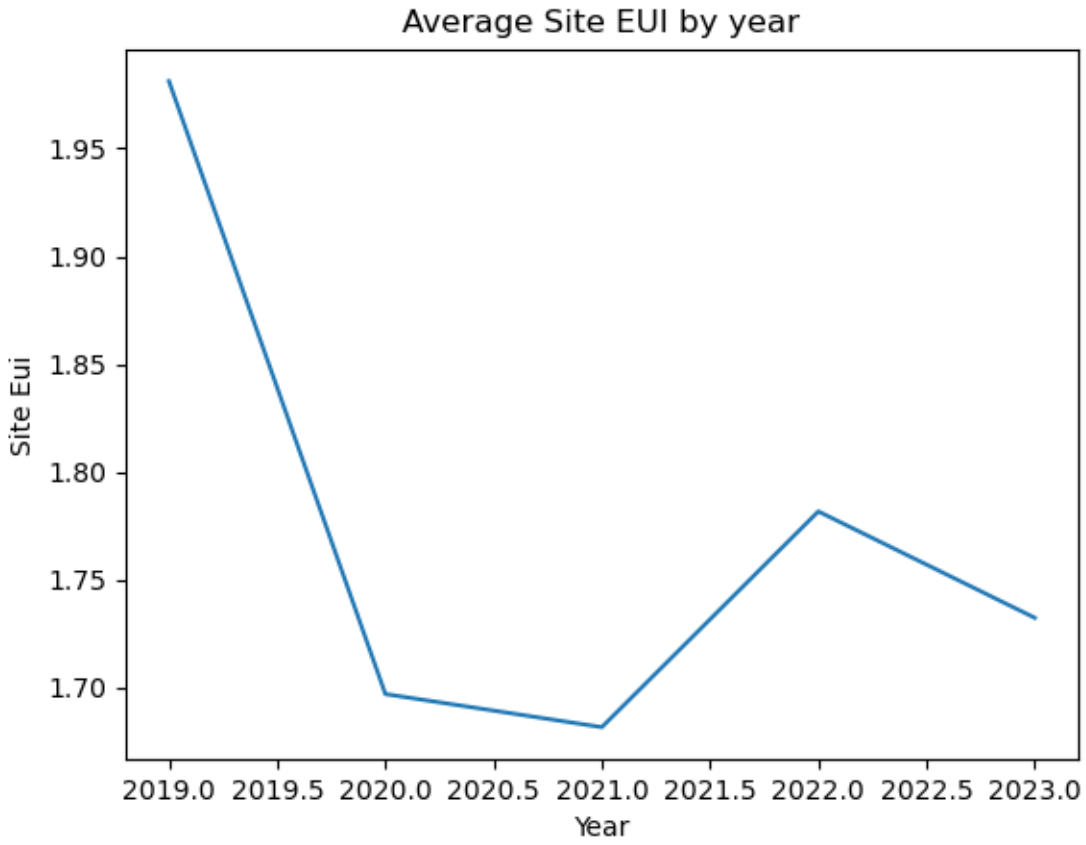
Data 601 HW4 Report

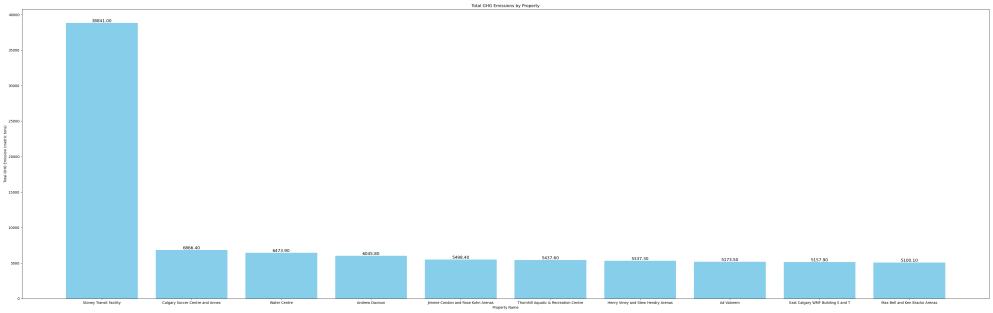
2025-02-02

HW4 REPORT

The Building_Energy_Benchmarking dataset represents the use and emission of energy, natural gas, etc by different properties in Calgary, Alberta.
Some columns containing a large amount of missing values were removed, these columns include 'Energy STAR Score','District Hot Water Use (GJ)','Electricity Use – Generated from Onsite Renewable Systems (kWh)','Green Power - Onsite and Offsite (kWh)' and 'Avoided Emissions - Onsite and Offsite Green Power (Metric Tons CO2e)'.
In evaluating the summary statistics for the numerical data of the dataset, certain trends are observed. The mean size of each property is about $4753m^2$ and the average amount of energy used annually is about $8266GJ$. The difference between the site and source energy use annually is about $2000GJ$.
The top 5 properties with the highest energy consumption are Stoney Transit Facility, Municipal Complex, Village Square Leisure Centre, Southland Leisure Centre and Foothills Aquatic Centre and Bauer and Bush Arenas.
Regex was then used to extract numeric values from text based numeric columns in the dataset. This was done by finding characters like , and - and replacing them with whitespace in order to extract just the numeric values of each relevant column. Regex was also used to standardize the text in the address and property name columns by removing special characters and unnecessary words.
The average site energy use divided by square meters shows a sharp decline after 2019, and while it does not reach that high again, there is an uptick in the average in 2022.



As observed in the bar chart, there is a massive discrepancy in the total ghg emission by each property as the Stoney Transit Facility emits almost 40,000 metric tons of GHG, which is almost 8 times more than the next highest facility. The Stoney Transit Facility is hub for all transit services in the city so this kind of discrepancy makes sense.



As emphasized by the correlation table, it is shown that there is a positive correlation between the site energy use and the property size. This indicates that the larger the property size, the more energy is being consumed. On the other hand, there isn't a strong positive correlation between the total emissions and the property size, so it cannot be assumed that the larger the property size the more the energy emission is. However, the correlation between the site energy use and the total emission would indicate that a good way to reduce the amount of emission would be to reduce the the amount of energy being used. According to the data, the best way to reduce the energy being used, and in turn, the amount of emission, would be to reduce the size of any given property.
A T-test was not performed using the feature 'Energy STAR Score' but the feature Site Energy use instead as

	Property GFA - Self-Reported (m^2)	Site Energ
Property GFA - Self-Reported (m^2)	1.000000	
Site Energy Use (GJ)	0.725977	
Total GHG Emissions (Metric Tons CO2e)	0.478433	

Energy STAR score was removed as it had too many missing values.