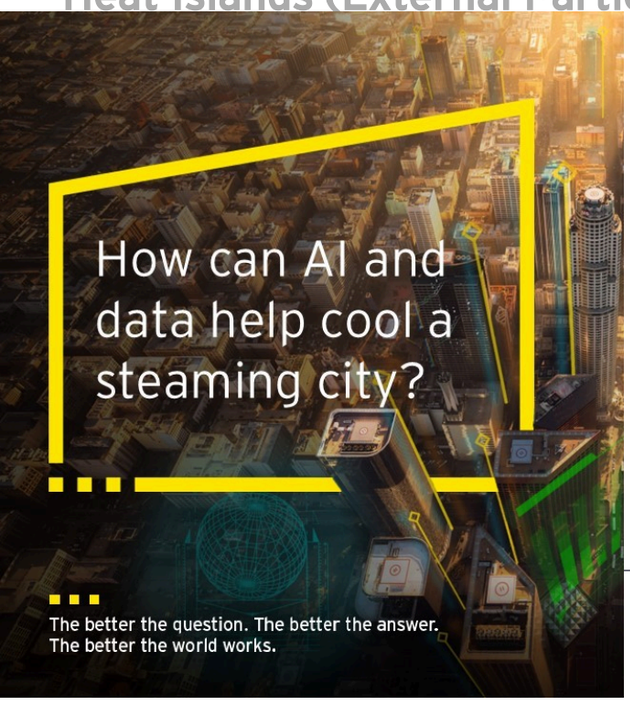
Back

The 2025 EY Open Science AI and Data Challenge: Cooling Urban Heat Islands (External Participants)



The EY Open Science AI & Data Challenge calls for innovators to address the Urban Heat Island effect using AI. Develop ML models to predict city temperatures and aid urban design for cooler, sustainable environments. Contribute to global efforts against climate change and enhance urban resilience. Join us in shaping a livable future for city dwellers.

CHALLENGE STARTS | CHALLENGE ENDS | ANNOUNCEMENT
Jan 20, 2025 | Mar 20, 2025 | May 15, 2025

Enrolled

Phase Details	Data Description	Forum	Marketplace	Submission	Ranking

Data Description

Target Dataset:

Near-surface air temperature data in an index format was collected on 24 July 2021 across the Bronx and Manhattan regions of New York City in the United States. The data was collected in the afternoon between 3:00 pm and 4:00 pm. This dataset includes time stamps, traverse points (latitude and longitude) and the corresponding Urban Heat Island (UHI) Index values for 11229 data points. These UHI Index values are the target parameters for your model.

Please find the dataset here.

Note: Participants are strictly prohibited from using Longitude and Latitude values as features in building their machine learning models. Submissions that employ longitude and latitude values as model features will be disqualified. These values should only be utilized for understanding the attributes and characteristics of the locations.

Incorporating latitude and longitude data in their raw forms or through any form of manipulation—including multiplication, embedding, or conversion to polar coordinates—as predictive features in your model is strictly prohibited, as it can compromise the adaptability of your model across diverse scenarios. This prohibition extends to calculating the distance from a reference point and using it as a feature, which is essentially a transformation of the original geographical coordinates into a new feature form. Submissions that include these types of features will be considered non-compliant and will be disqualified.

Feature Datasets:

- European Sentinel-2 optical satellite data
- NASA Landsat optical satellite data

These datasets can be extracted from Microsoft Planetary Computer Portal's data catalog. Please see the sample notebooks for more details.

Additional Datasets:

Participants can also explore the following datasets in their model development journey:

- Building footprints of the Bronx and Manhattan regions
- Detailed local weather dataset of the Bronx and Manhattan regions on 24 July 2021

Additionally, participants are allowed to use additional datasets for their models, provided those datasets are open and available to all public users and the source of such datasets are referenced in the model.

Validation Dataset:

After building the machine learning model, you need to predict the UHI index values on the locations identified in the validation dataset. Predictions on the validation dataset need to be saved in a CSV file and uploaded to the challenge platform to get a score on the ranking board.

Supporting Material:

Participants can refer to the following material before starting model development:

- Participants' guidance document, which provides a detailed overview of urban heat island concepts, relevant datasets, and suggestions for model development
- Jupyter notebook where a sample model has been built by using challenge training data
- Sample notebook to download a GeoTIFF image from the Sentinel-2 satellite dataset
- How to Get Started video
- Tips for Success video
- Orientation session video

This ZIP file contains all of the required content mentioned above. You will find datasets, sample notebooks and documentation to support the data challenge.

Terms of Use and Licensing requirements for the datasets:

Training Data:

- Description: Ground temperature data over New York City on July 24, 2021 (CSV format)
- Contributors: Climate, Adaptation, Planning, Analytics (CAPA) Strategies
- Data Host: Center for Open Science - https://www.cos.io
- Terms of Use: https://github.com/CenterForOpenScience/cos.io/blob/master/TERMS_OF_USE.md

Satellite Data (Sentinel-2 Sample Output)

- Description: Copernicus Sentinel-2 sample data from 2021 obtained from the Microsoft Planetary Computer (TIFF format)
- Contributors: European Space Agency (ESA), Microsoft
- Data Host: Microsoft Planetary Computer - https://planetarycomputer.microsoft.com/dataset/sentinel-2-l2a
- Terms of Use: https://sentinel.esa.int/documents/247904/690755/Sentinel_Data_Legal_Notice
- License: https://creativecommons.org/licenses/by-sa/3.0/igo/

Building Footprint Data

- Description: Building footprint polygons over the data challenge region of interest (KML format)
- Contributors: Open Data Team at the NYC Office of Technology and Innovation (OTI) - New York City Open Data Project
- Data Host: https://data.cityofnewyork.us/Housing-Development/Building-Footprints/nqwf-w8eh
- Terms of Use: https://www.nyc.gov/html/data/terms.html and https://www.nyc.gov/home/terms-of-use.page
- License: https://github.com/CityOfNewYork/nyc-geo-metadata#Apache-2.0-1-ov-file

Weather Data

- Description: Detailed weather data collected every 5 minutes at two locations (Bronx and Manhattan). Includes surface air temperature (2-meters), relative humidity, average wind speed, wind direction, and solar flux.
- Contributors: Contributors: New York State Mesonet
- Data Host: https://nysmesonet.org/
- Terms of Use: https://nysmesonet.org/about/data
- License: https://nysmesonet.org/documents/NYS_Mesonet_Data_Access_Policy.pdf