MANAGING BENCHMARKING DATA QUALITY

Action Team 2: Data Quality and Analysis

Zachary Hart

Institute for Market Transformation | May 2018

IMT
INSTITUTE
FOR MARKET
TRANSFORMATION

USDN | urban sustainability directors network

CONTENTS

INTRODUCTION

This guide was written to help energy and water benchmarking policy implementers understand current best practices in managing benchmarking data quality. It draws from works published by jurisdictions that are currently implementing benchmarking ordinances, academic publications, and interviews and discussions with members of the Urban Sustainability Directors Network (USDN) Benchmarking and Energy Data Collective Action Group.

Sections 1–2 of this guide help benchmarking policy implementers understand strategies they can use to improve the quality of the benchmarking data they collect and to prepare that data for publication and analysis.

DEFINITIONS

- **Analysis dataset:** The subset of data from benchmarking submissions that a jurisdiction uses for data analysis and for producing its annual benchmarking report.
- **Annual benchmarking report:** A report produced by a city summarizing and analyzing the data received in benchmarking submissions on a yearly basis.
- **Benchmarking submission:** The package of energy usage data, energy performance data, and building characteristics data submitted to the jurisdiction to comply with a benchmarking and transparency ordinance.
- **Benchmarking submitter:** The owner of a covered building or their designee, who is responsible for providing a benchmarking submission.
- **Data field:** A data category displayed in benchmarking submissions.
- **Data value:** The numerical or text information stored in a data field.
- **Default data:** Values (typically averages) derived from a sample population that are used to create an ENERGY STAR score, which can be applied until the benchmarking submitter has building-specific data.[1]
- **Data cleansing:** A systematic process of reviewing benchmarking submissions to identify and address omissions and suspected errors in the data.
- **ENERGY STAR Portfolio Manager:** A free online energy management tool provided by the United States Environmental Protection Agency (EPA) that allows users to track and assess energy and water consumption in their buildings.
- **Suspicious data field:** A field containing data values outside of an expected range as defined by the jurisdiction.
- **Transparency dataset:** The subset of data from benchmarking submissions that a jurisdiction releases to the public.

WHY DATA QUALITY MATTERS

The data produced by benchmarking and transparency ordinances is meant to help real estate market actors factor resource efficiency into the transactional and management decisions they make about ordinance-covered properties. Jurisdictions are already using benchmarking data to design additional policies and programs to improve energy efficiency and inform long-term climate planning. Benchmarking data can also be used by utilities to improve the marketing and design of their energy efficiency programs and by researchers to study the impact of energy efficiency policies and programs on building energy consumption.

All of these uses for benchmarking data rely on the assumption that the data is accurate and reliable. If benchmarking datasets are inaccurate to a significant degree, then they could lead real estate stakeholders to dismiss benchmarking data as useless, misinform the analyses of cities, utilities, and researchers alike, and erode trust in the implementing department. It is critical for each city managing a benchmarking and transparency ordinance to carefully consider the accuracy and reliability of the building performance data it generates and publishes.

SECTION ONE: IMPROVING DATA QUALITY THROUGH DATA CLEANSING

Errors and omissions in benchmarking submissions are inevitable, because benchmarking data is self-reported, often by non-energy experts. Every jurisdiction implementing a benchmarking and transparency ordinance should develop a data cleansing strategy for identifying and then correcting or removing poor-quality data.

Designing a data cleansing strategy begins with determining the criteria by which a jurisdiction will judge the quality of each benchmarking data submission. It also entails devising an efficient system for checking the data for possible errors, working with submitters to confirm or correct them and, finally, determining whether the submission warrants inclusion in the transparency or analysis datasets.

The remainder of this section recommends a system for categorizing data quality errors, suggests data quality checks a jurisdiction should run to flag potential errors, and proposes appropriate responses for each category of error.

CATEGORIZING SUSPECTED DATA QUALITY ERRORS

In benchmarking datasets there are different degrees of data quality problems. This guide proposes a system of categorizing suspected data errors based on three classes of errors.

- Class I: Compliance Errors. These are the most egregious errors in which key data fields are left blank or contain obvious mistakes. Examples of compliance errors are using the wrong building ID or not reporting 12 months of energy consumption data.
- Class II: Illegitimate Outliers for Energy Use Intensity (EUI). Benchmarking submissions that contain values outside the expected range (explained below) for important fields such as the EUI fields, Gross Floor Area, or ENERGY STAR score are called outliers. Outlier values may be correct or they may be the result of error.
- Class III: Illegitimate Outliers for the ENERGY STAR Score. These are incorrect or default data values in property use details data fields such as "Weekly Operating Hours" or "Number of Workers on Main Shift" that affect a building's ENERGY STAR score.

In creating a data cleansing strategy, an implementing jurisdiction must first define the conditions for which a benchmarking submission should be flagged for a suspected error. Once these definitions are in place, the jurisdiction needs to design a data checking system to scan its benchmarking submissions and flag and address possible data errors.

PREPARING THE TRANSPARENCY AND ANALYSIS DATASETS

Jurisdictions prepare benchmarking data for two broad purposes: transparency and analysis. The transparency dataset is what a city makes publicly available on a spreadsheet or visualization map. The real estate market uses this dataset to evaluate the energy performance of buildings. The overall accuracy of this dataset is important, as rampant errors could quickly compromise the credibility of the benchmarking dataset as a whole. The analysis dataset is what the

IMPROVING DATA QUALITY BEFORE BENCHMARKING DATA COLLECTION BEGINS

Prior to the ordinance's compliance deadline, a jurisdiction can employ several strategies to increase the chances that the benchmarking submissions it receives contain high-quality data. These include requiring in the ordinance that the benchmarking submission be completed by a qualified benchmarker, requiring verification of benchmarking submissions by a qualified third party, providing extensive benchmarking trainings, benchmarking how-to guides, and a fully staffed help center to provide one-on-one compliance assistance. These methods are discussed in detail in Chapter 4 of the Institute for Market Transformation report "Putting Data to Work: How Cities are Using Building Energy Data to Drive Efficiency."

IMPLEMENTING A DATA CHECKING SYSTEM

In general, a jurisdiction should automate as much of the data checking process as possible as doing this manually can be overwhelming, particularly for jurisdictions working with hundreds or thousands of covered buildings. Scanning the data, flagging suspected errors, and corresponding with submitters can take as much as half a full-time employee's (FTE) time for several months each year.

Some jurisdictions, such as Chicago, Denver, and Seattle, have employed contractors to run their benchmarking help centers. These contractors typically employ a Customer Relationship Management (CRM) software system to manage their compliance support services. CRMs can be set up with rules to automate many of the functions of a data checking process, including correspondence with those suspected of submitting erroneous data. CRMs can be programmed to scan benchmarking submissions for missing or suspicious data fields and automatically send appropriate email responses to submitters, notifying them when a submission contains suspected errors that must be corrected or explained in order to achieve compliance. CRMs can save time and effort for the city and enable a more comprehensive set of data checks to be included in the data cleansing process. Class III errors, for example, are easier to include in a highly automated data cleansing process than a manual one, as there is virtually no additional effort for the City to include them.

Still, a contractor-operated CRM system is not necessary to implement a strong data cleansing strategy. Those cities performing their benchmarking compliance support in-house with or without a CRM system can still automate many of the data checks described in this guide. Many of the data checks described in Tables 1-7 can be automated using Microsoft Excel functions.

city uses to analyze its benchmarking data, whether for an annual benchmarking report or for internal planning purposes. Data errors can skew the results of analysis and affect any decisions based on them. While data quality is critical to both datasets, participants in the USDN Benchmarking and Energy Data Collective Action Group recommended that jurisdictions handle suspected data errors differently depending on the dataset they are preparing.

In general, the group recommends that jurisdictions make an effort to confirm or correct every Class I, II, or III error they detect. This would ensure that both the transparency and analysis datasets are as accurate as possible; however, most jurisdictions have a limited number of staff hours to devote to resolving suspected benchmarking errors, making it impractical to reach out to the submitter of every suspected error. Recognizing this limitation, for each class of data error, this report includes a subsection discussing how a jurisdiction should handle submissions suspected of having that class of error depending on the dataset they are preparing.

SECTION TWO:
IDENTIFYING SUSPECTED DATA QUALITY ERRORS

The sub-sections below give examples of specific data errors that fall under the three error classes.

CHECKING FOR CLASS I ERRORS

Benchmarking submissions with Class I errors do not meet the most basic criteria for compliance. These submissions are missing data in critical fields or contain egregious errors. The first thing to look at when checking benchmarking submissions for Class I errors is the presence of ENERGY STAR Portfolio Manager's automatically generated data quality alerts. Table 1 below summarizes the eight alerts that can be collected in the ENERGY STAR Portfolio Manager custom reports jurisdictions receive from submitters.[2]

TABLE 1: ENERGY STAR PORTFOLIO MANAGER DATA QUALITY ALERTS

ALERT NAME	TYPE OF DATA FLAGGED
No meters selected for performance metrics	Energy, Water
There is not 12 full months of meter data	Energy, Water, Flow, IT Energy
Meter(s) with gaps in bills	Energy, Water, Flow, IT Energy
Meter(s) with overlaps in bills	Energy, Water, Flow, IT Energy
Data Center(s) without IT energy meters	IT Energy
There are no Property Uses	Use Details
Total Gross Floor Area is zero for the year selected	Use Details
One or more bills cover more than 65 days	Energy

Table 2 below contains additional recommended data checks that a jurisdiction should run to find Class I errors.[3]

TABLE 2: DATA CHECKS FOR CLASS I ERRORS

DATA CHECK	EXPLANATION	NOTES
Gross Floor Area (Buildings)	The field titled "Property GFA–Calculated (Buildings) (ft2)" should not contain the phrase "Not Available" and should contain a number greater than 0.	
No Energy Use Intensity	The field titled "Site EUI (kBtu/ft2)" should not contain the phrase "Not Available" and should contain a number greater than 0.	
No Water Use	The field titled "Water Use (All Water Sources) (kgal)" should not contain the phrase "Not Available" and should contain a number greater than 0.	Only for jurisdictions that require reporting of water data.
Incorrect Building ID or Property ID	Building or Property ID should be complete and correct. ID should have correct number of digits and correct syntax.	
Duplicate Record	There should only be one benchmarking submission per building ID.	Duplicate reports should be deleted, but this should not result in non-compliance.
No electricity use	The field titled "Electricity Use–Grid Purchase and Generated from Onsite Renewable Systems (kWh)" should not contain the phrase "Not Available" and should contain a number greater than 0.	
Whole Building Energy Use	The "Metered Areas (Energy)" fields should be "Whole Building" or, if a campus, "Another Configuration."	
Whole Building Water Use	The "Metered Areas (Water)" fields should be "Whole Building."	Only for jurisdictions that require reporting of water data.

TABLE 2 (CONTINUED)

DATA CHECK	EXPLANATION	NOTES
ENERGY STAR score not available	The "ENERGY STAR score" field should have a 1-100 value if the "Primary Property Type–EPA Calculated" is one of the property types currently eligible to receive an ENERGY STAR Score. Eligible Property Types: • Bank Branch • Barracks • Courthouse • Data Center • Distribution Center • Financial Office • Hospital (general medical & surgical) • Hotel • K-12 School • Medical Office • Multifamily Housing • Non-refrigerated warehouse • Office • Refrigerated Warehouse • Residence Hall/Dormitory • Retail Store • Senior Care Community • Supermarket/Grocery Store • Wastewater Treatment Plant • Wholesale Club/Supercenter • Worship Facility	There are instances in which the lack of an ENERGY STAR score for an eligible building does not indicate an error. ESPM has restrictions and conditions that a building must meet to get a score. These include maximum vacancy rates, the percentage of the building devoted to non-eligible use types, and various details about the property use. A submission lacking an ENERGY STAR score is not necessarily a Class I error and deserves further investigation before calling it non-compliant.
Water Score not available	The "Water Score" field should have a1-100 value if the "Primary Property Type–EPA Calculated" is "Multifamily Housing" Eligible Property Types: • Multifamily Housing	Only for jurisdictions that require reporting of water data.
No natural gas use	"Natural Gas Use" field should have a number > 0 and should not have "Not Available"	If using this alert, always check with the submitter to be sure that the building is not all-electric. This check may not be appropriate for jurisdictions with a high proportion of all-electric buildings.
No weather-normalized source energy use intensity	The "Weather Normalized Source EUI (kBtu/ft^2)" field should be a number > 0, and should not be "Not Available"	Buildings may correctly not receive a weather-normalized EUI if they have a "delivered fuel" such as fuel oil or diesel with very infrequent deliveries, or if a meter was added during that year (e.g., onsite renewables).[4]

Handling Class I Errors

Submissions with Class I errors should generally not be included in either the transparency or analysis datasets, though most jurisdictions will find that some exceptions should apply to this rule. In general, when a jurisdiction finds any of the Class I errors described in Tables 1 and 2, it should reject the submission as noncompliant, though again, some exceptions will inevitably apply. The recommended practice is to notify the submitter of their noncompliant status and instruct them to correct the error(s) and resubmit in order to achieve compliance. The jurisdiction may be able to correct some errors, such as a typo in the Property ID or address. While a jurisdiction might not deem a submission noncompliant in such cases, the jurisdiction should still notify the submitter and instruct them to correct the error so that it does not turn up again in the next year's submission.

CHECKING FOR CLASS II ERRORS

While Class I errors are generally the result of missing or incomplete data, benchmarking submissions with Class II errors are those that contain artificially high or low values for fields related to energy or water performance, such as ENERGY STAR score and the Energy and Water Use Intensity fields, or Gross Floor Area, a critical field that distorts the energy and water performance metrics if incorrect. Outlier data values may or may not be incorrect, as in any population of buildings there will be some that are very heavy or very light energy users; however, submissions containing extreme values are more likely to be erroneous and should be systematically checked.

The simplest and most easily applied method for identifying suspected Class II errors is to define an expected range for the performance-related data fields such as Energy Use Intensity, Water Use Intensity, ENERGY STAR score and Gross Floor Area. Values that are outside the expected range are flagged for follow up to see if their results are legitimate. In defining expected ranges that make sense for their jurisdictions, benchmarking implementers should consult the expertise of local real estate and energy experts. Tables 3–7 lists the expected ranges used by a number of jurisdictions to identify outliers for key data fields.

TABLE 3: GROSS FLOOR AREA

JURISDICTION	LOWER BOUND	UPPER BOUND	NOTES
Cambridge[5]	< 1,000 square feet	> 1,000,000 square feet	
Denver[6]	10% smaller than value in tax assessor data	10% greater than value in tax assessor data	An expected range based on variance from the value found in the tax assessor database depends on the quality of the assessor's data.
Minneapolis[7]	25% smaller than value in tax assessor data	25% larger than value in tax assessor data	An expected range based on variance from the value found in the tax assessor database depends on the quality of the assessor's data.
New York City[8]	30% smaller than the value in the Department of City Planning database	30% larger than the value in the Department of City Planning database	Method developed by Urban Green Council. An expected range based on variance from the value found in the tax assessor database depends on the quality of the assessor's data.
San Francisco[9]	< 100 square feet	> 7,000,000 square feet	

TABLE 4: ENERGY USE INTENSITY

JURISDICTION	LOWER BOUND	UPPER BOUND	NOTES
Boston[10]	< 5 kBtu/sq ft	> 1,000 kBtu/sq ft	
Cambridge[11]	< 1 kBtu/sq ft	> 1,000 kBtu/sq ft	Source EUI
Chicago[12]	< 3 kBtu/sq ft, more than 3 standard deviations below the median for the property use type	More than 3 standard deviations above the median for the property use type	Site EUI
Denver[13]	< 25 kBtu/sq ft	> 375 kBtu/sq ft	Site EUI
Minneapolis[14]	< 25 kBtu/sq ft	> 400 kBtu/sq ft	
New York City[15]	<50 kBtu/sq ft	> 1,000 kBtu/sq ft	Source EUI. Method developed by Urban Green Council.
San Francisco[16]	< 1 kBtu/sq ft	> 1,000 kBtu/sq ft	

TABLE 5: ENERGY STAR SCORE

JURISDICTION	LOWER BOUND	UPPER BOUND	NOTES
Chicago[17]	< 3	> 98, unless certified within past two years	
Denver[18]	< 3	> 97	
Washington, DC	–	> 95, for multifamily	

TABLE 6: WATER USE INTENSITY

JURISDICTION	LOWER BOUND	UPPER BOUND	NOTES
Boston[19]	–	> 400 gal/sq ft	

TABLE 7: ELECTRICITY USE

JURISDICTION	LOWER BOUND	UPPER BOUND	NOTES
Washington, DC	< 15% of total Site Energy Use	–	This suggests that not all electric meters in the building have been included in the benchmarking submission.

Identifying Outliers Using Statistical Methods

New York University's Center for Urban Science and Progress (NYU CUSP), New York City, and the District of Columbia developed a statistical method of identifying outlier values in the Weather Normalized Source Energy Use Intensity (EUI) and Water Use Intensity (WUI) fields for the Office and Multi-family property use types.[20] Using it at present requires a relatively high level of technical expertise, such as skill in the Python programming language, and may not be feasible for some cities, though software could be developed to address this. However, this method can be more precise as there is often a large gap between what is a normal EUI and an arbitrarily-defined upper or lower limit.

Flagging Outliers for Year-on-Year Changes

After the first year of benchmarking, jurisdictions should look for large year-to-year changes in a building's energy or water performance or gross floor area. Large increases or decreases in total energy or water used or EUI or ENERGY STAR score warrant some attention to make sure that the changes are legitimate. The City of San Francisco removes benchmarking submissions that have increased their total energy use by 100 percent or decreased it by 80 percent, as the magnitude of these changes is suspicious.[21] The cities of Chicago, Seattle, and Washington, DC deem changes in energy use of 50 percent or

greater unreasonable and do not include them in trend analysis.[22] This is similar to guidance from the U.S. Department of Energy's Better Buildings Challenge, which removes buildings with a year-on-year change in energy use of more than 40 percent, or a year-on-year change in floor area of more than 10 percent, pending clarification from the submitter. This guide recommends following up with submitters to give them a chance to explain or correct their data before removing their submissions from the analysis dataset.

Handling Class II Errors

When a jurisdiction detects suspected Class II errors, it should contact the submitter to either confirm the suspicious data or correct it. For submissions found to contain accurate data, the high energy and water users can be referred to local utility programs, while the low users can be encouraged to apply for ENERGY STAR certification or other recognitions opportunities. Jurisdictions should note the submission's unusual data to avoid flagging it in future years. If the jurisdiction is not able to confirm or correct the outlier data, it should keep the submission in the transparency dataset but remove it from the analysis dataset. Jurisdictions may wish to mark such submissions in the transparency dataset as containing "possible error(s)" to alert readers of the possible data quality issue.

CHECKING FOR CLASS III ERRORS

For Portfolio Manager's ENERGY STAR score algorithm to work properly, benchmarking submissions must include accurate information about the property use details. Benchmarking submissions with Class III errors contain inaccurate or default data in fields such as Number of Workers on Main Shift and Number of Computers that can distort the building's ENERGY STAR score. The new water score for multifamily buildings can also be distorted by inaccurate or default space use data.

Table 8 below lists data fields that the City of Denver checks to identify possible Class III errors; however, local conditions and norms often vary, so jurisdictions that wish to use these data checks should work with local expert advisers to determine reasonable parameters for these fields.

Additionally, Washington, DC deletes ENERGY STAR scores from both the transparency and analysis datasets for any buildings with occupancy under 50 percent, as any ENERGY STAR score for such a building is unlikely to be meaningful or representative of actual operational efficiency.

Handling Class III Errors

Jurisdictions should follow-up on submissions containing suspected Class III errors if resources allow. These errors affect the ENERGY STAR score, the most used and best understood energy efficiency metric for buildings. Therefore, ensuring that the transparency and analysis datasets both contain accurate ENERGY STAR data is important for informing the market of buildings' energy performance.

If resources are not available to confirm or correct suspected Class III errors, jurisdictions should include the suspected records in the transparency dataset, as the data may in fact be correct. In such cases, the jurisdiction should consider removing these submissions from the analysis dataset altogether, or from particular analyses having to do with ENERGY STAR energy or water scores, as those are the only metrics affected by the property use details fields.

TABLE 8: DATA CHECKS FOR CLASS III ERRORS IN DENVER

DATA CHECK	CITY OF DENVER'S EXPLANATION
Office—Weekly Operating Hours	Flag if weekly operating hours are greater than 70, or default data is detected by Portfolio Manager.
Office—Number of Workers on Main Shift	Flag if number of workers on the main shift is greater than the square footage of the building divided by 100, or default data is detected by Portfolio Manager.
Office—Number of Computers	Flag if number of computers are more than 30% higher than the number of workers on the main shift, or default data is detected by Portfolio Manager.
Financial Office—Weekly Operating hours	Flag if weekly operating hours are greater than 70, or default data is detected by Portfolio Manager.
Financial Office—Number of Workers on Main Shift	Flag if number of workers on the main shift is greater than the square footage of the building divided by 100, or default data is detected by Portfolio Manager.
Financial Office—Number of Computers	Flag if number of computers are more than 30% higher than the number of workers on the main shift, or default data is detected by Portfolio Manager.
Hotel—Number of Rooms	Flag if the number of rooms is more than the building square footage divided by 200, or default data is detected by Portfolio Manager.
Medical Office—Weekly Operating Hours	Flag if weekly operating hours are greater than 90, or default data is detected by Portfolio Manager.
Medical Office—Number of Workers on Main Shift	Flag if the number of rooms is more than the building square footage divided by 200, or default data is detected by Portfolio Manager.
Multifamily Housing—Total Number of Residential Living Units	Flag if the number of units is more than the building square footage divided by 400, or default data is detected by Portfolio Manager.
Multifamily Housing—Number of Units in Low-rise Setting (1–4 stories)	Flag if default data is detected by Portfolio Manager or there are units in multiple setting types for one building. This may happen on occasion, but is a common error for multi-family properties. For example, in a high-rise building, all units should be listed as high-rise even if they are on floors 1–4.

TABLE 8 (CONTINUED)

DATA CHECK	CITY OF DENVER'S EXPLANATION
Multifamily Housing—Number of Units in Mid-rise setting (5-9 stories)	Flag if default data is detected by Portfolio Manager or there are units in multiple setting types for one building. This may happen on occasion, but is a common error for multi-family properties. For example, in a high-rise building, all units should be listed as high-rise even if they are on floors 1–4.
Multifamily Housing—Number of Units in High-rise setting (10+ stories)	Flag if default data is detected by Portfolio Manager or there are units in multiple setting types for one building. This may happen on occasion, but is a common error for multi-family properties. For example, in a high-rise building, all units should be listed as high-rise even if they are on floors 1–4.)
Multifamily Housing—Number of Bedrooms	Flag if the number of rooms is more than the building square footage divided by 500, or default data is detected by Portfolio Manager.

NOTES

1. "What Are Default Values? When Are They Available?," *U.S. Environmental Protection Agency ENERGY STAR*, accessed January 25, 2018. https://portfoliomanager.zendesk.com/hc/en-us/articles/211696597-What-are-default-values-When-are-they-available-.
2. "List of ENERGY STAR® Portfolio Manager® Alerts," *U.S. Environmental Protection Agency ENERGY STAR*, last modified September 9, 2016, https://www.energystar.gov/buildings/tools-and-resources/list_portfolio_manager_alerts.
3. Many of these checks were adapted or taken from the City of Denver's list of data quality checks, created by Katrina Managan.
4. "Why is my Weather Normalized EUI N/A?," U.S. Environmental Protection Agency, ENERGY STAR. https://portfoliomanager.zendesk.com/hc/en-us/articles/211697047-Why-is-my-Weather-Normalized-EUI-N-A-
5. City of Cambridge Community Development Department, *2015 Building Energy & Water Use Report* (May 2016), http://www.cambridgema.gov/CDD/zoninganddevelopment/sustainablebldgs/~/media/B6C0D070B96D4A28A6E90C5DBD8C11A7.ashx.
6. City of Denver Department of Public Health & Environment, Energize Denver Annual Report Appendices 2017 (Feb 2018), https://www.denvergov.org/content/dam/denver-gov/Portals/771/documents/EQ/Energize%20Denver/2017%20Energize%20Denver%20Annual%20Report%20Appendices.pdf
7. City of Minneapolis, *2015 Energy Benchmarking Report* (Feb 2017), , http://www.minneapolismn.gov/www/groups/public/@health/documents/images/wcmsp-194743.pdf
8. City of New York, New York City's Energy and Water Use 2013 Report (August 2016), http://www.nyc.gov/html/gbee/downloads/pdf/nyc_energy_water_use_2013_report_final.pdf
9. SF Environment and ULI Greenprint Center for Building Performance, *San Francisco Existing Commercial Buildings Performance Report 2010-2014*, https://sfenvironment.org/sites/default/files/fliers/files/sfe_gb_ecb_performancereport.pdf.
10. City of Boston, *Energy and Water Use in Boston's Large Buildings, 2013* (August 2015), https://www.boston.gov/sites/default/files/document-file-05-2017/berdo_rprt_webfinal_tcm3-52025.pdf.
11. City of Cambridge Community Development Department, *supra* note 4.
12. City of Chicago, *City of Chicago Energy Benchmarking Report 2016* (December 2016), https://www.cityofchicago.org/content/dam/city/progs/env/EnergyBenchmark/2016_Chicago_Energy_Benchmarking_Report.pdf.
13. City of Denver Department of Public Health & Environment, *supra* note 5.
14. City of Minneapolis, *supra* note 6.
15. City of New York, *supra* note 8.
16. SF Environment and ULI Greenprint Center for Building Performance, *supra* note 7.
17. City of Chicago, *supra* note 10.
18. City of Denver Department of Public Health & Environment, *supra* note 5.
19. City of Boston, *supra* note 8.
20. Constantine Kontokosta, Bartosz Bonczak, and Marshall Duer-Balkind, "DataIQ – A Machine Learning Approach to Anomaly Detection for Energy Performance Data Quality and Reliability," *American Council for an Energy Efficient Economy* (2016), https://aceee.org/files/proceedings/2016/data/papers/12_1139.pdf.
21. SF Environment and ULI Greenprint Center for Building Performance, *supra* note 7.
22. City of Chicago, *supra* note 10; Seattle Office of Sustainability & Environment, *Seattle Building Energy Benchmarking Analysis Report 2013 Data* (September 2015), http://www.seattle.gov/Documents/Departments/Environment/ClimateChange/EBR-2013-report.pdf.

ACKNOWLEDGEMENTS

Funding for this report was provided by the Kresge Foundation. The author would like to thank the members of the Urban Sustainability Directors Network's Benchmarking and Energy Data Collective Action Group for their assistance in developing this report. In particular, the author would like to thank the following individuals for their contributions:

- Patrick Love, City of New York
- Rich Freeh, City of Philadelphia
- Ufei Chan, City of New York
- Katrina Managan, City of Denver
- Marshall Duer-Balkind, District of Columbia
- Lisa Timmerman, City of Portland

ABOUT IMT

The Institute for Market Transformation (IMT) is a national nonprofit organization focused on increasing energy efficiency in buildings to save money, drive economic growth and job creation, reduce harmful pollution, and tackle climate change. IMT ignites greater investment in energy-efficient buildings through hands-on expert guidance, technical and market research, policy and program development and deployment, and promotion of best practices and knowledge exchange. IMT's efforts lead to important new policy outcomes, widespread changes in practice, and ultimately, lasting market shifts toward greater energy efficiency, with substantial benefits for the economy and the environment. For more information, visit imt.org and follow us on Twitter at @IMT_speaks.

ABOUT USDN

The Urban Sustainability Directors Network (USDN), a project of Global Philanthropy Partnership, is a peer-to-peer network of local government professionals from cities and counties across the United States and Canada dedicated to creating a healthier environment, economic prosperity, and increased social equity. USDN's dynamic network enables sustainability directors and staff to share best practices and accelerate the application of good ideas both between North American cities, and between North America and the rest of the world. www.usdn.org.