**Problem Statement**

**Organization Name:** UMD Department of Architecture

**Dataset Name:** D.C. Building Energy Benchmarks**A close up of a logo

Description automatically generated**

**Difficulty:**

Level 2: Participants with basic data analysis knowledge.

The problem statement is open-ended yet straightforward. The dataset has a standard structure suitable for beginners. Creative and interdisciplinary solutions are welcomed.

# Background

The Clean and Affordable Energy Act of 2008 established that all private buildings over 50,000 gross square feet within the District of Columbia, including multifamily residences, must annually measure and disclose their energy and water consumption to the Department of Energy and Environment (DOEE). Benchmarking is defined as tracking a building’s energy and water use and using a standard metric to compare the building’s performance against past performance and to its peers nationwide. These comparisons have been shown to drive energy efficiency upgrades and increase occupancy rates and property values. The District of Columbia has chosen U.S. EPA’s free, industry-standard ENERGY STAR® Portfolio Manager® tool for benchmarking and reporting. DDOE is required to publicly disclose the ENERGY STAR® Benchmarking results for each publicly or privately owned building that is subject to the benchmarking law, beginning with the 2nd year of benchmarking data for that building.

# Questions

# The main research questions are: (1) identify the most influential factors/variables that contribute to building energy use intensity (EUI) in Washington DC; (2) understand whether there is correlation between energy efficiency and carbon emission; (3) find the energy consumption patterns (increase or decrease) between 2013-2017, for each zip code and building types; (4) identify what is the factor/ratio between source energy and site energy. Specifically

1. What are influential factors influencing **Weather Normalized Site EUI (kBtu/ft˝)**? Are those influential factors the same in each building types? Zip code? Potential influential factors include postal code, year built, Primary Property Type, Tax Record Floor Area, Reported Building Gross Floor Area, Water Use, Electricity Use, Natural Gas Use, and District Water-Based Energy Use.
2. Does Weather Normalized Site EUI, Weather Normalized Source EUI has direct correlation with Total GHG Emissions (Metric Tons CO2e)? Does Weather Normalized Site EUI, Weather Normalized Source EUI has direct correlation with Total GHG Emissions (Metric Tons CO2e)?
3. The Weather Normalized Site EUI (kBtu/ft˝) trend from 2013 to 2017
   1. Which zip codes are doing better and which are worse? Potential correlation to other variables?
   2. which building types are doing better and which are worse? Potential correlation to other variables?
   3. Total GHG Emissions trend from 2013 to 2017? Potential correlation to other variables?
4. What is the factor/ratio between the source energy and site energy?

# Data Considerations

The final dataset contains more than 3,800 records and are geocoded. The student teams are encouraged to consider integrating other locational factors that can potentially answer the questions above from other data sources, such as the social-demographic and environmental information from Open DC data. Think out of the box!

The online data portal is also available for the raw dataset and the GIS shapefile: <https://opendata.dc.gov/datasets/building-energy-benchmarks/data>. Raw data glossary can be found here: <https://www.arcgis.com/sharing/rest/content/items/aba010cff7fe4d4cb369a54b56cd7544/info/metadata/metadata.xml?format=default&output=html>