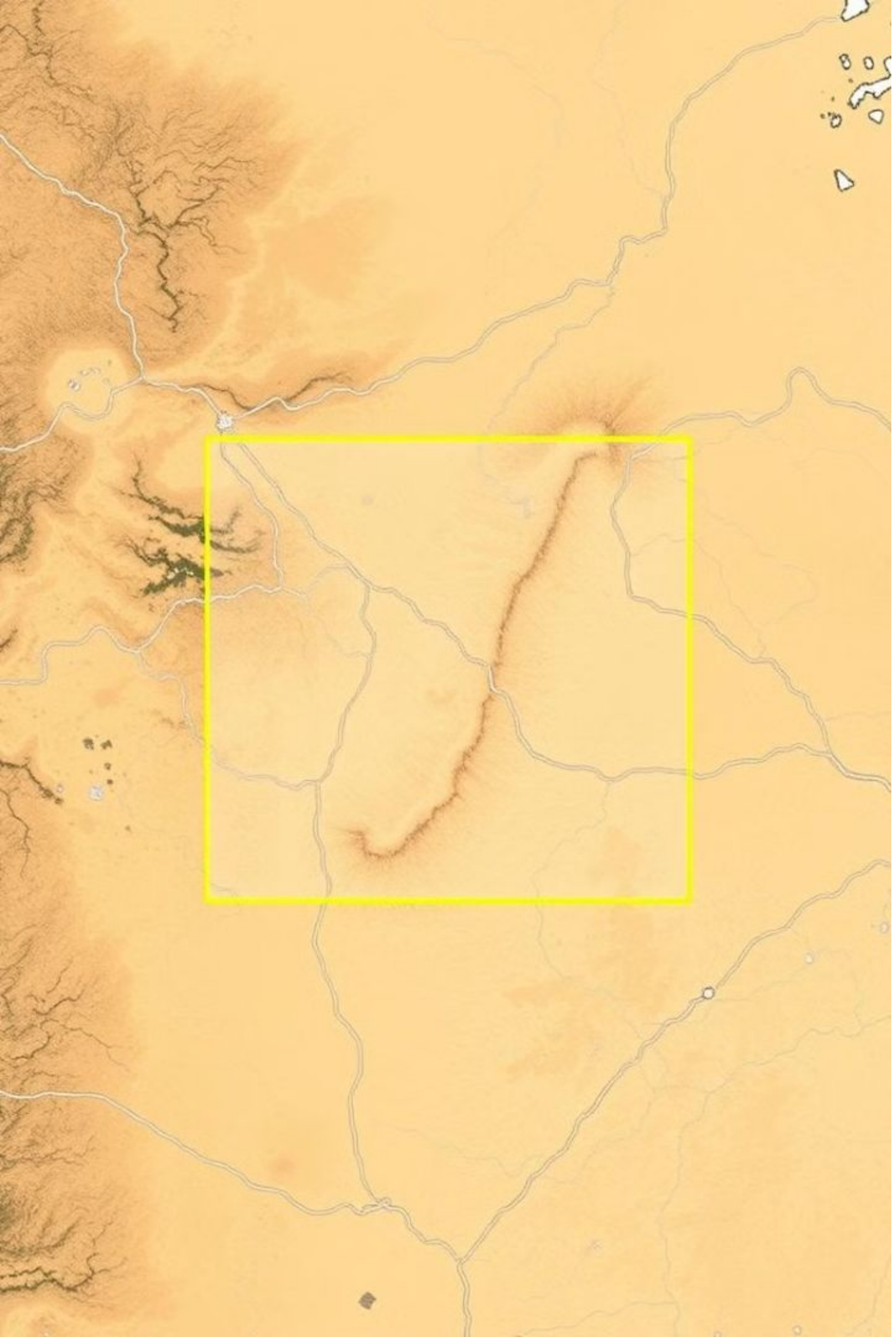
Feature Selection

Advantage

1. Eliminate the redundant variables
2. Improve the efficiency of training process
3. R-square: 0.4 to 0.426(slightly improved)

XGBoost Mode R-square on testing dataset : 0.426 Not Good Enough!

How to Break Through ?



Critical Improvement: Buffer Area Creation

1 Realize UHI Index is related to the environment

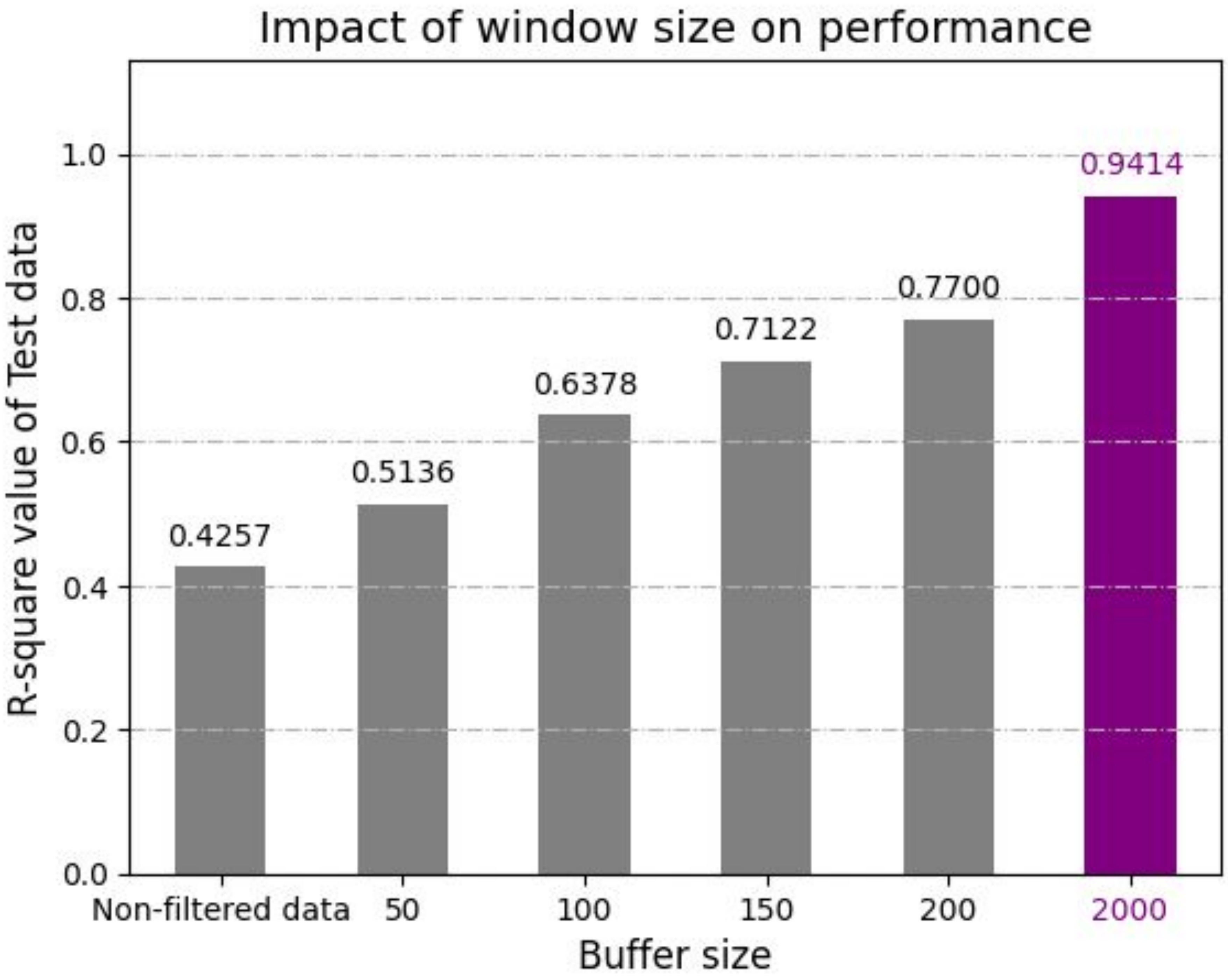
2 Average Spectral Band Data of Each Pixel

Capturing information of surrounding area by using average pixels of specific range.

3 Reduced Variability

Reduces pixel variability.

Decide Buffer Area



R-square:

Increased from 0.426 to 0.9414

Assumption:

Larger buffer area distance include more information

Final Result

XGBoost Model(After using 2000 meters Buffer Size):
R-square = **0.9414**, which is excellent for the future prediction

Limitation

Need to average the spectral band data within 2000 meters from each pixel, when applying this model to some larger areas, we have to consider the **time efficiency** of the process

Business Plan & Model Purpose

- Model vision: Empowering local governments and urban planners.

- Key function: High-resolution temperature mapping to identify Urban Heat Islands (UHI).

Applications: How to Use

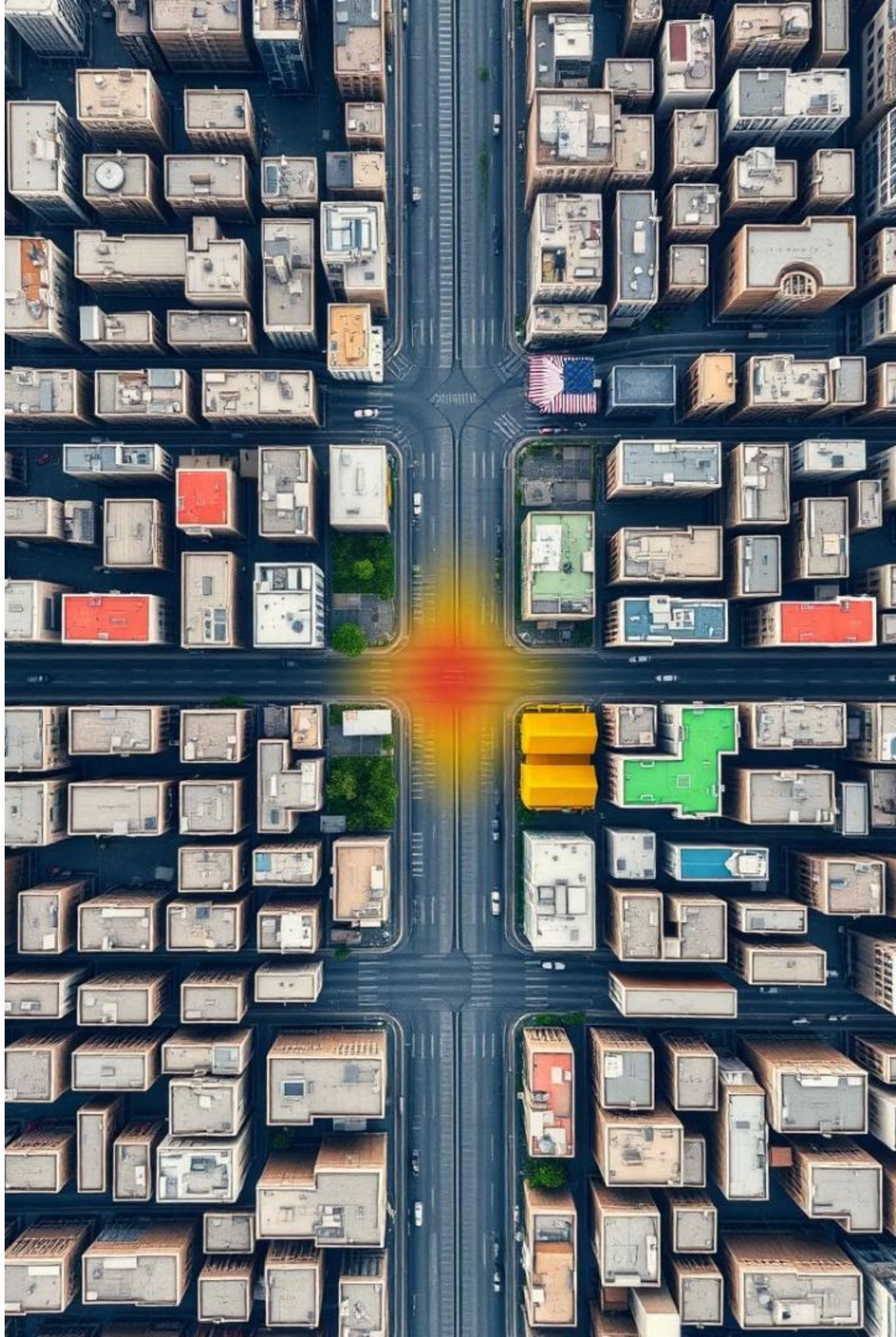
- Identify heat hotspots clearly
- Implement strategic cooling:
  - Green roofs
  - Reflective materials
  - Increased vegetation & water areas

Scaling & Potential for Impact

- Easily scalable using satellite imagery (Sentinel-2)

- Rapid adaptation through transfer learning

- Examples of similar successful approaches:
  - **Phoenix:** Cooling corridors reduced temperatures by 8°F
  - **Singapore:** Green roofs decreased cooling energy by 15%
  - **Barcelona:** Community cooling strategies reduced heat inequality



Addressing Socioeconomic Impacts

- Identify vulnerable communities
- Prioritize green infrastructure & cooling centers
- Promote fairness in climate adaptation policies

Reliable Energy Management

- Forecast heat-driven energy spikes
- Optimize grid resources proactively
- Avoid shortages, improve energy efficiency

Future Improvements

- Additional datasets to integrate:
    - Detailed weather records
    - Urban infrastructure data
    - Humidity, wind speed information
- Continuous model refinement

Q & A

Thank You for Your Attention !