Introduction:

Our modeling pipeline was designed to harness a diverse set of data sources that collectively capture the multifaceted nature of the urban heat island (UHI) effect. This document outlines our approach to feature engineering, data processing, model selection, and evaluation.

Data Sources:

In addition to the provided building footprint data and Sentinel-2 imagery, we updated the Landsat and Sentinel date range to span from March 2021 through July 2021, ensuring that our satellite-derived features better align with the UHI observations. To enhance the real-world applicability of our model, we computed the average satellite imagery from days prior to the prediction target, rather than relying solely on data from the exact day of observation. This approach ensures that our model can generalize beyond retrospective analysis and be applied in a predictive setting. By leveraging historical imagery leading up to the prediction date, we simulate a scenario where UHI conditions can be estimated in advance, such as forecasting urban heat patterns three days before they occur. Furthermore, using averaged imagery smooths out short-term anomalies caused by transient atmospheric conditions, ensuring that the model learns consistent spatial patterns rather than noise from single-day observations. This not only strengthens its predictive capability but also improves its adaptability across different time frames and urban environments..

Moreover, our process fuses the building footprint data provided by the challenge with a new dataset from the New York State Open Data portal, which enriches our feature space with detailed building-related metrics, such as volumetric measurements, building elevations, and footprint efficiencies. By merging these sources, we not only leverage the high-resolution spatial details of the challenge data but also incorporate enriched construction metrics from this new dataset.

Feature Engineering:

Our feature engineering process is built on a multi-scale, multi-source approach that enhances the model's capacity to capture the urban heat island phenomenon. For the Landsat dataset, we extracted Land Surface Temperature (LST) features by calculating the mean and standard deviation of the thermal infrared band over multiple buffer radii, including values such as 50, 200, 350, 700, and 1000 meters. This granular spatial sampling allows us to capture both localized temperature variations and broader thermal trends across urban landscapes. Similarly, for Sentinel-2 data, we generated spectral features by computing band-specific statistics and several key indices, such as NDVI, GCI, various built-up indices, across the same set of buffer radii. These indices were calculated for each buffer zone (ranging from 200 to 1000 meters) to reflect changes in vegetation, urban density, and surface materials.

Feature Selection:

Our model is built around a multi-stage pipeline designed to extract maximum predictive power from the data while minimizing overfitting and noise. First, we conduct thorough feature

engineering and selection by applying a series of filters to eliminate weak and redundant predictors. We begin with correlation and collinearity checks to remove features that show minimal relationship to our target variable or overlap excessively with one another, a step that significantly reduces the computational cost of the subsequent RFECV process. Following this, we employ recursive feature elimination with cross-validation (RFECV) using an ExtraTrees regressor, which we selected over a Decision Tree due to its higher robustness and over Random Forest because it achieves similar accuracy with significantly lower computational cost. ExtraTrees benefits from feature randomness, leading to better generalization, while RFECV ensures an automated, iterative selection process that identifies the most relevant features without manual tuning. This careful, multi-layered approach ensures that our final dataset is both lean and highly informative, ultimately leading to superior model performance.

As a result, we selected a total of **119 features**, categorized as follows:

- **Sentinel-2 and Landsat-Based Features (48 features)**: Spectral bands and indices such as NDVI, NBI, NBIA, BRBA, and BAEI, as well as temperature-related metrics from the Landsat thermal infrared band, computed across multiple buffer distances (50m to 1000m).
- **Building Characteristics (42 features)**: Metrics extracted from the New York State Open Data portal, including building count, area sums, volumetric measurements, and height statistics (mean, max, min, and standard deviation) across different spatial buffers.
- **Ground Elevation Features (15 features)**: Variations in ground elevation across different buffer zones, helping to capture the influence of terrain on heat distribution.
- **Spatial Aggregation Metrics (14 features)**: Statistical summaries (mean, standard deviation, min/max) of key features over different buffer distances, ensuring a multi-scale understanding of urban density and land cover interactions.

Model Training and Evaluation:
Once we have a streamlined set of high-impact features, we split the data into training and test subsets (maximizing the training portion) and apply a StandardScaler transformation to provide a uniform scale for all features. This preprocessed data then goes into an extensive model comparison phase, where we train and evaluate multiple ensemble regressors (Random Forest, Gradient Boosting, XGBoost, LightGBM, CatBoost, Extra Trees, and more), using 10-fold cross-validation to capture the reliability and robustness of each algorithm. Instead of relying on a single “best” model, we stack two top-performing learners (ExtraTrees and XGBoost) under an ElasticNetCV meta-learner. We chose ElasticNet over other options due to its ability to balance regularization by combining L1 (Lasso) and L2 (Ridge) penalties. Using cross-validation, ElasticNetCV optimally tunes its regularization parameters, ensuring a stable and well-generalized final model.

Conclusion:
In contrast to more straightforward pipelines, our solution stands out due to its disciplined approach to feature selection and the strategic use of stacking. By rigorously pruning irrelevant or redundant variables, we prevent the model from wasting capacity on noise. And by fusing

diverse algorithms within the stacking framework, we capture both linear and complex non-linear relationships, while the meta-learner merges these insights into final predictions. As a result, our model exhibits strong generalization and resilience to overfitting, setting it apart from simpler, single-model solutions that may be prone to either underfitting or chasing spurious correlations.