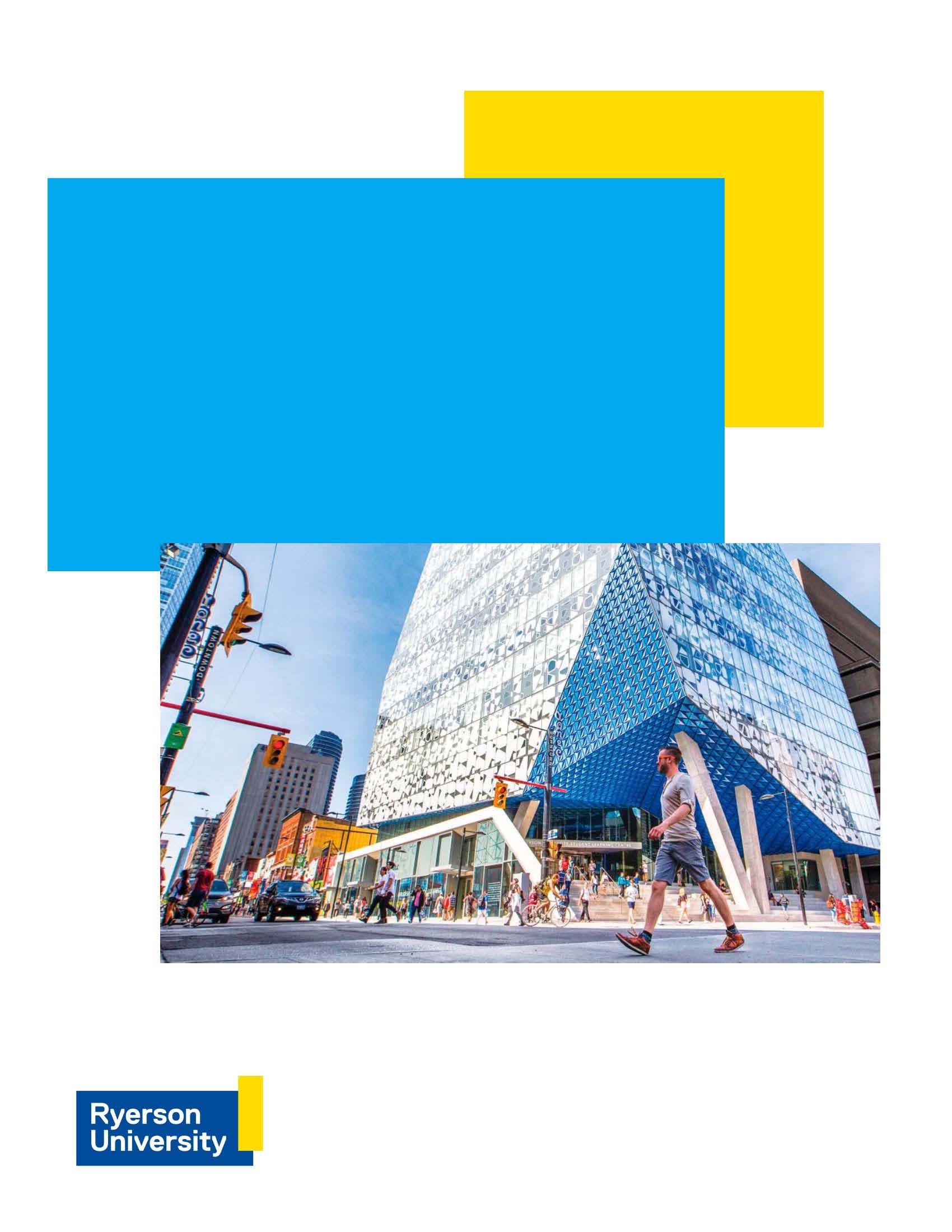
CIND820 D1H – Big Data Analytics Project

Literature Review

Shankar Ghimire

Student ID#: 501177872

Data Analytics, Big Data, and Predictive Analytics

Supervisor: Ceni Babaoglu, PhD

**Contents**

[**Abstract 3**](#_heading=h.gjdgxs)

[**Literature Review 5**](#_heading=h.30j0zll)

[Introduction 5](#_heading=h.1fob9te)

[What are the key drivers of energy efficiency across the different property types in Ontario, Canada? 6](#_heading=h.3znysh7)

[How do greenhouse gas (GHG) emissions vary by location and property type in Ontario, Canada? 7](#_heading=h.2et92p0)

[What is the relationship between water usage and energy consumption in large buildings in Ontario, Canada? 8](#_heading=h.tyjcwt)

[**References: 9**](#_heading=h.3dy6vkm)

[**A brief description of dataset attributes 11**](#_heading=h.1t3h5sf)

[**Missing Values 16**](#_heading=h.4d34og8)

[**A brief descriptive statistics of the selected datasets 18**](#_heading=h.2s8eyo1)

# Abstract

The dataset on ‘Energy and water usage of large buildings in Ontario’ provides comprehensive information on various aspects including property types, energy and water consumption, greenhouse gas emissions, and energy efficiency scores etc. The dataset contains 6,864 rows and 32 columns.

This research will explore the dataset and aims to find the following insights:

1. What are the key drivers of energy efficiency across different property types?

2. How do greenhouse gas (GHG) emissions vary by location and property type?

3. What is the relationship between water usage and energy consumption in large buildings?

Firstly, my analysis will explore the key factors of energy efficiency across the various property types. The dataset includes information such as proper type, electricity and gas consumption and energy star score for finding the key insights for energy efficiency.

Secondly, I will explore the greenhouse gas (GHG) emissions across the various property types located in the various cities and postal codes which will help find out the pattern and trends of the greenhouse gas emissions based on the building location in Ontario, Canada.

Lastly, I will analyze the relationship between the water usage and energy consumption in large buildings.

For the classification and prediction tasks on the given dataset, a systematic data analysis approach will be employed. Initially, the entire dataset will go a cleaning process, followed by preliminary analysis utilizing various exploratory data analysis tools. Then, experimental design and model constructions will be undertaken. Finally, the performance of the model will be evaluated with recommendations and conclusions.

# Literature Review

## Introduction

The dataset on ‘Energy and water usage of large buildings in Ontario’ includes a detailed data on energy (electricity and gas) consumption, water usage, and greenhouse gas (GHG) emission. Besides these measures, dataset also includes other vital information such as building certifications and energy efficiency measures. Understanding the key drivers of energy efficiency on different property types, greenhouse gas (GHG) emission by location and property types and the relationship between the water usage and energy (electricity and gas) consumption will help to develop effective strategies and take better decisions to reduce the environmental impacts of buildings in the community.

Energy efficiency is an important factor in reducing the operational cost of large buildings in Ontario, Canada and the effects of climate change and global warming in the environment. The dataset includes detailed information on energy usage of the various property types and other relevant attributes, providing a comprehensive basis for this analysis. Using dataset metrics, understanding the key drivers of energy efficiency in the various property types is very important for developing better strategies to enhance the performance of large buildings.

Greenhouse gases (GHGs) such as Carbon Dioxide (CO2), Methane (CH4), Nitrous Oxide (N2O) etc. are the gases in the atmosphere that raise the surface temperature of the Earth. Greenhouse gas emissions produced from large buildings in Ontario, Canada have direct effects in contributing to the climate change and global warming in the environment. So, its reduction has been the global priority. This research aims to highlight the areas and property types with the highest emissions. Hence, this analysis will provide a valuable insight for policy makers and stakeholders of large building in developing better strategies to reduce the greenhouse gas emissions in large buildings in Ontario, Canada.

Every large building uses a substantial amount of energy and water in its daily operations. The relationship between water usage and energy (electricity and gas) consumption in large buildings is complex and closely interconnected with each other. This research aims to explore the relationship by analyzing data on water and energy usage in various types of buildings. By identifying their correlation, this analysis aims to find insights into integrated resource management strategies that can enhance efficiency and reduce environmental effects.

The literature review explores the various existing research on related topics and provides detailed insight into the various factors influencing energy efficiency, patterns of greenhouse gas (GHG) emissions by location and property types and the relationship between the water and energy in large buildings.

## What are the key drivers of energy efficiency across the different property types in Ontario, Canada?

Energy efficiency in large buildings has been a critical concern globally. Large buildings consume a substantial portion of global energy and contribute a lot to the greenhouse gas (GHG) emissions leading to climate change and global warming issues across the earth. Understanding the key factors that influence energy efficiency in various buildings is crucial to mitigate these impacts. This research aims to identify the key drivers of energy efficiency on various property types such as ‘Retail Store’, ‘Manufacturing/Industrial Plant’, ‘Multifamily Housing’ etc. I will explore the range of factors such as proper types, energy use intensity (EUI) and third-party certification like Energy Star that can impact energy efficiency in the various property types.

The article entitled ‘Analysis of Improvement in the Energy Efficiency of Office Buildings Based on Energy Performance Certificates’ concludes that building energy performance certificates issued to office buildings demonstrated the significant impact of implementing the requirements of the Building Energy Performance Directive in the normative base of the Lithuanian construction sector in the period from 2006 to 2021 on the energy efficiency of office buildings (Bliūdžius et al., 2024).

## How do greenhouse gas (GHG) emissions vary by location and property type in Ontario, Canada?

Greenhouse gas (GHG) emissions from large buildings have a substantial impact on climate change and global warming across the globe. Greenhouse gas (GHG) emissions from large buildings in Ontario, Canada vary significantly based on location and property types. Understanding the different factors that influence greenhouse gas emissions from different property types and location play a very crucial role in developing effective mitigation strategies. This research aims to investigate the variation of the greenhouse gas emissions on various property types such as ‘Retail Store’, ‘Manufacturing/Industrial Plant’, ‘Multifamily Housing’ etc. and different geographical locations based on city or postal code. By analyzing the existing literature, I will explore the factors like energy use intensity, energy types (electricity and gas) that contribute to the greenhouse gas emissions. This research will help to identify the hotspots for the greenhouse gas emissions and prioritize the mitigation efforts. The dataset contains the information such as City, Postal Code and GHG\_Emission of various buildings located in various cities in Ontario, Canada.

The article entitled ‘Variations in direct greenhouse gas emissions across neighbourhoods: A case of Edmonton in Canada’ demonstrates that greenhouse gas (GHG) emissions vary significantly by property type and location due to factors like building configurations, population density, and transportation behavior. For example, developing neighborhoods in Edmonton exhibit higher per capita emissions than core areas, primarily due to larger single-detached dwellings and greater reliance on personal vehicles (Welegedara et al., 2021).

## What is the relationship between water usage and energy consumption in large buildings in Ontario, Canada?

Water usage and energy consumption are closely related to each other in large buildings in Ontario, Canada where water heating and pumping system consume substantial amounts of energy. Water is used in large buildings for various purposes. Understanding the relationship between these two vital resources in large buildings can help identify opportunities for efficient water usage and energy consumption. The dataset includes the various data like property types, electricity, gas and water consumption in large buildings located in the various cities in Ontario, Canada.

The article entitled "Modeling the determinants of large-scale building water use: Implications for data-driven urban sustainability policy" demonstrates the relationship between water usage and energy consumption in large buildings is significant (Kontokosta & Jain2015). Their study, which analyzed water and energy use data from over 2300 multi-family buildings in New York City, revealed that energy use intensity (EUI) has a statistically significant effect on water use intensity (WUI). Specifically, a 10% increase in EUI is associated with a 2.8% increase in WUI. This correlation suggests that buildings with higher energy consumption also tend to have higher water consumption, highlighting the potential for integrated policies that address both energy and water efficiency to achieve greater resource conservation in urban environments (Kontokosta & Jain, 2015).

# References:

Analysis of Improvement in the Energy Efficiency of Office Buildings Based on Energy Performance Certificates Bliūdžius, Raimondas; Banionis, Karolis; Monstvilas, Edmundas; Norvaišienė, Rosita; Adilova, Dinarr; et al.  Buildings; Basel Vol. 14, Iss. 9, (2024): 2791. <https://www.proquest.com/docview/3110416435/88312EBB29EC4E30PQ/9?accountid=13631>

Assessing greenhouse gas emissions and energy consumption in SMEs OECD SME and Entrepreneurship Papers; Paris, 42 pp.Oct 17, 2023. <https://www.proquest.com/docview/3051090090/D418F23D789B4B2EPQ/35?accountid=13631>

Assessing water, energy and emissions reduction from water conservation measures in buildings: a methodological approach Meireles Inês; Sousa Vitor.  Environmental Science and Pollution Research; Heidelberg Vol. 27, Iss. 5, (Feb 2020): 4612-4629.https://www.proquest.com/docview/2288312158/88312EBB29EC4E30PQ/4?accountid=13631

Building energy consumption analysis and measures: a case study from an administration building in Chengdu, China Zhang, Junye.  Energy Informatics; Heidelberg Vol. 7, Iss. 1, (Dec 2024): 78. <https://www.proquest.com/docview/3099952516/88312EBB29EC4E30PQ/13?accountid=13631>

Energy consumption and greenhouse gas emissions by buildings: A multi-scale perspective YL Li, MY Han, SY Liu, GQ Chen - Building and Environment, 2019 - Elsevier <https://www.sciencedirect.com/science/article/pii/S0360132318306917>

Energy consumption in residential buildings: Comparison between three different building styles MA Umbark, SK Alghoul, EI Dekam - Sustainable Development …, 2020 - core.ac.uk <https://core.ac.uk/download/pdf/288300515.pdf>

Modeling the determinants of large-scale building water use: Implications for data-driven urban sustainability policy CE Kontokosta, RK Jain - Sustainable Cities and Society, 2015 - Elsevier <https://www.sciencedirect.com/science/article/pii/S2210670715000633>

Variations in direct greenhouse gas emissions across neighbourhoods: A case of Edmonton in Canada NPY Welegedara, SK Agrawal, S Gajjar… - Environmental Challenges, 2021 - Elsevier <https://www.sciencedirect.com/science/article/pii/S2667010021002900>

# A brief description of dataset attributes

| **Variable Name** | **Variable Label** | **Variable Definition** | **Data Type** |
| --- | --- | --- | --- |
| EWRB\_ID | Ontario Energy and Water Reporting and Benchmarking Identifier | An EWRB ID is a unique six-digit identifier assigned by the Ontario government to each building that may be required to report data under Ontario Regulation 506/18. The EWRB ID for a building remains the same each year, even if a building is bought or sold. | Integer |
| City | City | The municipal location of a building, provided by the reporter. | Text |
| Postal\_Code | Postal Code | The first three digits of the postal code for a building (not necessarily the owning organization's mailing address), provided by the reporter. | Text |
| PrimPropTypCalc | Primary Property Type - Portfolio Manager-Calculated | Primary use type is the activity that occupies the largest gross floor area in a building. It is calculated by the Portfolio Manager software according to the gross floor areas reported by the reporter for each use type in a building. | Text |
| PrimPropTypSelf | Primary Property Type - Self-selected | The primary use type of a building, selected by the reporter (not calculated by Portfolio Manager). | Text |
| Largest\_PropTyp | Largest Property Use Type | The largest property use type by area in a building. | Text |
| All\_Prop\_Types | List of All Property Use Types at a building | A list of all property use types, provided by the reporter. | Text |
| Thrd\_Party\_Cert | Third-party Certifications | A list of building certifications, provided by the reporter. | Text |
| WN\_Sit\_Elc\_Int1 | Weather-normalized Site Electricity Intensity (GJ/m2) | The annual average electricity, measured in gigajoules, used per square meter at a building, provided by the reporter. These values are weather-normalized and not climate-normalized. | Decimal |
| WN\_Sit\_Elc\_Int2 | Weather-normalized Site Electricity Intensity (kWh/ft2) | The annual average electricity, measured in kilowatt hours, used per square foot at a building, provided by the reporter. These values are weather-normalized and not climate-normalized. | Decimal |
| WN\_Sit\_Gas\_Int1 | Weather-normalized Site Natural Gas Intensity (GJ/m2) | The annual average natural gas, measured in gigajoules, used per square metre at a building, provided by the reporter. These values are weather-normalized and not climate-normalized. | Decimal |
| WN\_Sit\_Gas\_Int2 | Weather-normalized Site Natural Gas Intensity (m3/m2) | The annual average natural gas, measured in cubic metres, used per square metre at a building, provided by the reporter. These values are weather-normalized and not climate-normalized. | Decimal |
| WN\_Sit\_Gas\_Int3 | Weather-normalized Site Natural Gas Intensity (m3/ft2) | The annual average natural gas, measured in cubic metres, used per square foot at a building, provided by the reporter. These values are weather-normalized and not climate-normalized. | Decimal |
| All\_Water\_Int1 | All Water Use Intensity (m3/m2) | The annual average water, measured in cubic metres, used per square metre at a building, provided by the reporter. These values are not weather-normalized (or climate-normalized). | Decimal |
| All\_Water\_Int2 | All Water Use Intensity (m3/ft2) | The annual average water, measured in cubic metres, used per square foot at a building, providedy by the reporter. These values are not weather-normalized (or climate-normalized). | Decimal |
| Ind\_Water\_Int1 | Indoor Water Use Intensity (m3/m2) | The annual average water, measured in cubic metres, used per square metre inside a building, provided by the reporter. These values are not weather-normalized (or climate-normalized). | Decimal |
| Ind\_Water\_Int2 | Indoor Water Use Intensity (m3/ft2) | The annual average water, measured in cubic metres, used per square foot inside a building, provided by the reporter. These values are not weather-normalized (or climate-normalized). | Decimal |
| Site\_EUI1 | Site Energy Use Intensity (GJ/m2) | The total annual average energy use of all types, measured in gigajoules, per square metre at a building, calculated by Portfolio Manager. These values are not weather-normalized (or climate-normalized). | Decimal |
| Site\_EUI2 | Site Energy Use Intensity (ekWh/ft2) | The total annual average energy use of all types, measured in equivalent kilowatt hours, per square foot at a building, calculated by Portfolio Manager. These values are not weather-normalized (or climate-normalized). | Decimal |
| Source\_EUI1 | Source Energy Use Intensity (GJ/m2) | The total annual average energy use of all types by a building plus all of the production and delivery losses, measured in gigajoules, per square metre of a building, calculated by Portfolio Manager. These values are not weather-normalized (or climate-normalized). | Decimal |
| Source\_EUI2 | Source Energy Use Intensity (ekWh/ft2) | The total annual average energy use of all types by a building plus all of the production and delivery losses, measured in equivalent kilowatt hours, per square foot of a building, calculated by Portfolio Manager. These values are not weather-normalized (or climate-normalized). | Decimal |
| WN\_Site\_EUI1 | Weather-normalized Site Energy Use Intensity (GJ/m2) | The total annual average energy use of all types, measured in gigajoules, per square metre at a building, calculated by Portfolio Manager. These values are weather-normalized but not climate-normalized. | Decimal |
| WN\_Site\_EUI2 | Weather-normalized Site Energy Use Intensity (ekWh/ft2) | The total annual average energy use of all types, measured in equivalent kilowatt hours, per square foot at a building, calculated by Portfolio Manager. These values are weather-normalized but not climate-normalized. | Decimal |
| WN\_Source\_EUI1 | Weather-normalized Source Energy Use Intensity (GJ/m2) | The total annual average energy use of all types by a building plus all of the production and delivery losses, measured in gigajoules, per square metre of a building, calculated by Portfolio Manager. These values are weather-normalized but not climate-normalized. | Decimal |
| WN\_Source\_EUI2 | Weather-normalized Source Energy Use Intensity (ekWh/ft2) | The total annual average energy use of all types by a building plus all of the production and delivery losses, measured in equivalent kilowatt hours, per square foot of a building, calculated by Portfolio Manager. These values are weather-normalized but not climate-normalized. | Decimal |
| GHG\_Emiss\_Int1 | GHG Emissions Intensity (kgCO2e/m2) | The estimated greenhouse gases emitted due to direct (site) and indirect (transmission and production) energy use of all types, measured in kilograms of carbon dioxide, per square meter of a building, calculated by Portfolio Manager. | Decimal |
| GHG\_Emiss\_Int2 | GHG Emissions Intensity (kgCO2e/ft2) | The estimated greenhouse gases emitted due to direct (site) and indirect (transmission and production) energy use of all types, measured in kilograms of carbon dioxide, per square foot of a building, calculated by Portfolio Manager. | Decimal |
| Ener\_Star\_Score | ENERGY STAR Score | The 1 – 100 ENERGY STAR score is a percentile score comparing a building's performance to national Canadian survey data of the same Portfolio Manager building type, calculated by Portfolio Manager. | Integer |
| Ener\_Star\_Certs | ENERGY STAR Certification - Year(s) Certified | The year(s) in which the reporter reports that a building has received ENERGY STAR certification. | Date |
| Data\_Qual\_Check | Data Quality Checker Run | Whether the reporter ran Portfolio Manager's Data Quality Checker tool on the data they reported, reported by Portfolio Manager. | Binary |
| Data\_Qual\_Date | Data Quality Checker Date Run | The date on which the Data Quality Checker tool was run, if it was run. | Date |

# Missing Values

| **EWRB\_ID** | 0 |
| --- | --- |
| **City** | 0 |
| **Postal\_Code** | 0 |
| **PrimPropTypCalc** | 0 |
| **PrimPropTypSelf** | 0 |
| **Largest\_PropTyp** | 0 |
| **All\_Prop\_Types** | 0 |
| **Thrd\_Party\_Cert** | 0 |
| **WN\_Sit\_Elc\_Int1** | 442 |
| **WN\_Sit\_Elc\_Int2** | 442 |
| **WN\_Sit\_Gas\_Int1** | 821 |
| **WN\_Sit\_Gas\_Int2** | 821 |
| **WN\_Sit\_Gas\_Int3** | 821 |
| **All\_Water\_Int1** | 255 |
| **All\_Water\_Int2** | 255 |
| **Ind\_Water\_Int1** | 4510 |
| **Ind\_Water\_Int2** | 4510 |
| **Site\_EUI1** | 209 |
| **Site\_EUI2** | 209 |
| **Source\_EUI1** | 209 |
| **Source\_EUI2** | 209 |
| **WN\_Site\_EUI1** | 650 |
| **WN\_Site\_EUI2** | 650 |
| **WN\_Source\_EUI1** | 650 |
| **WN\_Source\_EUI2** | 650 |
| **GHG\_Emiss\_Int1** | 0 |
| **GHG\_Emiss\_Int2** | 0 |
| **Ener\_Star\_Score** | 0 |
| **Ener\_Star\_Certs** | 0 |
| **Data\_Qual\_Check** | 0 |
| **Data\_Qual\_Date** | 0 |
| **Calculated with new source factors (Y/N)** | 0 |

# A brief descriptive statistics of the selected datasets

EWRB\_ID WN\_Sit\_Elc\_Int1 WN\_Sit\_Elc\_Int2 WN\_Sit\_Gas\_Int1 \

count 6864.000000 6422.000000 6422.000000 6043.000000

mean 115905.140297 0.432220 11.154049 0.721659

std 12789.243106 0.595779 15.374901 2.749726

min 100036.000000 0.000000 0.000000 0.000000

25% 106215.500000 0.193750 5.000000 0.340613

50% 111981.500000 0.294501 7.600000 0.454151

75% 120758.500000 0.461126 11.900000 0.681226

max 146495.000000 18.607788 480.200000 97.756377

WN\_Sit\_Gas\_Int2 WN\_Sit\_Gas\_Int3 All\_Water\_Int1 All\_Water\_Int2 \

count 6043.000000 6043.000000 6609.000000 6609.000000

mean 19.347429 1.797435 7.384968 0.686086

std 73.719185 6.848736 218.440899 20.293822

min 0.000000 0.000000 0.000000 0.000000

25% 9.131721 0.848365 0.314968 0.029261

50% 12.175628 1.131153 0.894786 0.083128

75% 18.263443 1.696729 1.574839 0.146307

max 2620.814391 243.481609 13788.335021 1280.978158

Ind\_Water\_Int1 Ind\_Water\_Int2 ... Site\_EUI2 Source\_EUI1 \

count 2354.000000 2354.000000 ... 6655.000000 6655.000000

mean 1.414857 1.414857 ... 78.613708 3.722794

std 33.109779 33.109779 ... 3280.238945 136.625572

min 0.000000 0.000000 ... 0.000000 0.000000

25% 0.027728 0.027728 ... 15.575000 0.838164

50% 0.084415 0.084415 ... 21.190000 1.117552

75% 0.145181 0.145181 ... 27.715000 1.485678

max 1280.978158 1280.978158 ... 260996.220000 10720.754840

Source\_EUI2 WN\_Site\_EUI1 WN\_Site\_EUI2 WN\_Source\_EUI1 \

count 6655.000000 6214.000000 6214.000000 6214.000000

mean 96.071896 1.530658 39.500776 2.115864

std 3525.813928 29.482479 760.836606 39.244789

min 0.000000 0.000000 0.000000 0.000000

25% 21.630000 0.614576 15.860000 0.854342

50% 28.840000 0.831770 21.465000 1.130534

75% 38.340000 1.084324 27.982500 1.485678

max 276664.070000 1756.900339 45339.270000 2339.960780

WN\_Source\_EUI2 GHG\_Emiss\_Int1 GHG\_Emiss\_Int2 Ener\_Star\_Score

count 6214.000000 6864.000000 6864.000000 6864.000000

mean 54.602819 127.502949 11.845411 36.519376

std 1012.766650 6377.020709 592.444572 35.256636

min 0.000000 0.000000 0.000000 0.000000

25% 22.047500 17.222258 1.600000 0.000000

50% 29.175000 25.833387 2.400000 31.000000

75% 38.340000 37.673689 3.500000 69.000000

max 60385.960000 508441.952202 47235.800000 100.000000