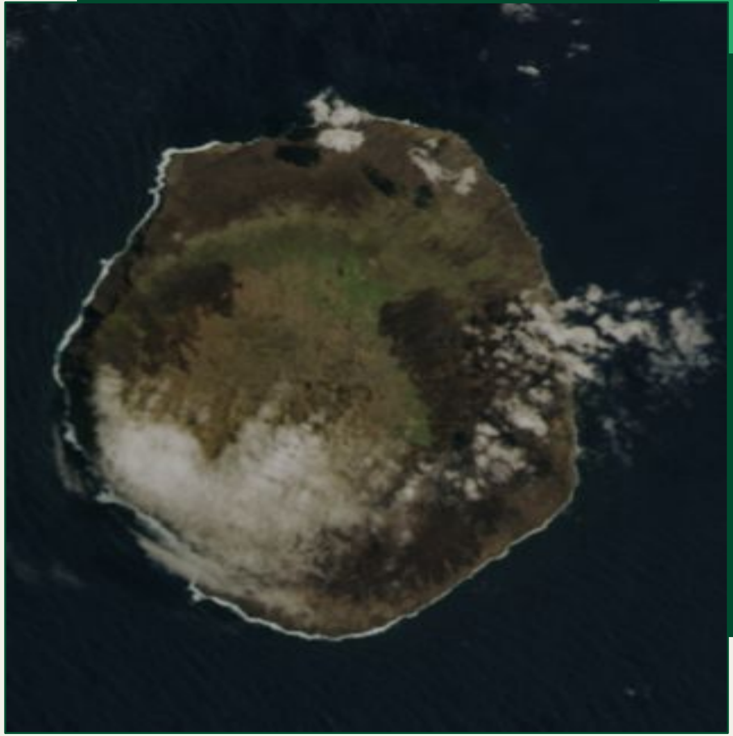
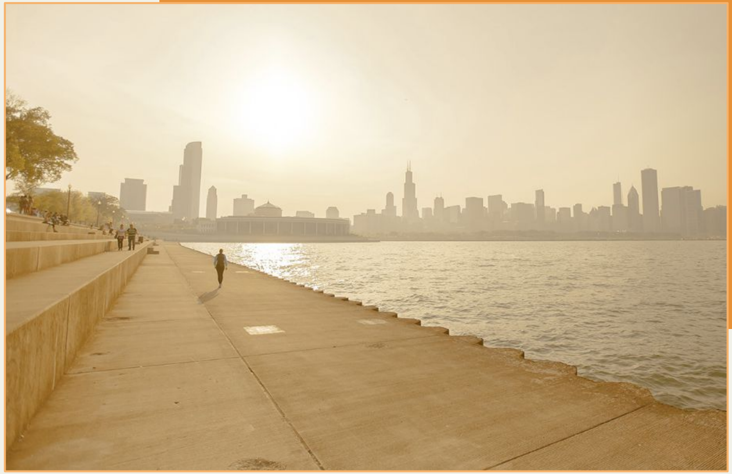
# Urban Heat Island (UHI) in NYC



Yanjun (Clivia) Zhou
Yifei (Lexie) Li
Yunqing (Kelly) Sun

# Contents

# Introduction



## BACKGROUND

- <u>Urban Heat Island (UHI) Effect:</u>
  - Elevated temperatures in urban areas
  - Cause significant health, social, and energy-related issues
  - Vulnerable populations are disproportionately affected
- <u>Problem Statement:</u> Develop a ML Model to *predict* UHI hotspots in urban areas & *identify* key contributing factors.

# Dataset Overview

**Source**

- Near-surface air temperature data in the Bronx and Manhattan, NYC
- Collected on 24th July 2021, 3:00 pm - 4:00 pm

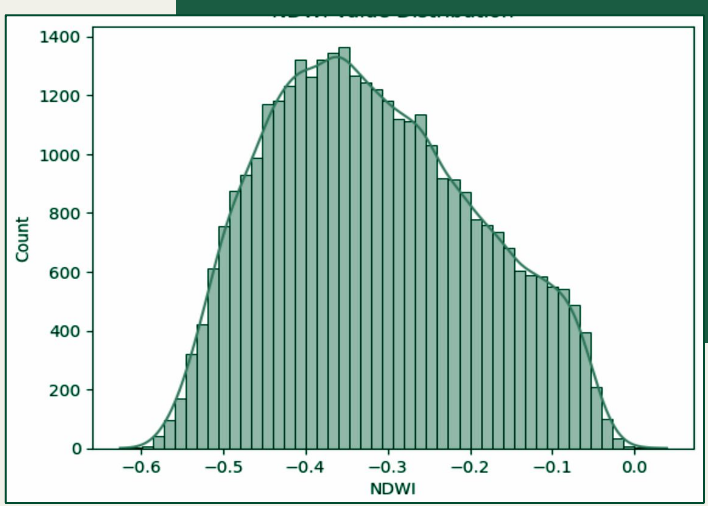
**Feature & Size**

- 11229 data points
- Latitude, Longitude, Time, UHI Index
- Satellite-derived features

**Target Variable**

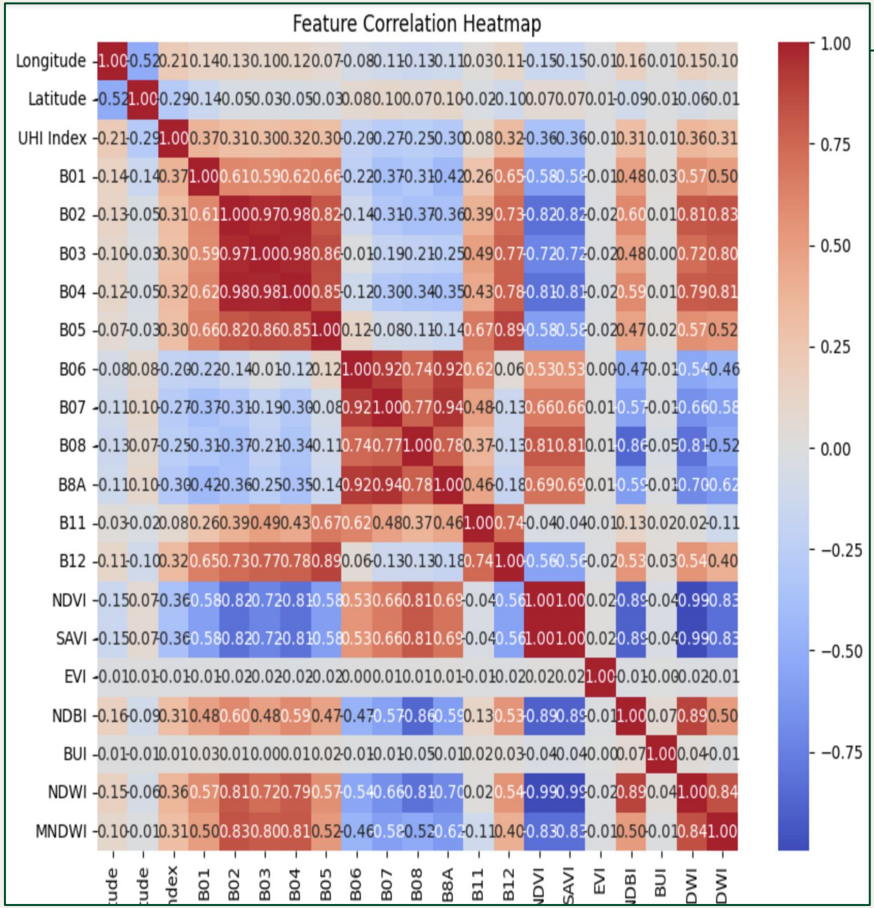
- UHI Index: Relative temperature difference compared to the city's average → Predict the values and highlight factors driving UHI intensity

# Data Cleaning



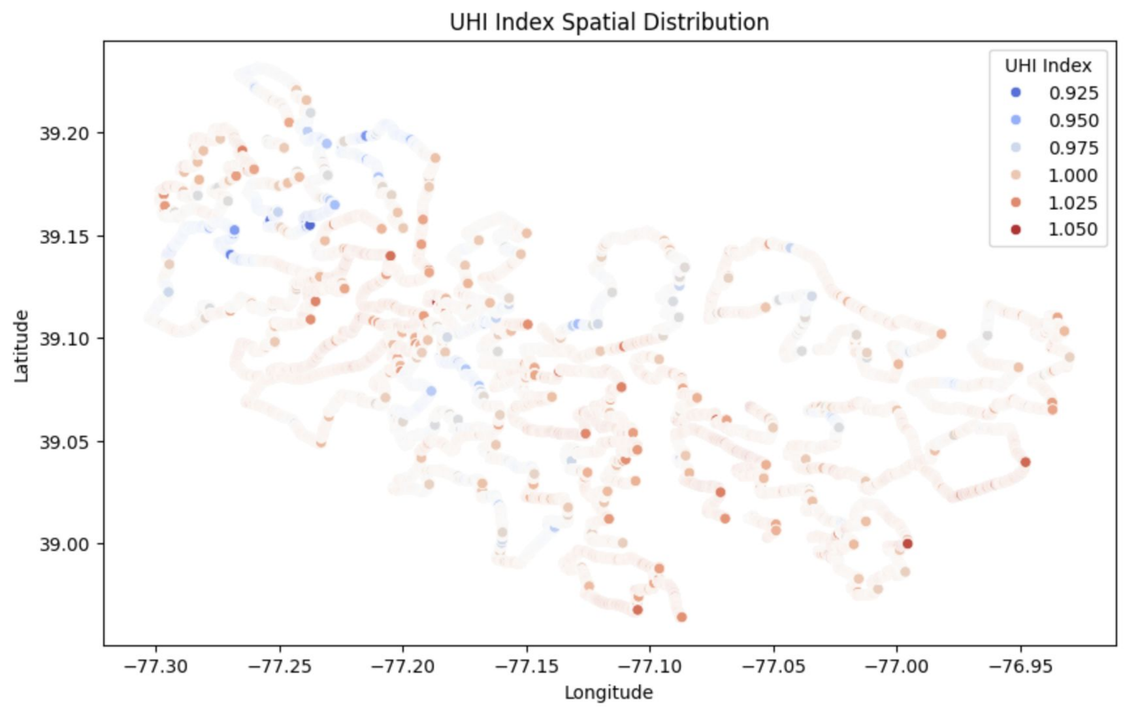

## NDWI Value Distribution

- Index for detecting water bodies in satellite imagery
- Positive: Nearby water bodies (eg. Lakes)
- What the graph reflect:
  - Cities & Vegetations
  - Water Bodies V.S. Cooling Factor




## Correlation Matrix

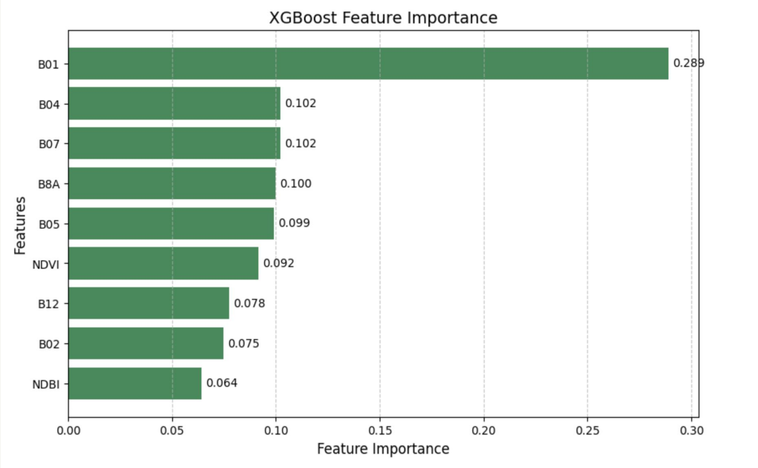
- Removal for highly correlated variables
  - NDVI & SAVI (0.99): Vegetation Index
  - NDWI & MNDWI (0.84): Normalized Difference Water Index
- Insights of correlation with the target variable:
  - B01 (0.37): Building reflections
  - B12 (0.33): Surface dryness
  - B04 (0.32): Red light band --> NDVI




## UHI Index Distribution

- ➢ Red: Dense buildings or surface materials → heat accumulation
- ➢ Orange to Grey: Urban edges or low-density building areas
- ➢ Blue: More vegetation cover, water bodies, or open green spaces → **Mitigate** the UHI effect

# Methods & Analysis




## Satellite data and surface Characteristics

- High Dimensionality
- Nonlinear Relationships
- Feature Collinearity

## Methods used

- **Random Forest** VS. **XGBoost** VS. **LightGBM**
- Train models & compare performance metrics
- Identify key features influencing predictions
- Visualize feature importance with bar charts

## Model Results

| Model | $R^2$ |
|---|---|
| Random Forest | 0.2608 |
| XGBoost | 0.3136 |
| LightGBM | 0.2766 |

Model Evaluation:
**MAE: 0.0111**
**MSE: 0.0002**
**RMSE: 0.0141**
**$R^2$: 0.3136**

| Feature | Absolute Coefficient |
|---|---|
| B01 | 0.289 |
| B04 | 0.102 |
| B07 | 0.102 |
| B8A | 0.100 |
| B05 | 0.099 |
| NDVI | 0.092 |
| B12 | 0.078 |
| B02 | 0.075 |
| NDBI | 0.064 |

- XGBoost is the ideal model based on $R^2$
- **All** features modelled and **variables selected by model**
- Model's predictions deviate by **0.0111 UHI units** from the true values on average.
- Model explains **31.36% of UHI variance**, showing moderate performance.

# Results

## Key Insights

- High B01, low NDVI, low NDBI
  - Primarily driven by urban heat-retaining surfaces and atmospheric conditions
  - NDVI is less effective in dense urban heat zones.
  - NDBI modifies buildings but doesn’t measure their UHI impact
- High B04, B8A, B12
  - Urbanization, low vegetation, and reduced surface moisture.
- Urban planning strategies
  - Increasing vegetation
  - **Modifying building materials and surface reflectivity**

# Real World Implication

## Real World Implications

- Use cooler materials: eg. Reflective or light-colored materials
- Reduce air pollution: Implement policy that limits emissions
- Plant more greenery: installing green roofs, and creating water bodies

# Future Work

## Future Work

- ➢ Expand to other cities and time periods
- ➢ Incorporate additional environmental data (e.g., humidity, wind speed) and urban mobility data
- ➢ Include long-term trend analysis: Analyze how the UHI effect and other environmental conditions change over longer periods

# Thank you!