Urban Heat Island (UHI) in NYC

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





Contents

01. Introduction

- Background
- Dataset Overview
- Data Cleaning & Visualizations

02. Analysis

- Methods & Analysis
- Results

03. Conclusion

- Recommendations
- Extensions

Introduction



BACKGROUND

- Urban Heat Island (UHI) Effect:
 - Elevated temperatures in urban areas
 - Cause significant health, social, and energy-related issues
 - Vulnerable populations are disproportionately affected
- Problem Statement: 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

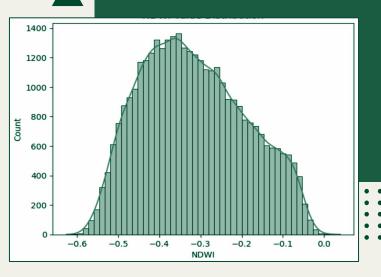
- 11229 data points
- Latitude, Longitude, Time, UHI Index
- Satellite-derived features

Target Variable

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



Data Cleaning



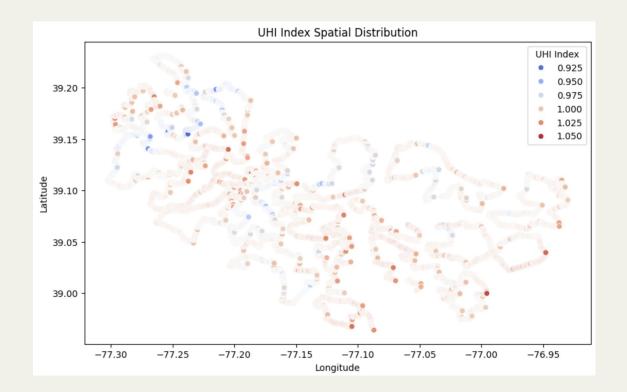
NDWI Value Distribution

- 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

Feature Correlation Heatmap 1.00 Longitude -1.000.520.210.140.130.100.120.070.080.110.130.110.030.11-0.150.150.010.160.010.150.10 Latitude -0.521.000.290.140.050.030.050.030.080.100.070.100.020.100.070.070.010.090.010.060.01 UHI Index -0.21-0.291.000.370.310.300.320.300.200.270.250.300.080.32-0.360.360.010.310.010.360.31 - 0.75 B01 -0.140.140.371.000.610.590.620.660.220.370.310.420.260.650.580.580.010.480.030.570.50 B02 -0.130.050.310.611.000.970.980.820.140.310.370.360.390.730.820.820.020.600.010.810.83 - 0.50 803 -0.100.030.300.590.971.000.980.860.010.190.210.250.490.77-0.720.720.020.480.000.720.80 B04 -0.120.050.32 0.62 0.98 0.98 1.00 0.85 0.12 0.30 0.34 0.35 0.43 0.78 0.8 10.8 10.0 20.5 9 0.0 10.7 9 0.8 1 805 -0.07-0.030.300.660.820.860.851.000.120.080.110.140.670.890.580.580.020.470.020.570.52 - 0.25 B06 -0.080.080.200.220.140.010.120.121.000.920.740.920.620.060. 0.000.470.010.540.46 807 -0.110.100.270.370.310.190.300.080.921.000.770.940.480.130.660.660.010.570.010.660.5 808 -0.130.07-0.250.310.370.210.340.110.740.771.000.780.37-0.130.810.810.010.860.050.810.52 - 0.00 B8A -0.110.100.300.420.360.250.350.140.920.940.781.000.460.180.690.690.010.590.010.700.6 B11 -0.030.020.080.260.390.490.430.670.620.480.370.461.000.740.040.040.010.130.020.020.11 - -0.25 B12 -0.11-0.100.32 0.650.730.770.780.890.060.130.130.180.741.000.560.560.020.530.030.540.40 NDVI -0.150.070.360.580.820.720.810.580.530.660.810.690.040.561.001.000.020.880.040.990.8 SAVI -0.150.070.360.580.820.720.810.580.530.660.810.690.040.561.001.000.020.890.040.990.8 - -0.50 NDBI -0.160.090.310.480.600.480.590.47-0.470.570.860.590.130.530.890.890.011.000.070.890.50 BUI -0.01-0.010.010.030.010.000.010.02-0.010.010.050.010.020.03-0.040.040.000.071.000.040.01 - -0.75 NDWI -0.150.060.360.570.810.720.790.570.540.660.810.700.020.540.990.990.020.890.041.000.8 MNDWI -0.100.010.310.500.830.800.810.520.460.580.520.620.110.400.830.820.010.500.010.841.00

Correlation Matrix

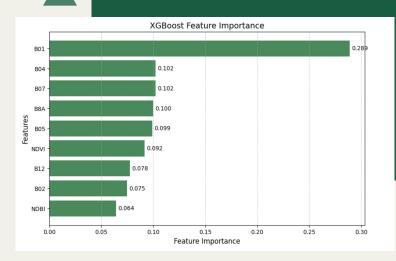
- > Removal for highly correlated variables
 - NDVI & SAVI (0.99): Vegetation Index
 - NDWI & MNDWI (0.84): Normalized
 Difference Water Index
- Insights of correlation with the target variable:
 - o B01 (0.37): Building reflections
 - B12 (0.33): Surface dryness
 - o B04 (0.32): Red light band --> NDVI



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 Random Forest		R ²	>	Model Evaluation: MAE: 0.0111 MSE: 0.0002 RMSE: 0.0141 R ² : 0.3136	>	Feature	Absolute Coefficient
		0.2608				B01	0.289
					B04	0.102	
XGBoost		0.3136		R-: 0.3130		B07	0.102
LightGBM		0.2766				B8A	0.100
						B05	0.099
>	> XGBoost is the ideal model based on R ²					NDVI	0.092
>	All features modelled and variables selected by model					B12	0.078
\triangleright	Model's predictions deviate by 0.0111 UHI units from the true values on					n B02	0.075
	average.					NDBI	0.064

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!