Ontario Broader Public Sector Energy Usage and GHG Emissions

CIND 820 – DJ0

Kavin Paul - 501138932

Ryerson University

Table of Contents

Abstract

Agencies under the broader public sector in Ontario are required to report energy use and greenhouse gas emissions under Ontario Regulation 507/18 made under the electricity Act, 1998. Under this regulation municipalities, their service boards, school boards, post-secondary educational institutions, and public hospitals are all required to report. The reports are summarized to include energy consumption, water, sewage, and greenhouse gas emissions. Reporting agencies are also required on every 5th anniversary of July 1, 2019, to submit and publish plans to the public for energy consumption, forecasted results, results achieved and changes to be made to reach the forecasted results.

The objective of this project is to look at the data for 6 years and see what this regulation has accomplished in that time. It will look at all the energy consumption and emissions for all the operations under the Ontario broader public sector and see which ones have been successful in reducing them. Next visualizations will be created to show what the energy consumption and GHG emissions looks like in the broader public sector of Ontario.

This project will focus on the data that has been reported for the years 2014 to 2019 available on the Ontario Government data catalogue (https://data.ontario.ca/en/dataset/energy-use-and-greenhouse-gas-emissions-for-the-broader-public-sector). After detecting any anomaly and cleaning the data, decision tree classification model will be used to create two predictive model to see which variables available in the data have the biggest contribution to the success of reducing emissions and energy consumption respectively. The analytics and modelling were done mainly on python using pandas and other libraries with some visualization done on R. Overall, this work will determine whether there have been any significant changes and if the regulation is on track to accomplishing its goals of reducing the energy consumption and greenhouse gas emissions by agencies under the Ontario broader public sector.

Literature Review

The following contains individual summaries of Articles/Journals/Papers looking at Emissions and Energy consumption and their significance on our lives today. Detailed citation provided in the references section.

CO2 Emissions

CO2 emissions have been widely recognized as the primary driver of climate change and this work provides a summary of historic and current data and look at three primary questions. Which countries are responsible for the highest emissions today? Including historical emissions which countries have emitted the most in total? And who emits the most on a per person average? Emissions have been continually going up year after year from a global perspective, we currently emit 51 billion tons of CO2 equivalent annually. Of the 36 billion tons of CO2 emitted in 2017, 53% came from Asia, 18% from North America, 17% from Europe, and 8% from Africa, South America, and Oceania. More than 50% of emissions comes from China, USA and European union with 27%, 15%, and 10% respectively. The top 10 emitters are responsible for 75% of the global emissions. Looking at historical emissions USA is responsible for 25%, EU 22%, and China 13%. Looking at per person averages can be misleading because the picture changes entirely, and it is noted that the major oil and gas producers in the world show the highest per person average for emissions. They also found that wealth is a huge indicator of our carbon footprint and as we gain wealth, we get access to a lot of modern amenities which increases our carbon footprint. The authors conclude that the richest countries in the world should take responsibility and use their resources and educated workforces to create low cost, low carbon solutions and that historically this will spread to other parts of the world who wish to trade with them. (Ritchie, H., & Roser, M. May 11, 2020)

World Greenhouse Gas Emissions in 2005

This paper summarizes the change in emissions by sector, industry, land use etc. in the year 2005 as compared to the year 2000. 12.7% emissions increase was observed Globally between 2000 and 2005. It also shows growth by different sectors that range from close to 0% to 40%. Emissions from change in land use increased by 4.1%. Transportation, cement, iron & steel were among the most significant increases. Changes due to increased electricity use and manufacturing saw emissions increase of 44%. Overall, this data shows consistent increase in emission across the board from the year 200 to 2005 (Herzog, T, 2009)

Greenhouse Gas Emissions from Global Cities

With more than 50% of the world population residing in urban areas in 2009 this study looks at data for 10 global cities: Bangkok, Barcelona, Cape Town, Denver, Geneva, London, Los Angeles, NY City, Prague, Toronto. It summarizes emissions in these cities with respect to ground transportation fuels, GHG emissions from waste and Methane emissions from landfill waste, and other sources some which seem negligible. (Kennedy, C., Steinberger, J., Gasson, B., Hansen, Y., Hillman, T., HavránekM., Pataki, D., Phdungsilp, A., Ramaswami, A., & Mendez, G. V. , 2009)

ICT Energy Consumption – Trends and Challanges

With the growth of information and communication technology this work looks at emissions as a result of ICT energy consumption and finds its responsible for the same amount as global air travel. Cell phone use has grown by a large factor from 1991 over the next 15 years and internet servers went up by a factor of 1000. They found electricity consumption doubling in households over a 4-to-5-year period. In 2005, 3% of the global energy consumption was due to cell phones, networks, and the internet. At this rapid rate several solutions are proposed to manage this growth ecologically to reduce the energy consumption (Fettweis and Zimmermann, 2008).

Estimation of Energy Consumption in Machine Learning

This work looks at including energy efficiency as a factor in Machine Learning and deep learning. It states that researchers have been focused on producing highly accurate models without considering energy consumption as a factor. They believe in existing machine learning frameworks like Tensorflow, Caffe2, and Pytorch it is possible to look at the energy efficiency as a factor and gain insight on this as well. (García-Martín, E., Rodrigues, C. F., Riley, G., & Grahn, H. ,2019)

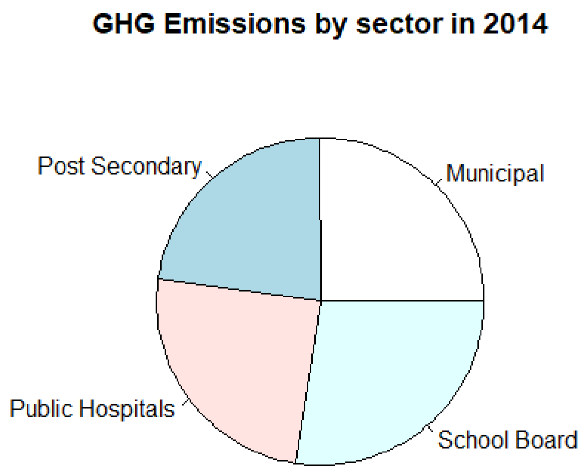
Data Description

Let’s start by looking at all the attributes in the 6 datasets and what they represent:

- Sector: The data contains 4 sectors, municipal, post-secondary, public hospital, and School Board
- Sub Sector: The data contains 8 sub sectors, Acute Hospital, Acute/Chronic Hospital, Chronic Hospital, College, University, Municipal Service Board, Municipality, School Board.
- Organization: Contains several hundred organization including, municipality names, townships, city, School Board etc.
- Operation: Column with unique values containing the names of each Operation
- Operation Type: The data contains 34 Operation Types including, Art galleries, Auditoriums, Sports Arenas, community centres, parking garages, long term care etc.
- Address: Addresses of each operation
- City: City of each operation
- Postal Code: Postal code of each operation
- Total Indoor Space, Unit: Total Indoor space of each operation listed in square feet and square metres
- Weekly Average Hours: Average hours of operation weekly
- Number of Portables: Shows the number of portables units being used by each operation

- Swimming Pool: Yes/No column showing whether there is a swimming pool
- Electricity Quantity, Unit: Contains the electricity used by each operation in kWh
- Natural Gas Quantity, Unit: Contains the natural gas used by each operation in Cubic metre, Giga Joule, and ekWh
- Fuel Oil 12 Quantity, Unit: For operations using the Oil 12 heating systems shows the quantity used in litres.
- Fuel Oil 46 Quantity, Unit: For operations using the Oil 46 heating systems shows the quantity used in litres.
- Propane Quantity, Unit: For operations using propane heating systems shows the quantity used in litres.
- Coal Quantity, Unit: For operations using coal shows quantity used. (This contains 0 for all values and has blanks for all units)
- Wood Quantity, Unit: For operations using wood for heating shows quantity used in Metric Tonnes.
- District Heating Quantity, Unit: For operation running hot water through pipes for heating shows quantity in Giga Joules and Metric Tonnes.
- District Cooling Quantity, Unit: For operations using water in pipes for cooling, shows quantity in Giga Joules and Kilo Litre of Chilled water.
- District Cooling is Renewable: Yes/No column stating if it is renewable or not.
- District heating/cooling renewable emissions factor: 2 columns for each showing the factor value for CO2 equivalent per kWh
- GHG Emissions KG: Greenhouse gas emissions in KG
- Energy Intensity: 4 columns showing energy intensity in ekWh/sqft, ekWh/ML, GJ/ML, and GJ/m2 (The columns containing GJ/m2 contains values for all Operations converted to this unit and this was used for creating the results column described in methodology).

The following is a summary of the data and a look at the year-by-year emissions and energy consumption by sector.



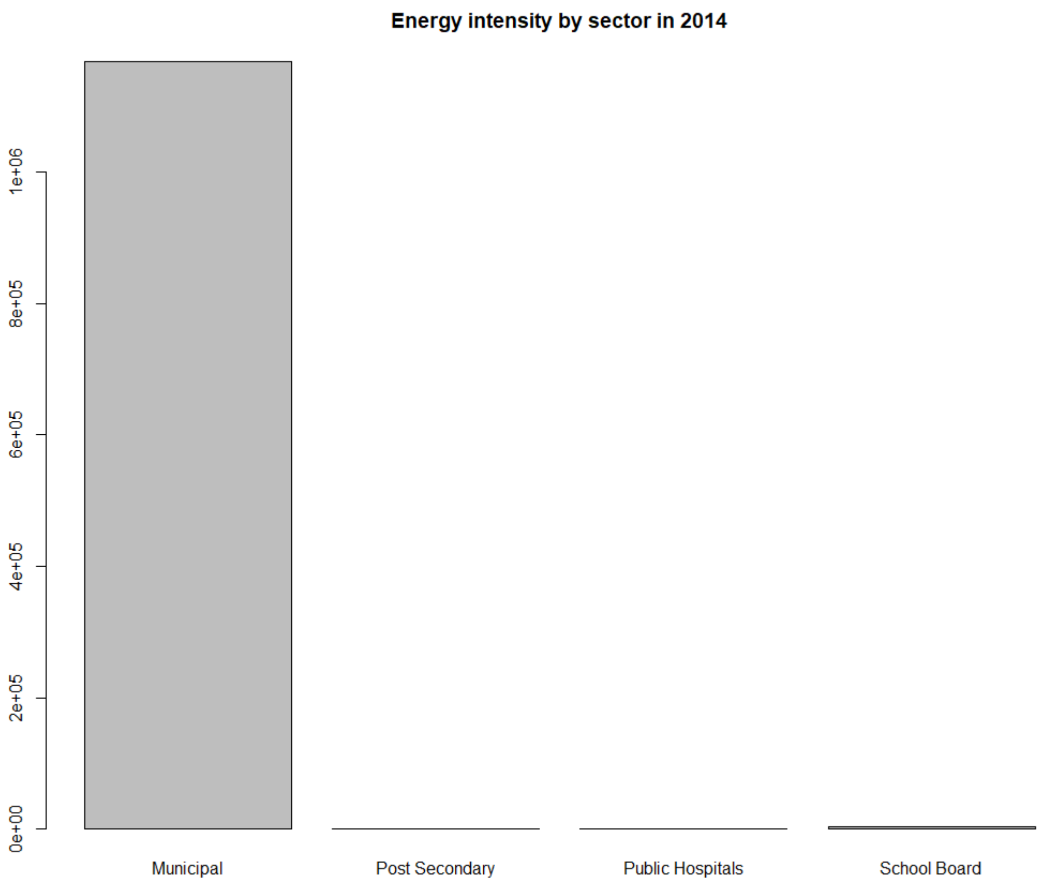
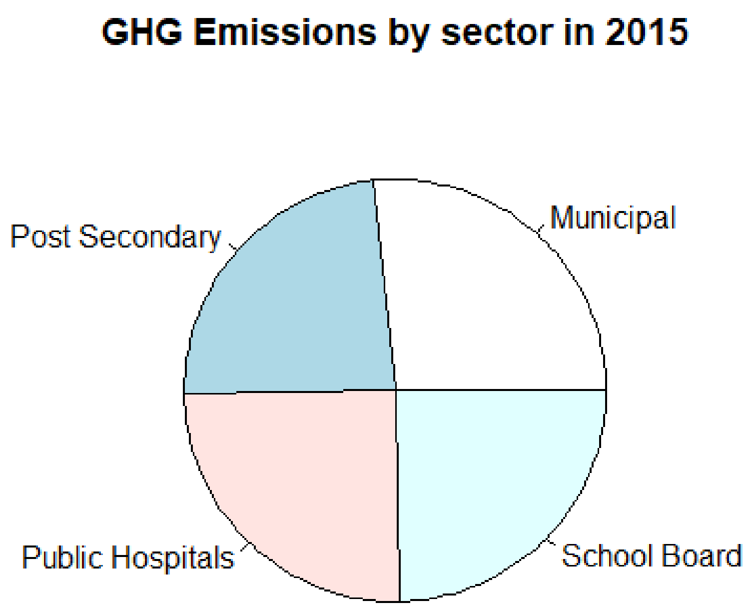


Figure 1.1 – 2014

Sector	GHG Emissions (Kg)	Energy Intensity (GJ/m^2)
Municipal	682118386.43	1167830.31
Post-Secondary Educational Institution	611856657.31	833.59
Public Hospital	674795770.64	781.78
School Board	742206125.76	3712.77

Table 1.1 – 2014



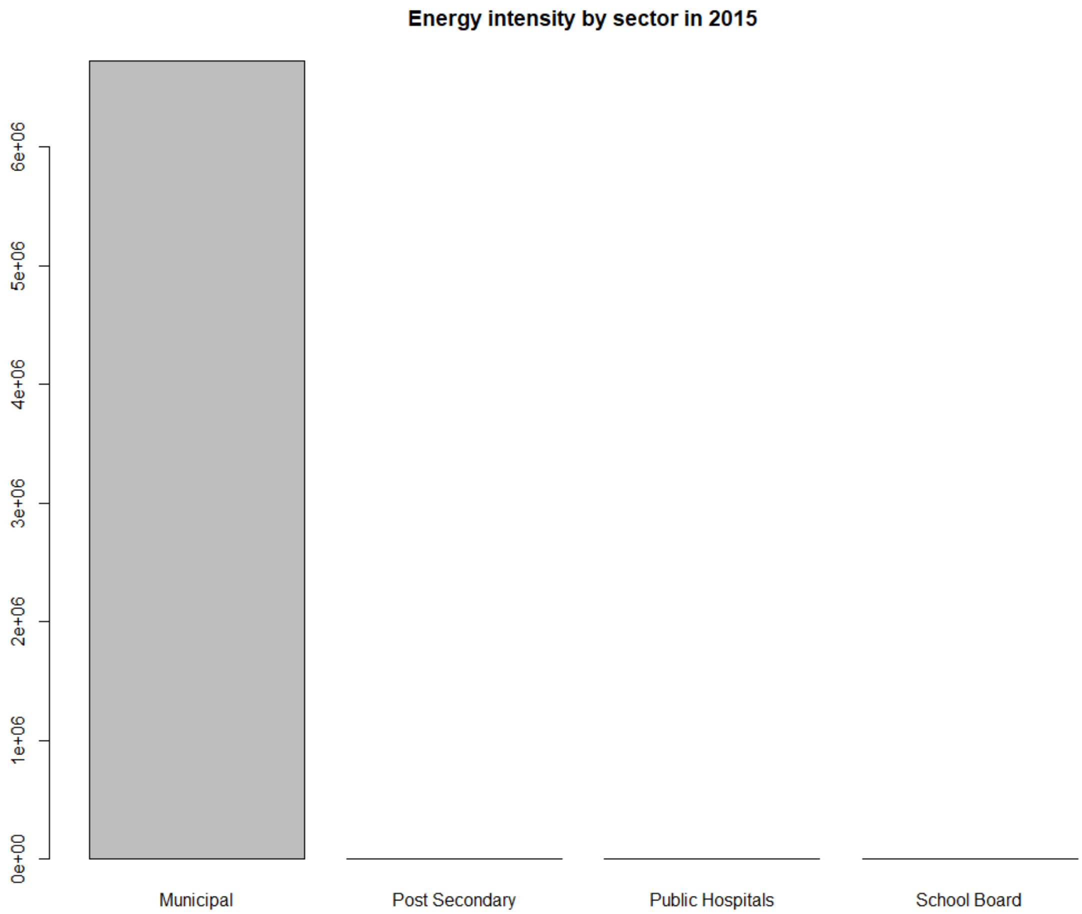
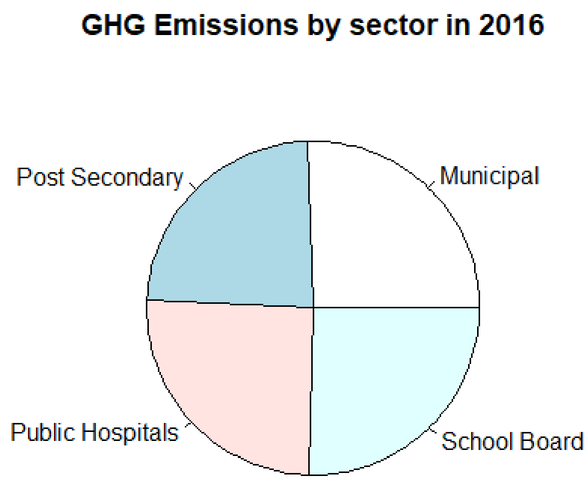


Figure 1.2 – 2015

Sector	**GHG Emissions (Kg)**	**Energy Intensity (GJ/m^2)**
Municipal	676416391.98	6720864.08
Post-Secondary Educational Institution	599681253.09	788.16
Public Hospital	636306529.09	777.45
School Board	624766467.88	3194.58

Table 1.2 – 2015



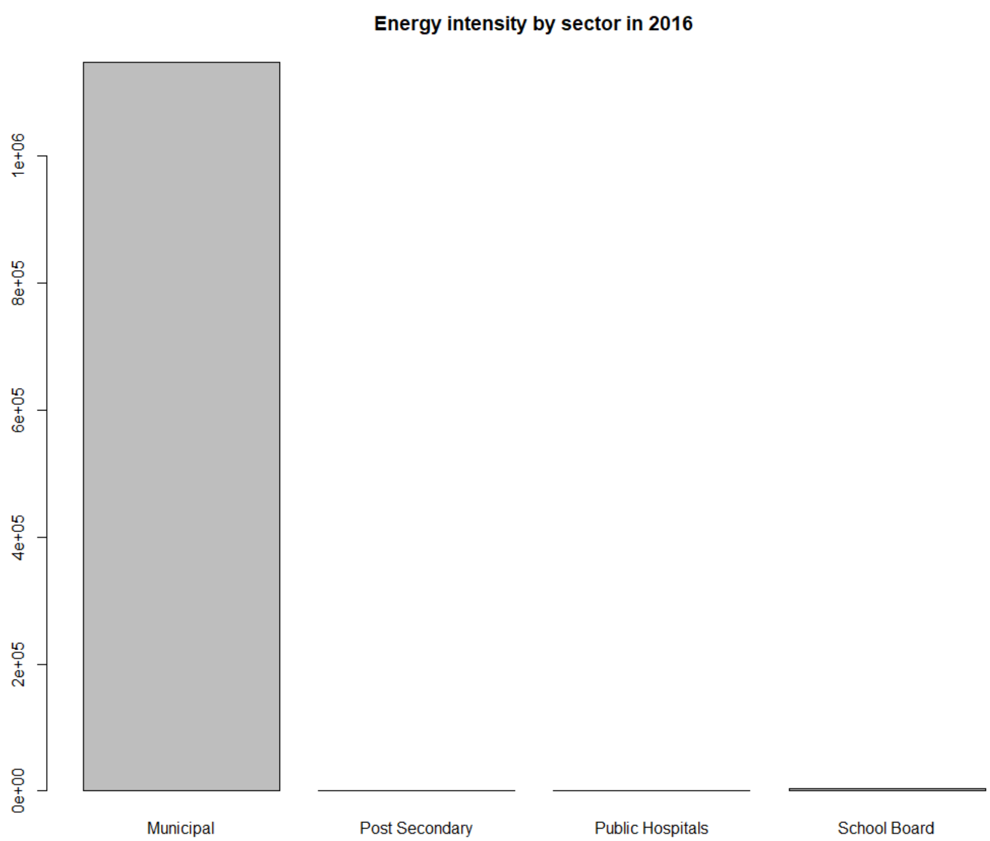
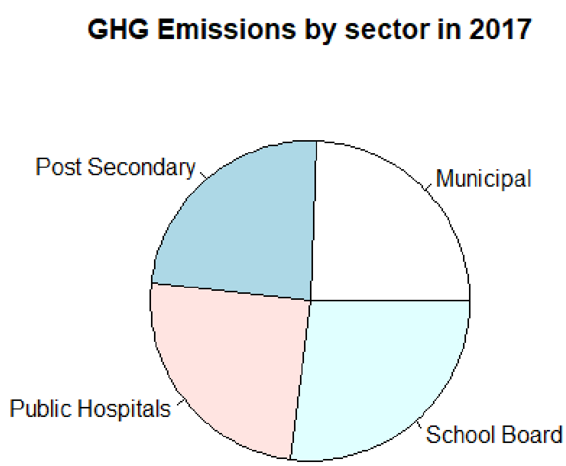


Figure 1.3 – 2016

Sector	GHG Emissions (Kg)	Energy Intensity (GJ/m^2)
Municipal	617763359.08	1147539.15
Post-Secondary Educational Institution	570626560.19	723.74
Public Hospital	613329925.18	770.76
School Board	613339172.91	3169.86

Table 1.3 – 2016



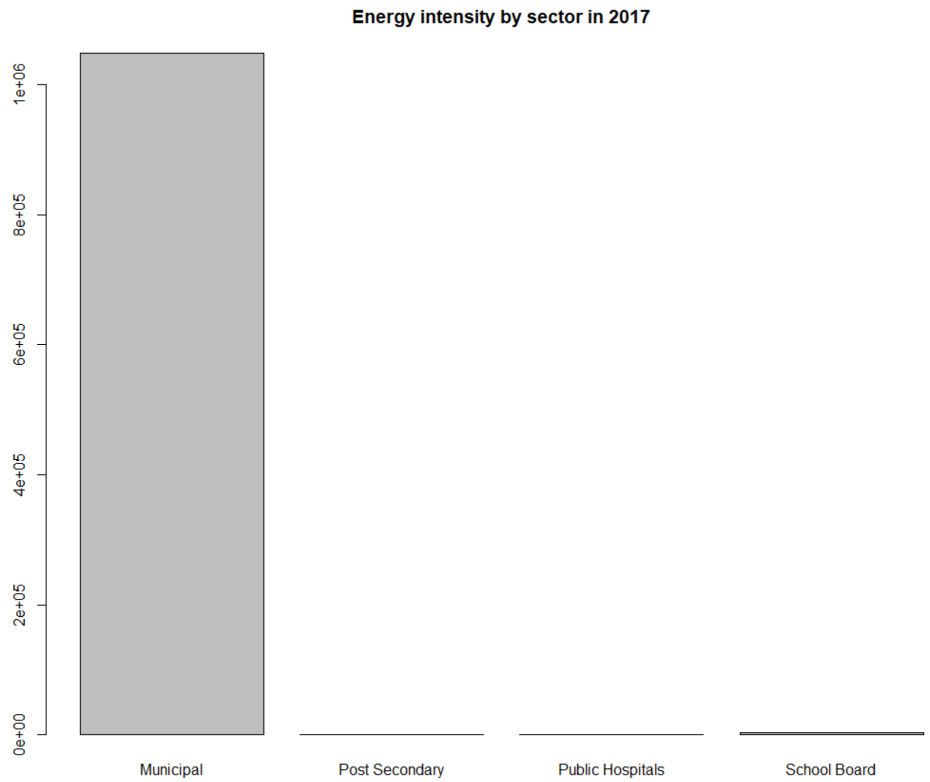
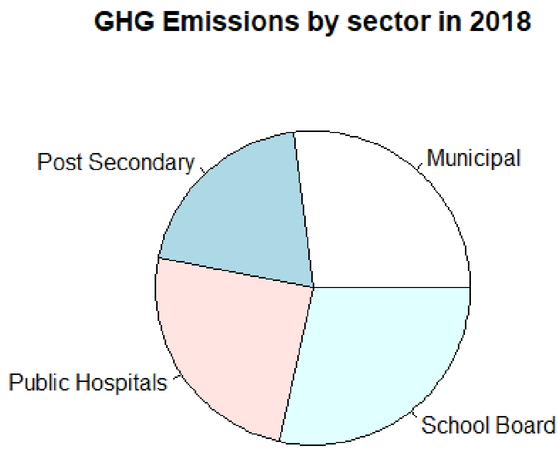


Figure 1.4 – 2017

Sector	**GHG Emissions (Kg)**	**Energy Intensity (GJ/m^2)**
Municipal	565199945.42	1047849.19
Post-Secondary Educational Institution	554407926.38	781.18
Public Hospital	575695976.75	727.22
School Board	628557569.13	3466.84

Table 1.4 – 2017



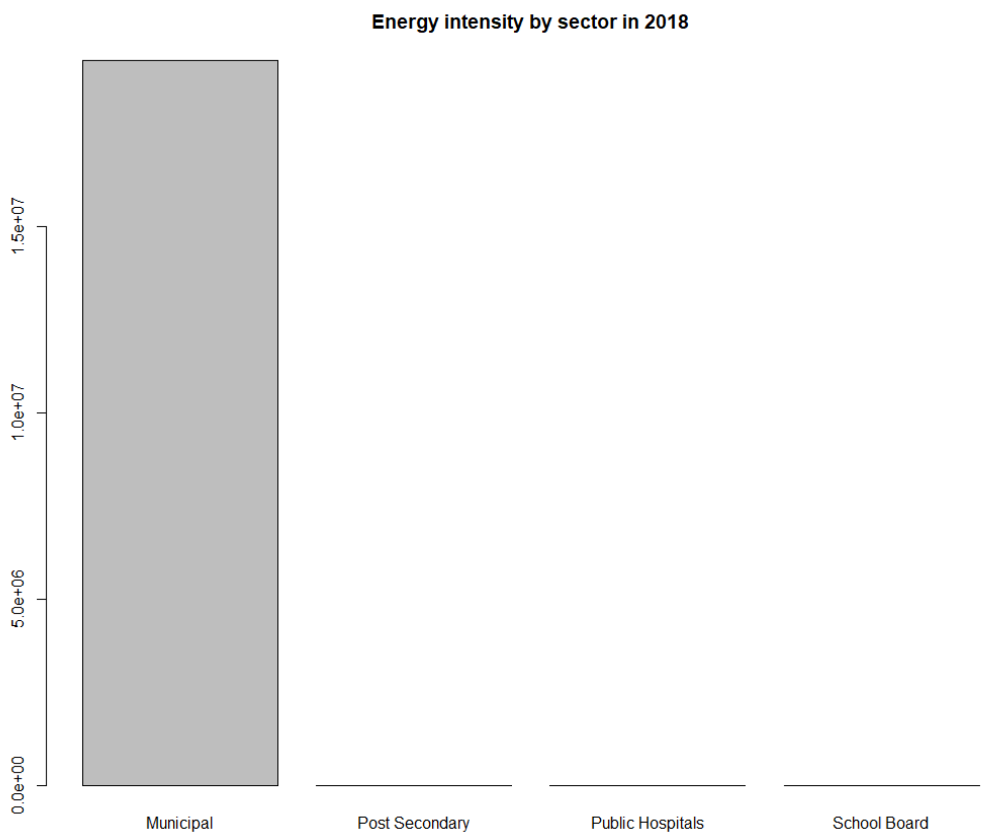
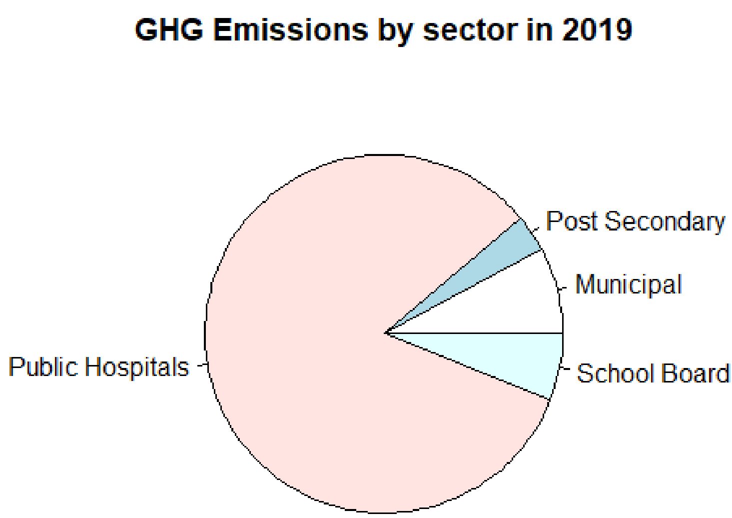


Figure 1.5 – 2018

Sector	GHG Emissions (Kg)	Energy Intensity (GJ/m^2)
Municipal	651970741.95	19436704.70
Post-Secondary Educational Institution	479295294.37	699.97
Public Hospital	593799473.49	7066.28
School Board	688019602.78	3531.42

Table 1.5 – 2018



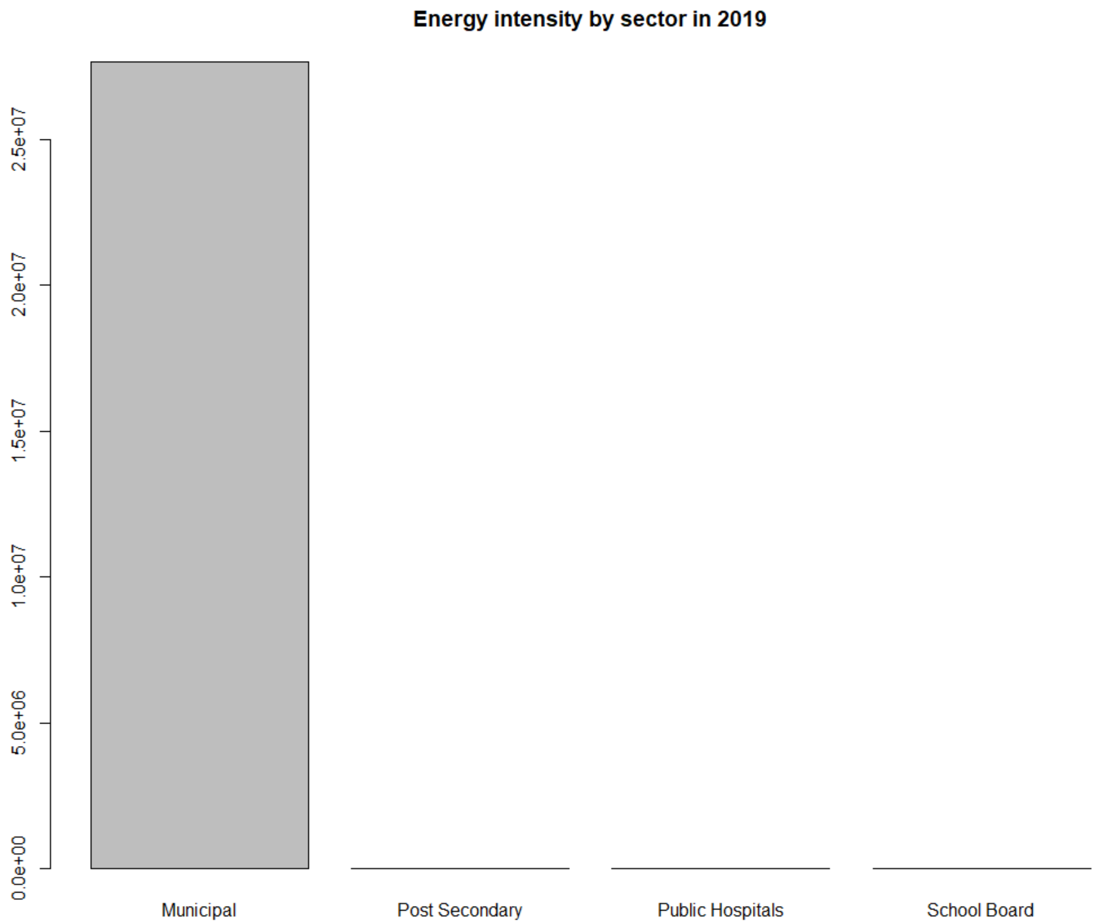


Figure 1.6 – 2019

Sector	GHG Emissions (Kg)	Energy Intensity (GJ/m^2)
Municipal	749643014.45	27656742.24
Post-Secondary Educational Institution	317813145.81	723.89
Public Hospital	7837062279.98	1802.98
School Board	566183978.03	3038.06

Table 1.6 – 2019

Figures and Tables show Emissions and Energy Consumption by year. Municipal sector typically has the highest energy consumption. In the year 2019 there is a significant increase in the GHG emissions in the Public Hospital sector. The following is a summary of both these variables over the 6 years.

2014 GHG emissions summary

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0	6126	38226	179369	116875	99665401

2015 GHG emissions summary

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0	5068	31971	163572	100313	99444636

2016 GHG emissions summary

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0	4387	28603	155740	96057	92778381

2017 GHG emissions summary

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0	3370	28988	153502	98378	87235316

2018 GHG emissions summary

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0	5146	37671	165235	111357	76117439

2019 GHG emissions summary

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0	4216	31330	628000	98290	7.217e+09

Tables 1.7

2014 Energy Intensity summary

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0	0.6	0.9	85.0	1.4	366653.4

2015 Energy Intensity summary

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0	0.6	0.8	475.0	1.4	1207413.2

2016 Energy Intensity summary

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0	0.46	0.74	74.30	1.22	176446.61

2017 Energy Intensity summary

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0	0.49	0.77	69.54	1.23	176923.73

2018 Energy Intensity summary

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0	0.5	0.8	1345.2	1.3	1904607.9

2019 Energy Intensity summary

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0	0.5	0.8	1780	1.4	1440000

Tables 1.8

Outliers

What seem to be outliers in this dataset were not removed, with exceptions. The data contains everything from a small office in a cemetery to large universities and hospitals and therefore the numbers vary from very small to very large and removing outliers for each year leaves us with majority of the data removed and very little to work with. Some values were removed, for example values above 168 for average weekly hours as it does not make sense.

Methodology

- Data Preprocessing:
    - From the 41 columns in each of the 6 datasets extract the columns to work with. These columns are Operation, Sector, Operation Type, City, Total Indoor space, Weekly Average hours, Sub Sector, Organization, Swimming Pool, Number of Portables, Oil12, Oil46, Propane, Coal, Wood, the Energy intensity (GJ/m2) and GHG Emissions(KG) (All 6 years for the last two).
    - Drop duplicates based on unique column “Operation”, convert indoor space to the same unit Sq.m, drop unit of measure columns
    - Replace all division by zero errors and nan values to zero in the numeric columns for year 2014 and create a clean dataframe, add Emissions and Energy Intensity columns by year to clean dataframe with empty values.
    - Rename columns in all dataframes to match with the names in clean dataframe, set index to unique value column “Operation” and update clean dataframe with values from each dataframe.
    - Clean commas, errors and convert numeric columns to numeric datatype, remove rows containing 0 in the 2014 or 2019 GHG and Energy Intensity Columns as we will not be able to get a result value for them.
    - For the columns Oil12, Oil46, Propane, Coal, Wood replace all non zero and non nan values with 1 to reflect which Operations are using each of these types of heating and cooling
    - For the categorical variables use pandas get_dummies to create individual columns for each category filled with 1’s and 0’s to show if they are or are not. This step will increase the number of columns drastically to 1816.
    - Create two new columns for GHG results, and Energy Intensity results and fill with zeroes. Compare values in 2014 to 2019 and change value to 1 if they have been succesful in having less emissions and Energy usage in 2019 respectively.
- Using the newly created result columns for energy intensity and GHG emissions:
    - From the dataframe drop the two result columns and the 4 columns for 2014 and 2019 GHG emissions. (The result columns were created based on these 4 columns and without dropping them all the branches of the tree will simply contain these 4 attributes.)
    - Create a 70/30 train/test split with each of the result columns seperately and create two decision trees predicitng successful reduction in energy consumption and successful reduction in greenhouse gas emissions.
- Evaluate the performance of the two models.

Results

The R file used for data description and the Python file used for data preprocessing and modelling can be found at the github repository: https://github.com/CodingKavin/Capstone.git

Out of the 9979 Operations, 7345 were able to produce less emissions in 2019 compared to 2014 and 6957 Operations were able to reduce their energy consumption. This shows the regulation has been successful in its goal of reducing emissions and energy usage in the Ontario Broader Public Sector. We will now look at each of the prediction models. A max of 10 leaf nodes were set for the modelling as going beyond that did not improve the model at all.

Emissions

First let's look at the tree and what each X label represents followed by the confusion matrix and the performance evaluations.

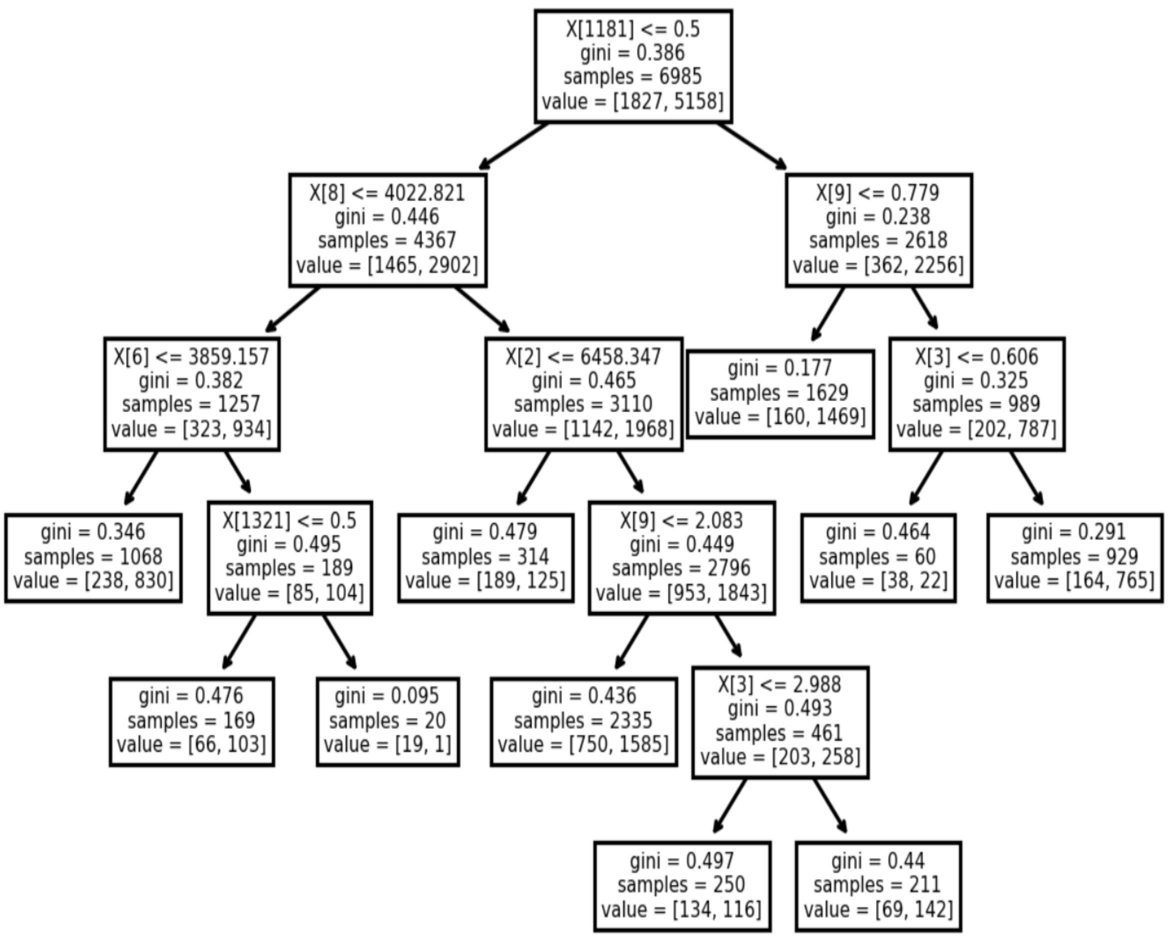


Figure 2.1 – Decision Tree for predicting emissions results

X[1181] = Sub Sector – Acute Hospital (Yes/No)

X[8] = 2017 GHG Emissions(KG)

X[9] = 2017 Energy Intensity(GJ/m2)

X[6] = 2016 GHG Emissions(KG)

X[2] = 2014 Energy Intensity(GJ/m2)

X[3] = Weekly Average Hours

X[1321] = Organization – Goderich Alexandra Marine and General Hospital (Yes/No)

	Predicted No (0)	Predicted Yes (1)
True No (0)	161 = TN	646 = FP
True Yes (1)	137 = FN	2050 = TP

Table 2.1 – Confusion Matrix for Emissions Predictor

Accuracy = (2050 + 161) / 2994 = 73.85%

Recall/Sensitivity = 2050 / (2050 + 137) = 93.74%

Specificity = 161 / (161 + 646) = 19.95%

Precision = 2050 / (2050 + 646) = 76.04%

Based on the results the model works moderately well in predicting those who have been successful in reducing emissions but is not very good for predicting those who have been unsuccessful in doing so.

Energy Consumption

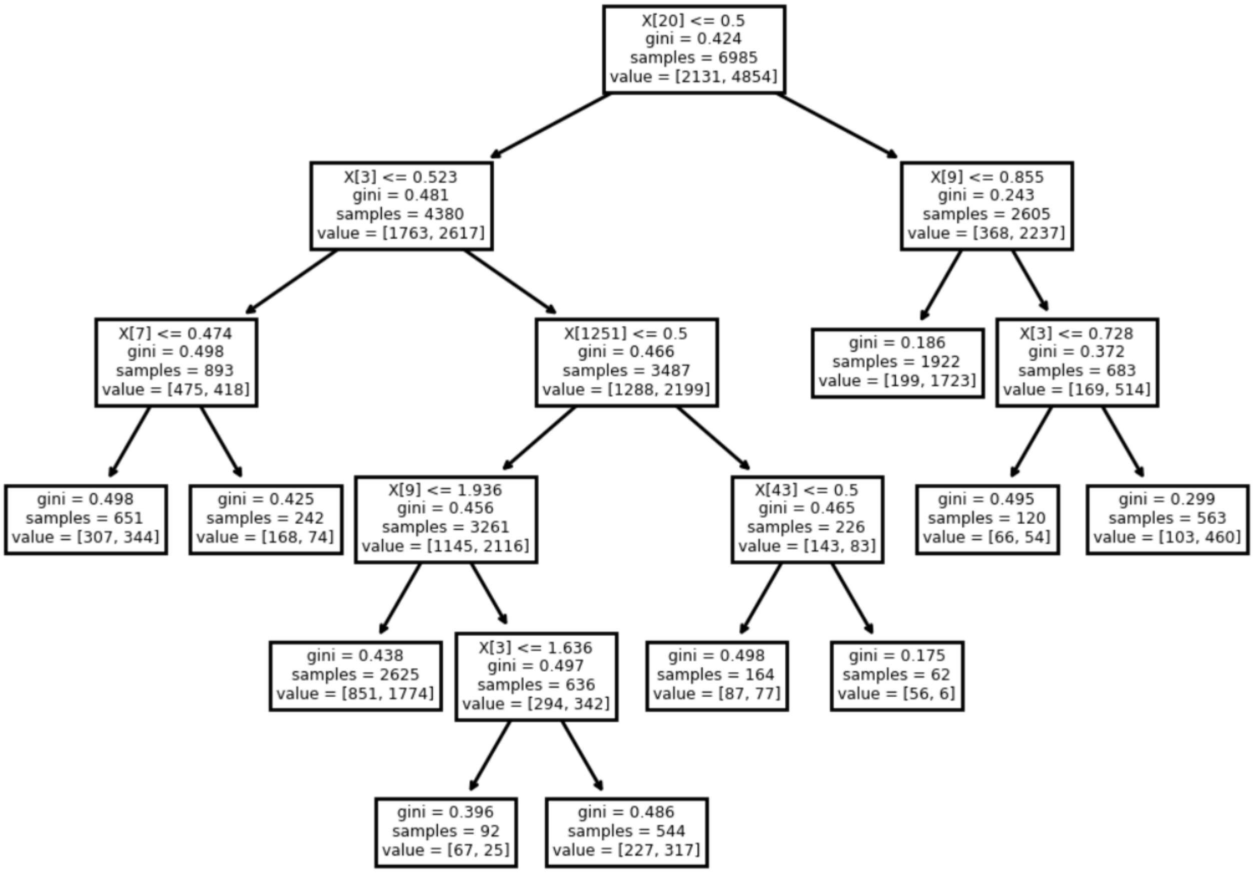


Figure 2.2 – Decision Tree predicting energy intensity results

X[20] = Propane heating type (Yes/No)

X[3] = Weekly Average Hours

X[9] = 2017 Energy Intensity (GJ/m2)

X[7] = 2016 Energy Intensity (GJ/m2)

X[1251] = Organization – City of St. Thomas (Yes/No)

X[43] = Operation Type – Indoor recreational Facilities (Yes/No)

	Predicted No (0)	Predicted Yes (1)
True No (0)	157 = TN	734 = FP
True Yes (1)	103 = FN	2000 = TP

Table 2.2 – Confusion Matrix for Emissions Predictor

Accuracy = (2000 + 157) / 2994 = 72.04%

Recall/Sensitivity = 2000 / (2000 + 103) = 951%

Specificity = 157 / (157 + 734) = 17.62%

Precision = 2000 / (2000 + 734) = 73.15

Similar results on the evaluations here and based on the results the model works moderately well in predicting those who have been successful in reducing energy consumption but is not very good for predicting those who have been unsuccessful in doing so.

Conclusions

From the work above we are able to see that the regulation intended to reduce energy consumption and greenhouse gas emissions has been able to accomplish results. We can see that majority of the Operations under the Ontario Broader public Sector were able to reduce their energy consumption and energy usage over the period of 6 years that we looked at. We can see that typically most of the energy consumption comes from the municipal sector while emissions are even across all sectors. We can see that the weekly average hours of operation is important in both models and intuitively makes sense that it plays a part in energy consumption and emissions. Finally, we have two models that help us predict who will be successful in reducing emissions and energy consumption.

Shortcomings and Continuity

Although we were able to predict moderately accurately based on the variables in the data who is successful in reducing emissions and energy usage, the models are not very well suited to predict who is unable to do so due to its slow specificity. Since we are dealing with a lot of numeric attributes a regression model should be created as well to see if we can create a better model for this purpose to perhaps improve upon our results. Even after dropping the duplicates in the data about 30% of the data was lost during the transformation and cleaning and we can look at how to retain more of the data to potentially improve the results.

References

Energy use and greenhouse gas emissions for the Broader Public Sector - Ontario Data Catalogue. (n.d.). Data.ontario.ca. Retrieved February 18, 2022, from https://data.ontario.ca/en/dataset/energy-use-and-greenhouse-gas-emissions-for-the-broader-public-sector

Ritchie, H., & Roser, M. (2020, May 11). *CO2 emissions*. Our World in Data. Retrieved February 17, 2022, from https://ourworldindata.org/co2-emissions?utm_source=squamish%20chief&utm_campaign=squamish%20chief&utm_medium=referral

Herzog, T. (2009). *World Greenhouse Gas Emissions in 2005*. http://pdf.wri.org/working_papers/world_greenhouse_gas_emissions_2005.pdf

Kennedy, C., Steinberger, J., Gasson, B., Hansen, Y., Hillman, T., HavránekM., Pataki, D., Phdungsilp, A., Ramaswami, A., & Mendez, G. V. (2009). Greenhouse Gas Emissions from Global Cities. *Environmental Science & Technology*, *43*(19), 7297–7302. https://doi.org/10.1021/es900213p

Fettweis, G. and Zimmermann, E., 2008. *ICT ENERGY CONSUMPTION – TRENDS AND CHALLENGES*. [online] Citeseerx.ist.psu.edu. Available at: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.139.4150&rep=rep1&type=pdf [Accessed 17 February 2022].

García-Martín, E., Rodrigues, C. F., Riley, G., & Grahn, H. (2019). Estimation of energy consumption in machine learning. *Journal of Parallel and Distributed Computing*, *134*, 75–88. https://doi.org/10.1016/j.jpdc.2019.07.007