

Data Explorer

OVERVIEW

The ENERGY STAR Portfolio Manager Data Explorer houses energy data for over 150,000 U.S. commercial, multifamily, and single-family properties. The Data Explorer allows users to query this data and explore energy metrics such as Energy Use Intensity (EUI), ENERGY STAR Score, and Percent Electricity across a wide range of property types and characteristics. A variety of filters and metric display options provide customizable results. The tool is available at <https://portfoliomanager.energystar.gov/dataExplorer/>.

Example use cases include:

- Researchers who want to see the distribution of ENERGY STAR scores for K-12 schools across the country and on a regional/state basis.
- Owners who want to see how their property compares with others of similar size and age.
- Policy makers who want to look at average EUI values in their state for each major property type.

Data in the Portfolio Manager Data Explorer comes from EPA’s ENERGY STAR Portfolio Manager benchmarking tool. In addition to energy use information, users of Portfolio Manager enter physical and operational property characteristics in order to generate certain metrics. Individual Portfolio Manager account holders enter data voluntarily in most cases. EPA does not verify data entered in Portfolio Manager except for properties applying for ENERGY STAR Certification (under 10,000 properties per year). For the Data Explorer, EPA conducted a rigorous data cleaning process to identify test properties, duplicate properties, and entries with inaccurate data.

The Data Explorer’s features include the ability to:

- Filter data by numerous characteristics, such as property type, size, location, age, and more.
- View results as a table or chart, including ranges, medians, and percentile values based on the filters applied.
- To ensure privacy, only aggregate results are displayed and searching for an individual property is not possible. A query must return 6 or more properties to show aggregated results.
- Download results in XLS, CSV, and JSON formats.

EPA will add a new year’s worth of data to the Data Explorer annually, with a 1-2-year lag. For example, 2022 calendar year data will be released in 2024.

This document provides an overview of the final dataset used in the Data Explorer and a technical explanation of the data cleaning process.

ABOUT THE DATA.....	2
DATA CLEANING.....	2
APPENDIX A: TABLE OF FILTERS FOR EXTREME VALUES.....	5

ABOUT THE DATA

Portfolio Manager is an interactive resource management tool that enables users to benchmark the energy use of any type of property in a secure online environment. Benchmarking in Portfolio Manager enables building owners and managers to save energy by measuring and comparing a property's energy to similar properties, past consumption, or a reference performance level. Nearly 25% of U.S. commercial and multifamily building space is already actively benchmarking in Portfolio Manager, making it the industry-leading benchmarking tool.

The data in Portfolio Manager is entered by individual account holders voluntarily or, in some cases, as required by a local or state law. Except for properties applying for ENERGY STAR Certification (under 10,000 properties per year), EPA does not verify data entered in Portfolio Manager. In addition, EPA has not determined that it is fully representative of the U.S. building stock. Since most organizations choose to use Portfolio Manager voluntarily, the data is prone to skew towards those more interested in efficiency, larger buildings, and certain geographical regions.

For the Portfolio Manager Data Explorer, EPA conducted a rigorous cleaning process (described in more detail in the next section of this document), with the final data for each year representing over 150,000 U.S. properties of all types.¹ While EPA has made every attempt to clean the dataset to remove duplicate entries, test properties, and entries with inaccurate data, EPA makes no claim of accuracy for the data included in the Portfolio Manager Data Explorer nor has EPA determined that this data is fully representative of the U.S. building stock.

DATA CLEANING

Six sets of filters (listed in the table below) were applied to the data pulled from Portfolio Manager during the data cleaning process. Each filter is intended to exclude properties which would adversely impact the quality of results from the Data Explorer tool. In addition to filtering the data, the properties in the dataset were geocoded for accurate geographical and climate zone allocation. Addresses from the dataset underwent a standardization process as well to ensure all ZIP codes are up to date and street names were standardized (e.g., "Avenue" was to be entered as "Ave") to ensure cleaning and coding measures were implemented accurately.

¹ An additional consideration is that data for each year may not consist of the same group of properties since not all properties benchmark in Portfolio Manager every year.

Summary of Filters for the Portfolio Manager Data Explorer Data Set

Properties Excluded from the Dataset	Rationale
Properties which do not include full calendar year data for a given data year	To ensure the key metrics are available for every property. Energy usage metrics such as EUI and ENERGY STAR Score are calculated on an annualized basis in Portfolio Manager and utilized as such by the Data Explorer.
Properties outside of the U.S. and territories	To limit results to U.S. properties only. While Portfolio Manager is available to properties outside of the U.S., they are outside of the scope of the Data Explorer tool.
Properties with a construction status of “test” or which include the words test, sample, fake, somewhere, anywhere or dummy in the address or name.	To exclude data likely to be invalid. Portfolio Manager users often benchmark hypothetical properties, which are not representative of a real property or its energy usage.
Duplicate properties	<p>To ensure the results include data for a given property no more than once in each year. There are several reasons a property may be entered more than once in Portfolio Manager. Examples of such instances may include when several parties (e.g., property manager and tenant) are interested in performance metrics, or when a property changes or shares ownership.</p> <p>EPA applied three duplicate filters. The first included matching variations of Property Name, Address, and GFA of each property. The second included matching Property Name, GFA, and EUI. Finally, a python dedupe library was applied to identify properties with similar property name, address, city, county, state/province, country, and postal code. Active learning and a predicted probability threshold were used to determine the likelihood of a duplicate. This process allows for the identification of duplicates even in instances where characters do not match exactly but properties are nevertheless likely to be duplicates. In instances where a duplicate was identified EPA removed all but the property with the most recent benchmarking activity from the data set.</p>

Properties Excluded from the Dataset	Rationale
<p>Properties with extreme values for energy use and property characteristics</p>	<p>To exclude data likely to be invalid. Properties with values outside of the normal bounds were considered likely inaccurate or representative of a hypothetical property.</p> <p>EPA applied filters on GFA and EUI to all properties since energy and square footage are among the required data for the Data Explorer regardless of property type. In addition, EPA applied a series of extreme value filters to various scorable property type-specific use details. Since the use details for the ENERGY STAR Score vary depending on property type, the bounds on use detail filters vary by property type. Appendix A includes a table of all EUI, GFA, and use detail filters. Note that some property types have blank entries because that use detail does not apply to them.</p>
<p>Properties using default values</p>	<p>To exclude data likely to be invalid. Portfolio Manager offers default use detail values for those who do not have full property use characteristic information but wish to estimate their ENERGY STAR score. This means that their property entries would not be based on measured data but rather estimates. The properties which used a default for a use detail that contributes to an ENERGY STAR Score were identified and removed from the Data Explorer data set.</p>

APPENDIX A: TABLE OF FILTERS FOR EXTREME VALUES

The following two tables detail the extreme value filters applied. Properties that do not meet any one of these criteria were removed from the Data Explorer data set.

Property Type	Gross Floor Area (square feet)		Energy Use Intensity (kBtu/ft ²)	
	Greater Than	Less Than	Greater Than	Less Than
Bank	1,000	5,000,000	5	2,000
Courthouse	5,000	1,000,000	5	2,000
Data Center	1,000	1,000,000	5	15,000
Dormitory	5,000	1,000,000	5	2,000
Hospital	20,000	5,000,000	5	2,000
Hotels	5,000	5,000,000	5	2,000
Medical Office	5,000	1,000,000	5	2,000
Office	5,000	5,000,000	5	2,000
Retail	1,000	500,000	5	2,000
Schools	5,000	1,000,000	5	2,000
Senior Care Community	5,000	1,000,000	5	2,000
Supermarket	5,000	500,000	5	2,000
Unrefrigerated Warehouse	5,000	5,000,000	5	2,000
Wastewater	No filter		N/A	N/A
Worship	1,000	1,000,000	5	2,000
Multifamily Housing	1,000	5,000,000	5	2,000
Refrigerated Warehouse	1,000	5,000,000	5	2,000
Financial Office	1,000	5,000,000	5	2,000
Other/Non-scorable	1,000	5,000,000	5	2,000

Property Type	Worker Density (number per 1,000 sq ft)		Computer Density (number per 1,000 sq ft)		Weekly Operating Hours	
	Greater Than	Less Than	Greater Than	Less Than	Greater Than	Less Than
Bank	0	10	0	10	30	168
Courthouse	0	10	0	10	30	168
Data Center	N/A	N/A	N/A	N/A	N/A	N/A
Dormitory	N/A	N/A	N/A	N/A	N/A	N/A
Hospital	0	10	N/A	N/A	N/A	N/A
Hotels	0	5	N/A	N/A	N/A	N/A
Medical Office	0	15	N/A	N/A	30	168
Office	0	15	0	15	30	168
Retail	0	5	0	5	30	168
Schools	N/A	N/A	0	10	N/A	N/A
Senior Care Community	0	5	0	5	N/A	N/A
Supermarket	0	5	N/A	N/A	30	168
Unrefrigerated Warehouse	0	5	N/A	N/A	30	168
Wastewater	N/A	N/A	N/A	N/A	N/A	N/A
Worship	N/A	N/A	0	5	N/A	N/A
Multifamily Housing	N/A	N/A	N/A	N/A	N/A	N/A
Refrigerated Warehouse	0	5	N/A	N/A	30	168
Financial Office	0	15	0	15	30	168