DATA 604

FINAL PROJECT REPORT

ANALYSIS OF ENERGY CONSUMPTION IN THE CITY OF CALGARY

Group 3: Code 404

Khushi Himanshu Dave

Jannatul Naeema

Shashank Kumar Srivastava

Zheyu Song

Introduction

The domain that we will be working with for this project is Energy and sustainability of Calgary. Energy plays a fundamental role in our lives; everything requires energy in one form or another. Canada is in the top five of natural gas producers in the world; two-thirds of which come from Alberta. In 2017, the energy sector made up 9.2%, or $175 Billion, of Canada's Gross Domestic Product (GDP) whereas, in Alberta the energy sector contributed 21.61% of provincial GDP. This is significantly more than in the rest of Canada; the oil and gas sector make up a major part of economic activity in Alberta. As one of the major cities in Alberta, Calgary has long been known as an energy city and took on a lot of initiatives to encourage in becoming more energy efficient eventually. In 2008, the City of Calgary developed the Sustainable Buildings Partnership Program to improve the performance of existing city infrastructure and support the sustainable building policy. This program's purpose is to identify and improve existing corporate infrastructure efficiency. These are proposed to be done using audits, alternative energy technologies, conservation, and energy efficiency upgrades. We focused on addressing this context and investigate into energy consumption situation in the City of Calgary.

Objectives

In our project, we want to analyze the energy consumption situations at different structures and facilities in Calgary. The goal of this project is to understand energy use for buildings and investigate the effects of it. We will use Python and SQL to perform these functions and analyze the energy use of buildings and structures aligned with the sustainable building policy. We have performed this investigation on each individual data set, as well as on merged datasets. This study is important to assess if the energy use of buildings and structures are in alignment with the sustainable building policy. Through this investigation, we aim to understand better energy efficiency and we aim to provide new insights as to whether energy efficiency needs to be improved.

Individual Datasets

We have selected four datasets for this project representing the energy consumptions at different capacities within the City of Calgary.

Dataset 1: Building energy benchmarking:

This dataset is an annual dataset from the year 2019 to 2021 and is the energy distribution for each individual building in the City of Calgary. It has 297 rows and 23 columns and contains data about each property. It has a property ID for each, and a classification of which type of property that specific property belongs to, which can be “Office,” “Residential,” “Fire Station,” etc. and which Year it was built in. It also has the Area of each building. For each of these properties, we have the total Site Energy use, the Electricity use, The Natural Gas use, and the District Hot water use, among others.

Dataset 2: Corporate Energy Consumption – City of Calgary:

We wanted to explore more into the energy consumption at city facilities and chose corporate energy consumption dataset. This dataset is open data available at city of Calgary webpage. This is in tabular format with 300k rows, and 9 columns collected over the period of 2014 to 2021. The total consumption of energy is recorded every month; these data are collected at 20 different business facilities; e.g., Calgary Housing, Calgary Transit, Fire Department, Park and open spaces, Waste and Recycling Services etc.

Dataset 3: Energy consumption of 6 Building Data - University of Calgary:

The data for this project was retrieved through the Office of Sustainability Campus as a Learning Lab initiative collected as energy consumption at different buildings daily. Due to the non-disclosure agreement, we could not present any link to this dataset. This dataset contains different sources of energy namely, heating, cooling, electricity, natural gas and domestic water. The data is available for six different buildings throughout the campus over the three years of 2018 to 2021.

Dataset 4: Current and Historical Alberta Weather Station Data – ACIS:

The dataset was retrieved from Alberta Agriculture, Forestry and Rural Economic Development, Alberta Climate Information Service (ACIS) while agreeing to the "ACIS Data Disclaimer & Terms of Use" https://acis.alberta.ca/acis/data-disclaimer.jsp allows us to operate this dataset with fully acknowledged and cited. In this part of study, we will use "Current and Historical Alberta Weather Station Data" (ACIS, November 2022). The Alberta Climate Information Service (ACIS) collects weather and climate data from a variety of meteorological stations operated by various government agencies. For the investigation of this dataset, we will be focusing on the variables: Date, Air Temp. Avg. (°C) for our weather normalization analysis. The dataset recorded daily temperature from 2014 to 2022 at Calgary International CS weather station.

Dataset division for each team member

Khushi Himanshu Dave: Dataset-1
Jannatul Naeema: Dataset-2
Shashank Kumar Srivastava: Dataset-3
Zheyu Song: Dataset-4

Guiding Questions

The research questions that we intend to discuss in this report via analysis is:

1. What is the correlation between Natural Gas and Electricity consumption for Annual data, Monthly data, and Campus data.

We chose this guiding question to explore the energy consumption from two main sources, we have used dataset 1, 2 and 3 to answer this guiding question.

2. What is the correlation between Normalized Energy consumption for Annual data and Monthly data of Calgary properties?

This guiding question is to better understand energy consumption of heating degree days and cooling degree days using dataset 1 and 2.

3. What is the correlation between Temperature of each data and energy consumption for every month?

We wanted to observe whether energy consumption has any dependency on outside temperature and which kind property or building type use most energy on a certain weather; here we also investigated the source of energy mostly used for the heating or cooling.

4. How did COVID-19 pandemic impact the energy use for buildings on campus?

During the covid time with lockdown and restrictions most people stayed at home, we wished to observe the effect of covid on University of Calgary Campus by checking up the energy use during that time and comparing it with pre-covid and post-covid periods.

5. Has the efficiency of heating energy usage on campus buildings improved from 2018 to 2022?

We focused on the energy consumption at different campus buildings because the University of Calgary is known to be a leading university in sustainability while undertaking steps to reduce dependency on carbon-based fuel and improve energy efficiency. We have used datasets 3 and 4 to answer this guiding question.

Data Cleaning, Wrangling and SQL queries

Dataset 1: The 'Energy Benchmark of Calgary' dataset was chosen as it had a lot of data related to the individual buildings across Calgary, and other details which can be useful to join the dataset with other datasets. We have merged this dataset with the Weather Station data and compared the values of this dataset with the corporate energy data, and the Campus data.

Individual Exploration included checking for null values, and cleaning and structuring the data using Excel and SQL. The 'District Hot Water' column was dropped as it contained a lot of null values. There was an exploratory analysis on the dataset, which answered a few questions for us. We found the Property Type that uses the most energy, which, using queries, we get as "Office" building type. There was another analysis on how Age of the building affects the energy use, for which we found that the Age of the building has no effect, after which we did another analysis for how the Area of that building affects energy use, but concluded that that too, has no effect. We also conducted analysis on how each building was dependent on Natural Gas and Electricity, and for results we get that buildings are about 50% dependent on Natural Gas, and 20%-45% dependent on electricity. After all this, there was work done using merged datasets.

```
Q1='''SELECT
    `Primary Property Type - Self Selected`,
    SUM(`Site Energy Use (GJ)`)
FROM
    energy_bench_data
GROUP BY `Primary Property Type - Self Selected`
ORDER BY 2;'''
query_table1 = pd.read_sql_query(Q1, engine)
display(query_table1)

#Citations
#https://www.w3schools.com/sql/sql_groupby.asp
#https://www.w3schools.com/sql/sql_count_avg_sum.asp
#https://www.w3schools.com/sql/sql_orderby.asp#:~:text=The%20ORDER%20BY%20keyword%20is,order%2C%20use%20the%20DESC%20keyword.
```

	Primary Property Type - Self Selected	SUM(`Site Energy Use (GJ)`)
0	Other	974.5
1	Mixed Use Property	2905.3
2	Performing Arts	4564.1
3	Museum	9774.7
4	Social/Meeting Hall	11149.8
5	Other - Public Services	11230.2
6	Other - Recreation	12933.1
7	Self-Storage Facility	17638.1
8	Indoor Arena	57191.8
9	Heated Swimming Pool	68406.6
10	Repair Services (Vehicle, Shoe, Locksmith, etc.)	69309.2
11	Non-Refrigerated Warehouse	86556.3
12	Fire Station	190105.3
13	Distribution Center	322866.8
14	Ice/Curling Rink	329438.7
15	Fitness Center/Health Club/Gym	523202.6
16	Office	620236.9

```
#Newest building
Q2 = '''SELECT
    `Primary Property Type - Self Selected`,
    MAX(`Year Built`),`Property GFA - Self-Reported (m²)`,
    `Site Energy Use (GJ)`
FROM
    energy_bench_data
WHERE
    `Primary Property Type - Self Selected` = 'Office'
    ORDER BY 2;'''
query_table2 = pd.read_sql_query(Q2, engine)
display(query_table2)

#Oldest building
Q3 = '''SELECT
    `Primary Property Type - Self Selected`,
    MIN(`Year Built`),`Property GFA - Self-Reported (m²)`,
    `Site Energy Use (GJ)`
FROM
    energy_bench_data
WHERE
    `Primary Property Type - Self Selected` = 'Office'
    ORDER BY 2;'''
query_table3 = pd.read_sql_query(Q3, engine)
display(query_table3)


#Citations
#https://www.w3schools.com/sql/sql_min_max.asp
#https://learnsql.com/blog/sql-min-max-functions/
```

	Primary Property Type - Self Selected	MAX(`Year Built`)	Property GFA - Self-Reported (m²)	Site Energy Use (GJ)
0	Office	2018	7770.0	10118.0

	Primary Property Type - Self Selected	MIN(`Year Built`)	Property GFA - Self-Reported (m²)	Site Energy Use (GJ)
0	Office	1919	7770.0	10118.0

```
Q6 ='''SELECT `Property Id`,`Natural Gas Use (GJ)`, `Site Energy Use (GJ)`,
      ROUND(`Natural Gas Use (GJ)` * 100.0 / `Site Energy Use (GJ)`, 1) AS Percent
FROM energy_bench_data;'''
query_table6 = pd.read_sql_query(Q6, engine)
display(query_table6)

#Citations
#https://stackoverflow.com/questions/36531361/calculate-percentage-between-two-columns-in-sql-query-as-another-column
```

	Property Id	Natural Gas Use (GJ)	Site Energy Use (GJ)	Percent
0	6169481	6308.2	10118.0	62.3
1	6305956	24.5	4792.7	0.5
2	6506773	7052.5	11983.9	58.8
3	6731628	2159.1	3653.5	59.1
4	6867796	363.0	506.1	71.7
...	...	...	...	...
292	21988624	964.5	1069.9	90.1
293	21988625	936.5	1238.3	75.6
294	21988627	835.2	920.2	90.8
295	21988628	688.4	1080.6	63.7
296	21988629	251.7	335.5	75.0

297 rows × 4 columns

```
Q8= '''SELECT `Property Id`,`ConvertedValues`, `Site Energy Use (GJ)`,
      ROUND(`ConvertedValues` * 100.0 / `Site Energy Use (GJ)`, 1) AS Percent
FROM (SELECT `Property Id`, `Electricity Use - Grid Purchase (kWh)`,`Site Energy Use (GJ)`,
        ROUND(`Electricity Use - Grid Purchase (kWh)` * 0.0036, 1) AS ConvertedValues
FROM energy_bench_data)a;'''
query_table8 = pd.read_sql_query(Q8, engine)
display(query_table8)

#Citations
#https://stackoverflow.com/questions/36531361/calculate-percentage-between-two-columns-in-sql-query-as-another-column
```

	Property Id	ConvertedValues	Site Energy Use (GJ)	Percent
0	6169481	3809.9	10118.0	37.7
1	6305956	1929.8	4792.7	40.3
2	6506773	4931.6	11983.9	41.2
3	6731628	1429.9	3653.5	39.1
4	6867796	143.2	506.1	28.3
...	...	...	...	...
292	21988624	105.4	1069.9	9.9
293	21988625	301.9	1238.3	24.4
294	21988627	84.9	920.2	9.2
295	21988628	387.4	1080.6	35.9
296	21988629	83.7	335.5	24.9

297 rows × 4 columns

Dataset 2: At first, we imported the dataset 'Corporate Energy Consumption' in both jupyter notebook and MySQL and checked the data for any missing or NaN values. We imported the dataset to the jupyter,

```
corporate_energy_consumption = pd.read_csv("corporate_energy_consumption.csv")
display(corporate_energy_consumption.head())
corporate_energy_consumption.info()
```

```
/tmp/ipykernel_89/3232737626.py:1: DtypeWarning: Columns (2) have mixed types. Specify dtype option on import or set low_memory=False.
  corporate_energy_consumption = pd.read_csv("corporate_energy_consumption.csv")
```

	Business_Unit_Desc	Facility_Name	Site_ID	Facility_Address	Energy_Description	Year	Month	Total_Consumption	Unit
0	Calgary Fire Department	ATCO VILLAGE (HOUSE)	20003498361	6015 23 AV SE	Electricity	2014	1	1883	kWh
1	Calgary Fire Department	ATCO VILLAGE (HOUSE)	20003498361	6015 23 AV SE	Electricity	2014	2	2320	kWh
2	Calgary Fire Department	ATCO VILLAGE (HOUSE)	20003498361	6015 23 AV SE	Electricity	2014	3	1657	kWh
3	Calgary Fire Department	ATCO VILLAGE (HOUSE)	20003498361	6015 23 AV SE	Electricity	2014	4	1107	kWh
4	Calgary Fire Department	ATCO VILLAGE (HOUSE)	20003498361	6015 23 AV SE	Electricity	2014	5	972	kWh

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 293680 entries, 0 to 293679
Data columns (total 9 columns):
 #   Column              Non-Null Count   Dtype
---  ------              --------------   -----
 0   Business_Unit_Desc  293680 non-null  object
 1   Facility_Name       293646 non-null  object
 2   Site_ID             293680 non-null  object
 3   Facility_Address    293646 non-null  object
 4   Energy_Description  293680 non-null  object
 5   Year                293680 non-null  int64
```

We converted the name of months from month column to numbers in excel and assigned number 1 for 'Jan', 2 for 'Feb' and so on. Since there is no missing value, we moved forward to the next steps. Later we investigated total energy consumption as natural gas, converted Natural Gas energy unit from GJ to kWh and we found from SQL SUM query that the biggest consumer of Natural gas in 2021 is Calgary Transit, in 2020 is Water Services, in 2019 and 2018 is Facility Management. Similarly, the highest consumer of electricity in 2021, 2020, 2019 and 2018 is Water Services. One example of query is mentioned below, we ran five similar queries to reach this decision.

Query to get total Natural Gas consumption in the year 2018 for Different Business Unit is shown below:

```
query5 = pd.read_sql_query('''SELECT Business_Unit_Desc, SUM(Total_Consumption*277.778) AS Natural_Gas_AVG_kWh,
Energy_Description FROM corporate_energy_consumption WHERE Energy_Description='Natural Gas' AND Year = 2018
GROUP BY Business_Unit_Desc ORDER BY Natural_Gas_AVG_kWh DESC;''', engine)
print (query5)
```

```
            Business_Unit_Desc  Natural_Gas_AVG_kWh Energy_Description
0          Facility Management         1.596085e+08        Natural Gas
1               Water Services         1.508135e+08        Natural Gas
2              Calgary Transit         1.290337e+08        Natural Gas
3               CPS - Bureaus          2.780724e+07        Natural Gas
4     Waste and Recycling Srvcs        2.673085e+07        Natural Gas
5              Calgary Housing         1.435307e+07        Natural Gas
6          Calgary Parking Auth        1.284668e+07        Natural Gas
7      Real Estate and Dev Serv        1.262445e+07        Natural Gas
8                     Mobility         1.198362e+07        Natural Gas
9       Parks and Open Spaces-PK       3.290280e+06        Natural Gas
10   Recreation and Social Prgms       2.260280e+06        Natural Gas
11    City and Regional Planning       3.997225e+05        Natural Gas
12         Green Line Operations       2.700002e+05        Natural Gas
13        Public Spaces Delivery       2.458335e+05        Natural Gas
14        Information Technology       8.611118e+03        Natural Gas
15       Calgary Fire Department       0.000000e+00        Natural Gas
```

We calculated the average energy consumption from both electricity and Natural Gas, we also converted the Natural Gas energy unit from GJ to kWh. Average Natural gas use for different business units from 2014 to 2022; and the Mobility consumed most over the years and similarly we ran the query for electricity usage from 2014 to 2022 and again mobility is at the top of the list as the average energy use. In our next step we conducted the analysis of average consumption of natural gas and electricity separately for the year of 2019 to 2021 and we plotted a few bar graphs that are displayed below.

```python
# Average Natural gas use for business units in 2019

query1 = pd.read_sql_query('''select Business_Unit_Desc, AVG(Total_Consumption*277.778) AS Natural_Gas_AVG_kWh,
Energy_Description FROM corporate_energy_consumption WHERE Energy_Description='Natural Gas' and Year =2019 GROUP BY
Business_Unit_Desc ORDER BY Year, Natural_Gas_AVG_kWh DESC;''', engine)
print (query1)
```

```
              Business_Unit_Desc  Natural_Gas_AVG_kWh Energy_Description
0                       Mobility         1.010742e+06        Natural Gas
1                Calgary Transit         1.915250e+05        Natural Gas
2        Waste and Recycling Srvcs       1.600356e+05        Natural Gas
3                 Water Services         1.259773e+05        Natural Gas
4            Calgary Parking Auth        1.255383e+05        Natural Gas
5                  CPS - Bureaus         7.885690e+04        Natural Gas
6      Recreation and Social Prgms       7.820182e+04        Natural Gas
7             Facility Management        7.215157e+04        Natural Gas
8         Real Estate and Dev Serv       6.683220e+04        Natural Gas
9         Parks and Open Spaces-PK       4.827936e+04        Natural Gas
10     City and Regional Planning        2.851012e+04        Natural Gas
11         Public Spaces Delivery        1.435186e+04        Natural Gas
12          Green Line Operations        5.108029e+03        Natural Gas
13                Calgary Housing        3.937857e+03        Natural Gas
14        Information Technology         4.629633e+02        Natural Gas
```

```python
# Average Natural gas use for business units in 2020

q1 = pd.read_sql_query('''select Business_Unit_Desc, AVG(Total_Consumption*277.778) AS Natural_Gas_AVG_kWh,
Energy_Description FROM corporate_energy_consumption WHERE Energy_Description='Natural Gas' and Year =2020 GROUP BY
Business_Unit_Desc ORDER BY Year, Natural_Gas_AVG_kWh DESC;''', engine)
print (q1)
```

```
              Business_Unit_Desc  Natural_Gas_AVG_kWh Energy_Description
0                       Mobility         1.176181e+06        Natural Gas
1                Calgary Transit         1.858973e+05        Natural Gas
2                 Water Services         1.352709e+05        Natural Gas
3        Waste and Recycling Srvcs       1.333396e+05        Natural Gas
4            Calgary Parking Auth        9.917500e+04        Natural Gas
5                  CPS - Bureaus         7.145922e+04        Natural Gas
6             Facility Management        5.871356e+04        Natural Gas
7         Real Estate and Dev Serv       5.813910e+04        Natural Gas
8         Parks and Open Spaces-PK       4.934074e+04        Natural Gas
9      Recreation and Social Prgms       1.854168e+04        Natural Gas
10         Public Spaces Delivery        6.148153e+03        Natural Gas
11                Calgary Housing        3.742711e+03        Natural Gas
12          Green Line Operations        2.990454e+03        Natural Gas
13        Information Technology         4.282411e+02        Natural Gas
```

```python
# Average Natural gas use for business units in 2021

q2 = pd.read_sql_query('''select Business_Unit_Desc, AVG(Total_Consumption*277.778) AS Natural_Gas_AVG_kWh,
Energy_Description FROM corporate_energy_consumption WHERE Energy_Description='Natural Gas' and Year =2021 GROUP BY
Business_Unit_Desc ORDER BY Year, Natural_Gas_AVG_kWh DESC;''', engine)
print (q2)
```

```
             Business_Unit_Desc  Natural_Gas_AVG_kWh Energy_Description
0                      Mobility        966273.921167        Natural Gas
1               Calgary Transit        189772.781591        Natural Gas
2                Water Services        129529.884963        Natural Gas
3     Waste and Recycling Srvcs        122932.197111        Natural Gas
4          Calgary Parking Auth         88752.507649        Natural Gas
5                 CPS - Bureaus         82594.804854        Natural Gas
6           Facility Management         58745.269633        Natural Gas
7      Parks and Open Spaces-PK         45125.697476        Natural Gas
8      Real Estate and Dev Serv         31305.693061        Natural Gas
9        Public Spaces Delivery          9537.044667        Natural Gas
10      Calgary Fire Department          5000.004000        Natural Gas
11              Calgary Housing          3590.606000        Natural Gas
12        Green Line Operations          3223.596543        Natural Gas
13       Information Technology           416.667000        Natural Gas
```

Fig 2: Average Natural Gas Consumption for the year 2019-2021

```python
# Average Electricity use for business units in 2019

query2 = pd.read_sql_query('''select Business_Unit_Desc, AVG(Total_Consumption) AS Electricity_AVG_kWh, Energy_Description
FROM corporate_energy_consumption WHERE Energy_Description='Electricity' and Year =2019 GROUP BY Business_Unit_Desc ORDER BY
Electricity_AVG_kWh DESC;''', engine)
print (query2)
```

```
             Business_Unit_Desc  Electricity_AVG_kWh Energy_Description
0                      Mobility          123773.9708        Electricity
1                 CPS - Bureaus           57549.7270        Electricity
2                Water Services           46966.5089        Electricity
3   Recreation and Social Prgms           41938.4286        Electricity
4               Calgary Transit           36831.3518        Electricity
5     Waste and Recycling Srvcs           31404.8184        Electricity
6    City and Regional Planning           22731.1818        Electricity
7           Facility Management           22223.7243        Electricity
8          Calgary Parking Auth           19781.3576        Electricity
9      Real Estate and Dev Serv           15180.4185        Electricity
10            Downtown Strategy            6024.7500        Electricity
11       Information Technology            3917.4167        Electricity
12     Parks and Open Spaces-PK            1859.2228        Electricity
13        Green Line Operations            1508.6486        Electricity
14       Public Spaces Delivery             784.0000        Electricity
15      Calgary Fire Department             526.8750        Electricity
16          Utilities Delivery              300.6143        Electricity
```

```python
# Average Electricity use for business units in 2020

q4 = pd.read_sql_query('''select Business_Unit_Desc, AVG(Total_Consumption) AS Electricity_AVG_kWh, Energy_Description
FROM corporate_energy_consumption WHERE Energy_Description='Electricity' and Year =2020 GROUP BY Business_Unit_Desc ORDER BY
Electricity_AVG_kWh DESC;''', engine)
print (q4)
```

```
           Business_Unit_Desc  Electricity_AVG_kWh Energy_Description
0                    Mobility          107243.7352        Electricity
1               CPS - Bureaus           53939.8145        Electricity
2              Water Services           47576.6527        Electricity
3             Calgary Transit           28742.8617        Electricity
4   Waste and Recycling Srvcs           25258.5285        Electricity
5         Facility Management           20114.8874        Electricity
6        Calgary Parking Auth           18344.6495        Electricity
7    Real Estate and Dev Serv           13228.3321        Electricity
8           Downtown Strategy           10322.0000        Electricity
9      Information Technology            3868.4167        Electricity
10   Parks and Open Spaces-PK            1750.0864        Electricity
11     Public Spaces Delivery            1280.6800        Electricity
12      Green Line Operations             985.1923        Electricity
13    Calgary Fire Department             542.9167        Electricity
14         Utilities Delivery             229.5843        Electricity
```

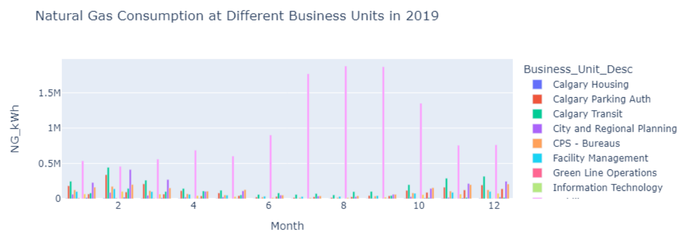
```python
# Average Electricity use for business units in 2021

q5 = pd.read_sql_query('''select Business_Unit_Desc, AVG(Total_Consumption) AS Electricity_AVG_kWh, Energy_Description
FROM corporate_energy_consumption WHERE Energy_Description='Electricity' and Year =2021 GROUP BY Business_Unit_Desc ORDER BY
Electricity_AVG_kWh DESC;''', engine)
print (q5)
```

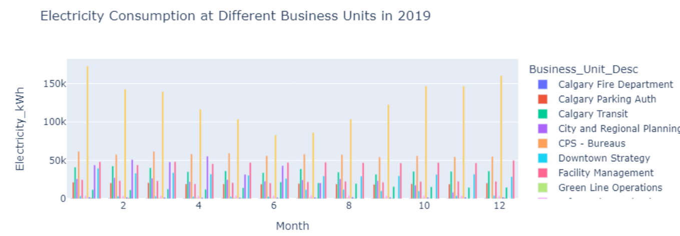
```
           Business_Unit_Desc  Electricity_AVG_kWh Energy_Description
0                    Mobility          105178.1905        Electricity
1               CPS - Bureaus           61047.8849        Electricity
2              Water Services           49741.1746        Electricity
3   Waste and Recycling Srvcs           27513.4187        Electricity
4             Calgary Transit           27011.5488        Electricity
5        Calgary Parking Auth           20983.5066        Electricity
6         Facility Management           19947.7679        Electricity
7           Downtown Strategy           11354.8824        Electricity
8    Real Estate and Dev Serv            6607.5335        Electricity
9      Information Technology            3966.4167        Electricity
10   Parks and Open Spaces-PK            1637.4993        Electricity
11     Public Spaces Delivery            1606.3030        Electricity
12      Green Line Operations             665.4706        Electricity
13    Calgary Fire Department             506.1667        Electricity
14         Utilities Delivery              90.7131        Electricity
```

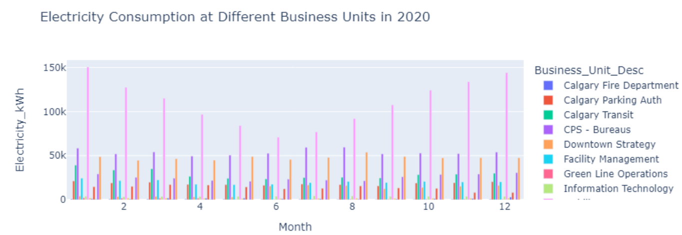
Fig 3: Average Electricity Consumption for the year 2019-2021

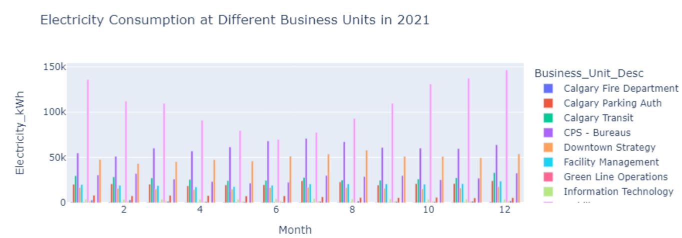
As the next step we plotted the average energy consumption query for different years; the queries of plots and plots are displayed below:

```python
import plotly.express as px
fig1 = px.line(normalize_energy_2019, x="Month", y="Normalized_Energy_Consumption",
              title="Normalized Energy in 2019")
fig1.show()

import plotly.express as px
fig1 = px.line(normalize_energy_2020, x="Month", y="Normalized_Energy_Consumption", title="Normalized Energy in 2020")
fig1.show()

import plotly.express as px
fig1 = px.line(normalize_energy_2021, x="Month", y="Normalized_Energy_Consumption", title="Normalized Energy in 2021")
fig1.show()
```









From all the graphs it is evident that if we consider average energy consumption as natural gas or electricity, the mobility is the highest consumer. Natural gas consumption is higher in the winter months and lower in the summer months for all business units except mobility. Similarly, electricity consumption is higher in the winter months and lower in the summer months; probably because of the excessive use of heating in the colder months for all different types of business units.

Dataset 3: The dataset retrieved for six buildings across the university contained the data for 5 types of energy in one spreadsheet. The initial data wrangling took place in Microsoft Excel. The data column was common to every building from 2018 to 2021. Each building's data was transferred into another spreadsheet with date, building and 5 types of energy as columns' name. The 5 types of energy were namely; Heating water, Cooling water, Electricity, Natural gas and Domestic water.

	index	Date	Building	Domestic_water	Electricity1	Chilled_wat	Heating_wat	Natural_g
0	0	01-10-2018	EEEL	6.39	13240.10	4920.66	16244.46	1319.45
1	1	02-10-2018	EEEL	49.96	13108.56	4882.56	17772.24	1413.89
2	2	03-10-2018	EEEL	101.98	13307.63	5163.91	16852.79	1288.89
3	3	04-10-2018	EEEL	110.85	12917.94	5117.02	15450.01	1422.22
4	4	05-10-2018	EEEL	106.50	10770.25	5225.46	13422.23	1413.89
...	...	...	...	...	...	...	...	...
6571	6571	26-09-2021	ADMIN	0.1	2658.78	2060.29	0.00	0.00
6572	6572	27-09-2021	ADMIN	6.1	3365.23	3443.59	0.00	0.00
6573	6573	28-09-2021	ADMIN	5.1	3323.29	1567.93	0.00	0.00
6574	6574	29-09-2021	ADMIN	4.5	3365.50	975.93	0.00	0.00
6575	6575	30-09-2021	ADMIN	1	3263.78	1523.97	0.00	0.00

When the data was loaded in python the numerical values were not read as numbers as they contained comma. Further wrangling was required. We used pandas to replace the commas with null value. After the replacement was done in every column, our dataset was then pushed onto the SQL server. We wanted to see the energy consumption trend for each building. A new table was created with columns as Date, Building and Total energy as the sum of all the types of energy. Arranging the maximum energy consumption in descending order we saw the overall consumption of energy of each building over three years.

	building	SUM(total)
0	ENGG	55136658.16
1	EEEL	35530751.76
2	ADMIN	20592558.92
3	SOSCI	19225006.65
4	PROFAC	18934155.18
5	AUHALL	4587295.89

The Engineering building used the most energy, which can be accounted for by the heavy machinery used by the students, instructors, and PhD scholars during the practical work. Which can be regarded for the next building as well. Aurora Hall is a residential building, which explains why the energy consumption is the least when compared to corporate or educational buildings. It consists of simple appliances such as bulbs, microwave, oven etc.

We compared water usage of the residential building because from the above inferences we could see that there is no match in energy consumption of residential building by educational buildings and corporate buildings. They consume far more energy than a residential building. Hence the study of water consumption was interesting.

	Building	TOTAL
0	ENGG	45798.11
1	SOSCI	41417.38
2	AUHALL	20737.91
3	EEEL	17933.85
4	PROFAC	9827.85
5	ADMIN	3328.05

We can see that water consumption by Aurora Hall was far more than most of the corporate as well as educational buildings. This can be regarded by students using a lot of water in their daily life, for cooking, washing, or bathing.

We even saw that the impact of COVID was quite visible in the university as well. There was a considerable dip in energy consumption of all educational type buildings and the residential building but not so notable change in the office type building which signifies that the offices were still in running condition, even in lockdown. This is further discussed in the data exploration section.

Dataset 4: We chose this dataset because it contains daily average temperature data, which can be used to calculate Heating Degree Days. This information can then be used to weather normalize heating energy data and analyze its efficiency. Through working with this dataset, we learned how to use SQL queries to calculate the Heating Degree Days by using the average temperature data and the methodology of weather normalization. Our knowledge of these concepts allowed us to understand better the relationship between weather conditions and heating energy usage. It has provided us with valuable insights that we can use to improve the efficiency of heating systems.

Due to the ACIS site's technical limitations, we could only download three years of data at a time. We downloaded the data for 2018 to 2020 and 2021 in two parts and merged them into the original four-year weather data after processing using Pandas. ACIS provided Source Flag and Recorded Completeness data based on rigour, and we confirmed the availability of these data after checking this information. We removed these columns for the convenience of our analysis. By converting the data to a Date Time data type and extracting only the date, we prepared the data for analysis and interpretation. In our energy efficiency analysis, we studied the change in heating energy efficiency for each campus building.

```python
csv1 = pd.read_csv("ACISDailyData-20180101-20201231-PID143937394.csv", encoding= 'unicode_escape')
#Separating datasets by Station Name
df_cspre21 = csv1[csv1['Station Name'] == 'Calgary Int\'L CS']
df_cupre21 = csv1[csv1['Station Name'] == 'Cop Upper']
csv2 = pd.read_csv("ACISDailyData-20210101-20211231-PID144244925.csv", encoding= 'unicode_escape')
df_cs21 = csv2[csv2['Station Name'] == 'Calgary Int\'L CS']
df_cu21 = csv2[csv2['Station Name'] == 'Cop Upper']
df_cs = pd.concat([df_cspre21, df_cs21])
df_cu = pd.concat([df_cupre21, df_cu21])
df_weather = pd.concat([df_cs, df_cu])
display(df_weather.head())
#df_weather.info()
```

	Station Name	Date (Local Standard Time)	Air Temp. Min. (°C)	Air Temp. Min. Source Flag	Air Temp. Min. Record Completeness (%)	Air Temp. Max. (°C)	Air Temp. Max. Source Flag	Air Temp. Max. Record Completeness (%)	Air Temp. Avg. (°C)	Air Temp. Avg. Source Flag	Air Temp. Avg. Record Completeness (%)	Relative Humidity Avg. (%)	Relative Humidity Avg. Source Flag	Relative Humidity Avg. Record Completeness (%)	Wind Speed 10 m Avg. (km/h)	Wind Speed 10 m Avg. Source Flag	Wind Speed 10 m Avg. Record Completeness (%)
0	Calgary Int'L CS	01-January-2018	-32.2	ACTUAL	100.0	-3.6	ACTUAL	100.0	-18.7	AGGREGATED	100.0	70.2	AGGREGATED	100.0	6.8	AGGREGATED	100.0
1	Calgary Int'L CS	02-January-2018	-11.6	ACTUAL	100.0	0.8	ACTUAL	100.0	-5.8	AGGREGATED	100.0	66.9	AGGREGATED	100.0	5.9	AGGREGATED	100.0
2	Calgary Int'L CS	03-January-2018	-10.1	ACTUAL	100.0	6.4	ACTUAL	100.0	-3.7	AGGREGATED	100.0	71.2	AGGREGATED	100.0	4.5	AGGREGATED	100.0
3	Calgary Int'L CS	04-January-2018	-12.0	ACTUAL	100.0	4.0	ACTUAL	100.0	-6.4	AGGREGATED	100.0	80.3	AGGREGATED	100.0	4.2	AGGREGATED	100.0

```python
df_weather['Date'] = pd.to_datetime(df_weather['Date']).dt.normalize()
df_weather.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2922 entries, 0 to 729
Data columns (total 17 columns):
 #   Column                                        Non-Null Count  Dtype
---  ------                                        --------------  -----
 0   Station_Name                                  2922 non-null   object
 1   Date                                          2922 non-null   datetime64[ns]
 2   Air_Temp_Min                                  2922 non-null   float64
 3   Air Temp. Min. Source Flag                    2922 non-null   object
 4   Air Temp. Min. Record Completeness (%)        2922 non-null   float64
 5   Air_Temp_Max                                  2922 non-null   float64
 6   Air Temp. Max. Source Flag                    2922 non-null   object
 7   Air Temp. Max. Record Completeness (%)        2922 non-null   float64
 8   Air_Temp_Avg                                  2922 non-null   float64
 9   Air Temp. Avg. Source Flag                    2922 non-null   object
 10  Air Temp. Avg. Record Completeness (%)        2922 non-null   float64
 11  Relative_Humidity_Avg                         2922 non-null   float64
 12  Relative Humidity Avg. Source Flag            2922 non-null   object
 13  Relative Humidity Avg. Record Completeness (%) 2922 non-null   float64
 14  Wind_Speed_10m_Avg                            2922 non-null   float64
 15  Wind Speed 10 m Avg. Source Flag              2922 non-null   object
 16  Wind Speed 10 m Avg. Record Completeness (%)  2922 non-null   float64
dtypes: datetime64[ns](1), float64(10), object(6)
memory usage: 410.9+ KB
```

```
create_statement2 = '''
ALTER TABLE weather
    DROP COLUMN IF EXISTS `Air Temp. Min. Source Flag`,
    DROP COLUMN IF EXISTS `Air Temp. Min. Record Completeness (%%)`,
    DROP COLUMN IF EXISTS `Air Temp. Max. Source Flag`,
    DROP COLUMN IF EXISTS `Air Temp. Max. Record Completeness (%%)`,
    DROP COLUMN IF EXISTS `Air Temp. Avg. Source Flag`,
    DROP COLUMN IF EXISTS `Air Temp. Avg. Record Completeness (%%)`,
    DROP COLUMN IF EXISTS `Relative Humidity Avg. Source Flag`,
    DROP COLUMN IF EXISTS `Relative Humidity Avg. Record Completeness (%%)`,
    DROP COLUMN IF EXISTS `Wind Speed 10 m Avg. Source Flag`,
    DROP COLUMN IF EXISTS `Wind Speed 10 m Avg. Record Completeness (%%)`;
'''
sql.execute(create_statement2,engine)
```

We used the weather table to calculate the daily Heating Degree Days from 2018 to 2021 by subtracting the baseline temperature of 18 degrees from the daily average temperature.

Data Exploration

Q1. What is the correlation between Natural Gas and Electricity consumption for Annual data, Monthly data, and Campus data

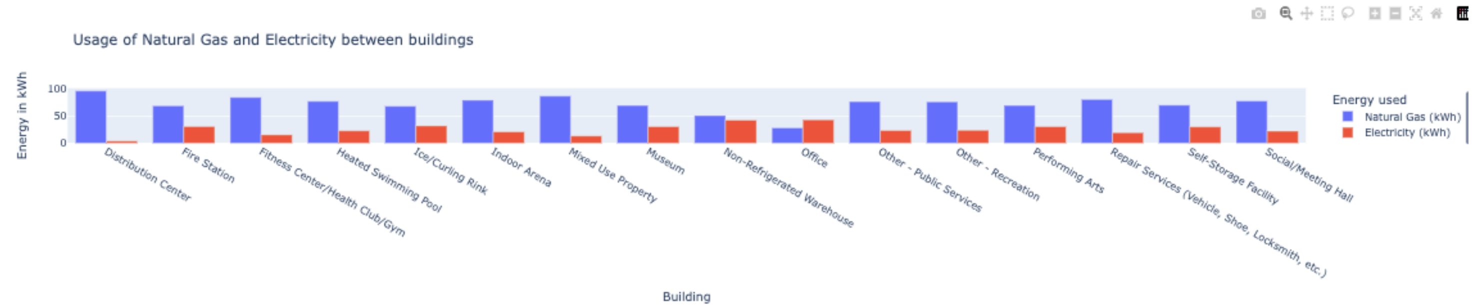
```
Q3= '''SELECT `Primary Property Type - Self Selected`, (ROUND(sum(`ConvertedValues`) * 100.0 / sum(`Site Energy Use (GJ)`), 1)) AS ElecPercent
FROM (SELECT `Primary Property Type - Self Selected`,`Property Id`, `Electricity Use - Grid Purchase (kWh)`,`Site Energy Use (GJ)`,
        ROUND(`Electricity Use - Grid Purchase (kWh)` * 0.0036, 1) AS ConvertedValues FROM energy_bench_data)a
        WHERE `Primary Property Type - Self Selected` not like 'Other' group by 1;'''


energypercent = pd.read_sql_query(Q3, engine)
display(energypercent)
```

	Primary Property Type - Self Selected	ElecPercent
0	Distribution Center	3.6
1	Fire Station	30.6
2	Fitness Center/Health Club/Gym	15.2
3	Heated Swimming Pool	22.7
4	Ice/Curling Rink	31.8
5	Indoor Arena	20.7
6	Mixed Use Property	13.1
7	Museum	30.4
8	Non-Refrigerated Warehouse	42.4
9	Office	43.0
10	Other - Public Services	23.3
11	Other - Recreation	23.7
12	Performing Arts	30.4
13	Repair Services (Vehicle, Shoe, Locksmith, etc.)	19.2
14	Self-Storage Facility	29.9
15	Social/Meeting Hall	22.1

```
Q4='''sELECT `Primary Property Type - Self Selected`,(ROUND(sum(`Natural Gas Use (GJ)`) * 100.0 / sum(`Site Energy Use (GJ)`), 1)) AS NGPercent
FROM energy_bench_data WHERE `Primary Property Type - Self Selected` not like 'Other' group by 1;'''
ngpercent = pd.read_sql_query(Q4, engine)
display(ngpercent)
```

	Primary Property Type - Self Selected	NGPercent
0	Distribution Center	96.4
1	Fire Station	69.0
2	Fitness Center/Health Club/Gym	84.5
3	Heated Swimming Pool	77.3
4	Ice/Curling Rink	68.2
5	Indoor Arena	79.3
6	Mixed Use Property	86.9
7	Museum	69.6
8	Non-Refrigerated Warehouse	50.8
9	Office	28.2
10	Other - Public Services	76.7
11	Other - Recreation	76.3
12	Performing Arts	69.6
13	Repair Services (Vehicle, Shoe, Locksmith, etc.)	80.8
14	Self-Storage Facility	70.1
15	Social/Meeting Hall	77.9

```python
import plotly.graph_objects as go
fig = go.Figure(data=[
    go.Bar(name='Natural Gas (kWh)', x=Newdf.Building_Type, y=Newdf.loc[:,"NGPercent"]),
    go.Bar(name='Electricity (kWh)', x=Newdf.Building_Type, y=Newdf.loc[:,"ElecPercent"]),
])

fig.update_layout(barmode='group')
fig.update_layout(legend_title_text = "Energy used",title="Usage of Natural Gas and Electricity between buildings")
fig.update_xaxes(title_text="Building")
fig.update_yaxes(title_text="Energy in kWh")
fig.show()
```



The percentage of energy calculated from dataset 2 over the years 2014 to 2022 for different corporate buildings

```python
percentage_energy1 = '''SELECT ng_yearly.Year,(SELECT sum(electricity_yearly.Electricity_kWh)*100/
(sum(electricity_yearly.Electricity_kWh)+sum(ng_yearly.NG_kWh)))
AS Percentage_Electricity, (SELECT sum(ng_yearly.NG_kWh)*100/(sum(electricity_yearly.Electricity_kWh)
+sum(ng_yearly.NG_kWh))) AS Percentage_NG FROM ng_yearly JOIN electricity_yearly ON
ng_yearly.Business_Unit_Desc = electricity_yearly.Business_Unit_Desc GROUP BY ng_yearly.Year;'''

percentage_energy1 = pd.read_sql_query(percentage_energy1, engine)
(percentage_energy1)
```

	Year	Percentage_Electricity	Percentage_NG
0	2014	17.077275	82.922725
1	2015	17.912601	82.087399
2	2016	17.662471	82.337529
3	2017	18.375692	81.624308
4	2018	18.125230	81.874770
5	2019	17.902627	82.097373
6	2020	17.131366	82.868634
7	2021	18.496247	81.503753
8	2022	20.071557	79.928443

In summary the percentage of energy consumption as natural gas and electricity from dataset 1 and dataset 2 is presented in the table below:

	Natural Gas Consumption Percentage	Electricity Consumption Percentage
Percentage of Dependency at Different Properties at City of Calgary	50%	20-45%
Percentage of Dependency at Different Corporate Buildings at City of Calgary	75-85%	15-20%

Natural gas consumption is significantly higher than electricity consumption at different kinds of properties and different types of business buildings. So, the dependency on natural gas is slightly higher at the corporate building in ratio of total energy usage than other kinds of properties.

Q2. What is the correlation between Normalized Energy consumption for Annual data and Monthly data of Calgary properties

Yearly Normalized Energy Consumption:

We joined dataset 1 and dataset 4 on SQL for the years of 2019, 2020 and 2021 on the year column. Once the datasets are joined, we calculated ratio of the total energy used and annual heating degree days, we conducted the same calculation for all three years individually to get normalized energy consumption for the perspective year. From our analysis we can note that the Normalized energy remains similar in all three years but is a little lower in 2020 which can be due to the impact of COVID-19 pandemic.

```
#Q9= '''CREATE TABLE annual_energy_dummy AS (SELECT SUM(`Site Energy Use (GJ)`), `Year Ending` FROM energy_bench_data GROUP BY `Year Ending`);select * from annual_energy1;

normalised='''select ROUND(T/A,2) as NormalisedEnergy,ROUND(T,2) as Total_energy,ROUND(A,2) as annual_hdd ,Y
from
(
SELECT
    annual_energy1.`Total Energy` as T,
    yearlyhdd.`Annual_Heating_Degree_Days` as A,
    annual_energy1.`Year Ending` as Y
FROM
    annual_energy1 inner join yearlyhdd
on yearlyhdd.`Year` = annual_energy1.`Year Ending`)S;'''
normalised = pd.read_sql_query(normalised, engine)
display(normalised)
```

	NormalisedEnergy	Total_energy	annual_hdd	Y
0	155.84	837892.7	5376.6	2019
1	147.10	731769.8	4974.7	2020
2	155.95	768821.5	4929.9	2021

Monthly Normalized Energy Consumption:

We joined dataset 2 and dataset 4 on SQL for the years of 2019, 2020 and 2021 on the month column. Here is one query to join the corporate energy consumption dataset and temperature dataset for the year of 2019:

```
# Joining corporate_energy_consumption and Temp_data

temp_elec_2019 = pd.read_sql_query('''SELECT corporate_energy_consumption.Business_Unit_Desc,
corporate_energy_consumption.Energy_Description, corporate_energy_consumption.Total_Consumption AS electricity_kWh,
corporate_energy_consumption.Month, temperature_data.Monthly_Air_Temp,temperature_data.Monthly_Cooling_Degree_Days,
temperature_data.Monthly_Heating_Degree_Days FROM corporate_energy_consumption JOIN temperature_data ON
corporate_energy_consumption.Month = temperature_data.Month WHERE corporate_energy_consumption.Energy_Description = 'Electricity'
AND corporate_energy_consumption.Year = 2019 GROUP BY corporate_energy_consumption.Business_Unit_Desc, MONTH;''', engine)
print (temp_elec_2019)
```

```
          Business_Unit_Desc Energy_Description  electricity_kWh  Month  \
0    Calgary Fire Department        Electricity              654      1
1    Calgary Fire Department        Electricity              543      2
2    Calgary Fire Department        Electricity              460      3
3    Calgary Fire Department        Electricity              381      4
4    Calgary Fire Department        Electricity              355      5
..                       ...                ...              ...    ...
193           Water Services        Electricity                8      8
194           Water Services        Electricity                7      9
195           Water Services        Electricity               65     10
196           Water Services        Electricity              104     11
197           Water Services        Electricity              136     12

     Monthly_Air_Temp  Monthly_Cooling_Degree_Days  \
0           -6.867742                          0.0
1          -12.400000                          0.0
2           -5.806452                          0.0
3            0.976667                          0.0
4           14.122581                         11.2
..                ...                          ...
```

Once the datasets are joined, we calculated ratio of the total energy used and monthly heating degree days, we conducted the same calculation for all three years individually to get normalized energy consumption for the perspective year. The steps to normalize the energy consumption of buildings based on HDD and CDD are as taking the sum of the total heating and cooling degree days for any required period and then divide the total kWh consumed in the same period by the total number of heating and cooling degree day.

The queries to calculate normalized energy consumption for the years 2019-2021 are shown below:

```
normalize_energy_2019 = pd.read_sql_query('''SELECT total_2019.Total_2019 AS Total_Energy_2019, temp_2019.Monthly_Heating_Degree_Days,
(total_2019.Total_2019/temp_2019.Monthly_Heating_Degree_Days) AS Normalized_Energy_Consumption, total_2019.Month
FROM total_2019 JOIN temp_2019 ON total_2019.Month = temp_2019.Month GROUP BY Month;''', engine)
print(normalize_energy_2019)
```

	Total_Energy_2019	Monthly_Heating_Degree_Days	Normalized_Energy_Consumption	Month
0	2.401771e+06	680.1	3531.497024	1
1	3.112509e+06	1022.7	3043.423247	2
2	2.443406e+06	660.1	3701.569851	3
3	1.946187e+06	395.2	4924.562951	4
4	1.712425e+06	281.8	6076.740352	5

```
normalize_energy_2020 = pd.read_sql_query('''SELECT total_2020.Total_2020 AS Total_Energy_2020, temp_2020.Monthly_Heating_Degree_Days,
(total_2020.Total_2020/temp_2020.Monthly_Heating_Degree_Days) AS Normalized_Energy_Consumption, total_2020.Month
FROM total_2020 JOIN temp_2020 ON total_2020.Month = temp_2020.Month GROUP BY Month;''', engine)
print(normalize_energy_2020)
```

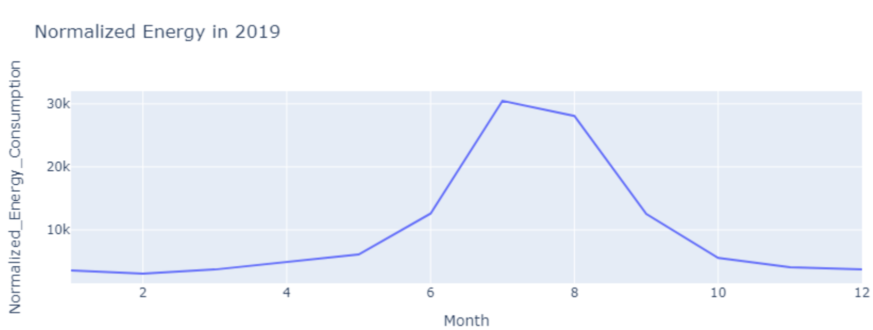
	Total_Energy_2020	Monthly_Heating_Degree_Days	Normalized_Energy_Consumption	Month
0	2.709753e+06	830.9	3261.226725	1
1	2.232144e+06	628.4	3552.106383	2
2	2.358481e+06	699.1	3373.596044	3
3	1.883781e+06	471.7	3993.599795	4
4	1.516940e+06	263.2	5763.449811	5

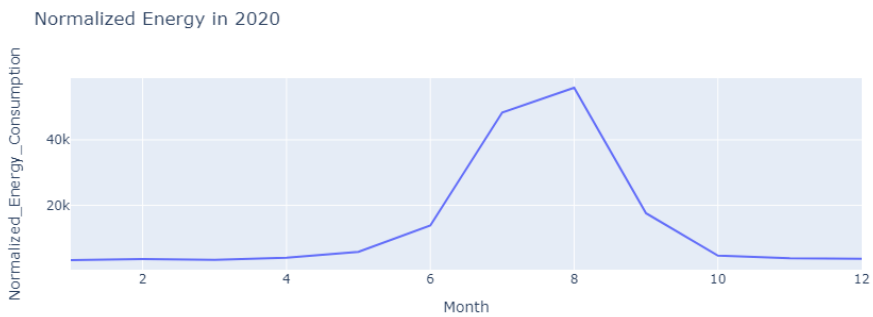
```
normalize_energy_2021 = pd.read_sql_query('''SELECT total_2021.Total_2021 AS Total_Energy_2021, temp_2021.Monthly_Heating_Degree_Days,
(total_2021.Total_2021/temp_2021.Monthly_Heating_Degree_Days) AS Normalized_Energy_Consumption, total_2021.Month
FROM total_2021 JOIN temp_2021 ON total_2021.Month = temp_2021.Month GROUP BY Month;''', engine)
print(normalize_energy_2021)
```

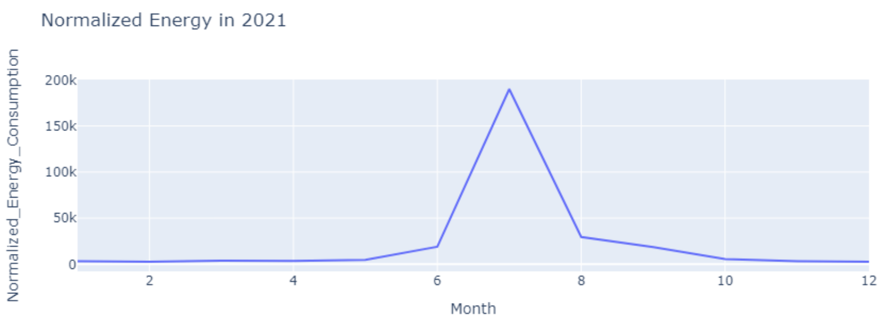
	Total_Energy_2021	Monthly_Heating_Degree_Days	Normalized_Energy_Consumption	Month
0	2.223702e+06	698.4	3183.994681	1
1	2.316584e+06	857.6	2701.240986	2
2	1.914181e+06	502.3	3810.832999	3

The graphs we plotted from the queries are displayed below:

```
import plotly.express as px
fig1 = px.line(normalize_energy_2019, x="Month", y="Normalized_Energy_Consumption",
              title="Normalized Energy in 2019")
fig1.show()

import plotly.express as px
fig1 = px.line(normalize_energy_2020, x="Month", y="Normalized_Energy_Consumption", title="Normalized Energy in 2020")
fig1.show()

import plotly.express as px
fig1 = px.line(normalize_energy_2021, x="Month", y="Normalized_Energy_Consumption", title="Normalized Energy in 2021")
fig1.show()
```







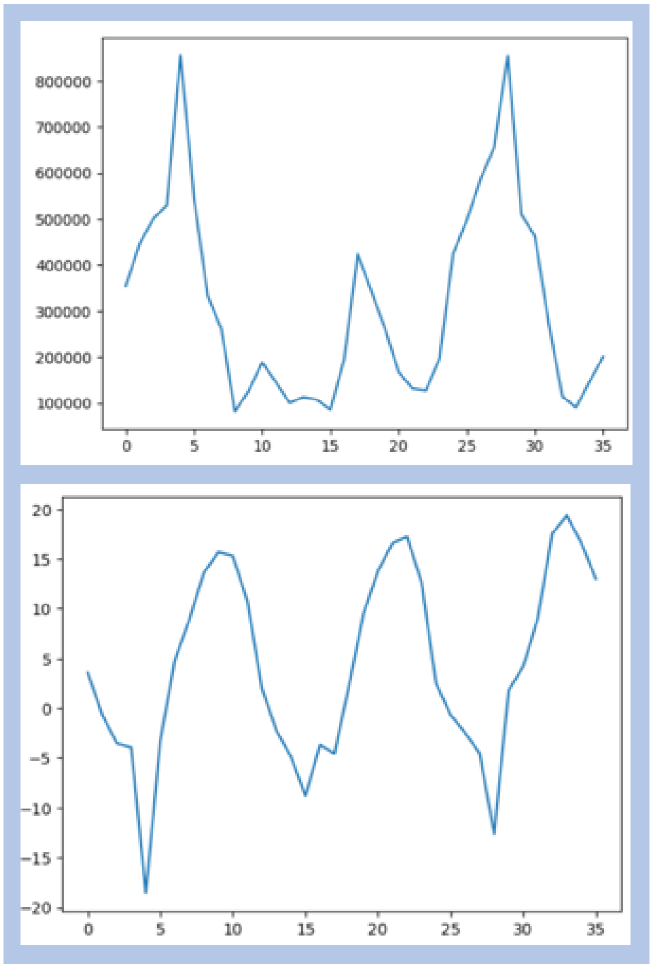
The normalized energy for the year 2019-2021 is summarized below:

	2019	2020	2021
Average Normalized Energy Consumption (kWh/Degree Days)	9855.05	13791.62	23905.64
Maximum Normalized Energy Consumption (kWh/Degree Days)	30491.41 in July	55901.20 in August	189984.42 in July
Minimum Normalized Energy Consumption (kWh/Degree Days)	3043.42 in February	3261.23 in January	2701.24 in February

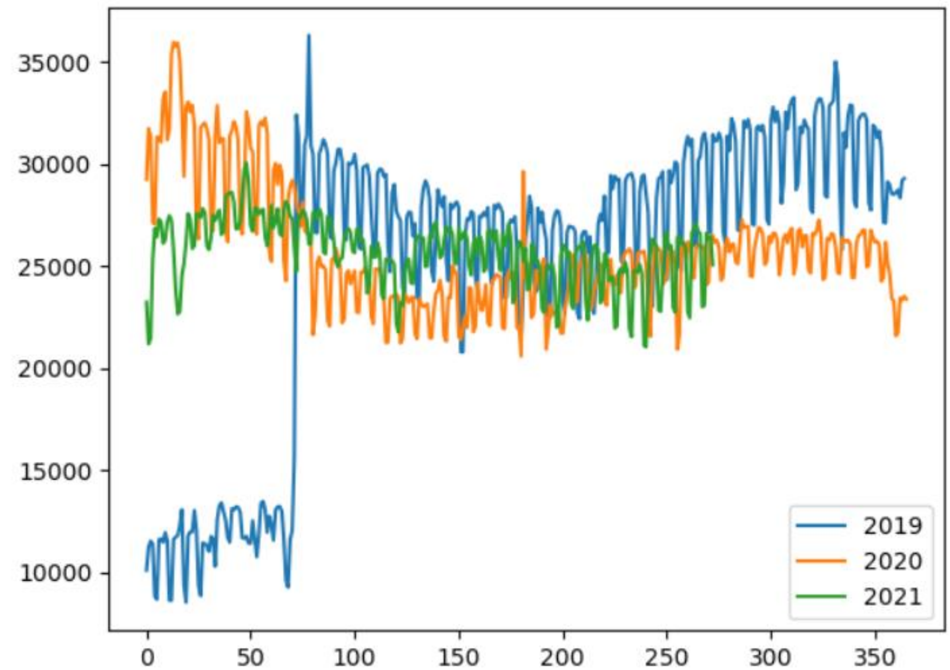
We can interpret from the graphs that normalized energy consumption is higher in summer months than in winter. The reason behind this lies in the definition of normalized energy consumption. In general weather normalization is a process that adjusts actual energy or peak outcomes to what would have happened under normal weather conditions. We can observe from the graph of each year that normalized energy consumption is at peaks over the summer months which indicates lower efficiency. In conclusion of this section, it is evident from our analysis that average normalized energy consumption is increasing over the years from 2019 to 2021 which indicates that the energy efficiency is decreasing over the years.

Q3. What is the correlation between Temperature of each data and energy?

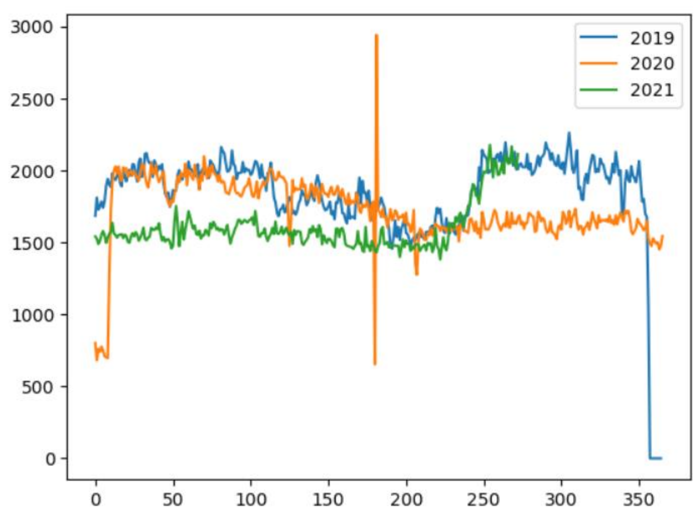
When we tried to find out the correlation between the temperature and a few of the top consumed energy types such as electricity, there was a negative or no correlation present. This can be explained by taking a close look at the spread of the data. There was no anomaly in the data before and after the COVID phase but during that the consumption considerably dropped but the outside temperature its trend as it is used to every year. Nonetheless some buildings do exhibit some correlation where when the temperature dips the energy consumption increases.



Q4. How did COVID-19 pandemic impact the energy use for buildings on campus?

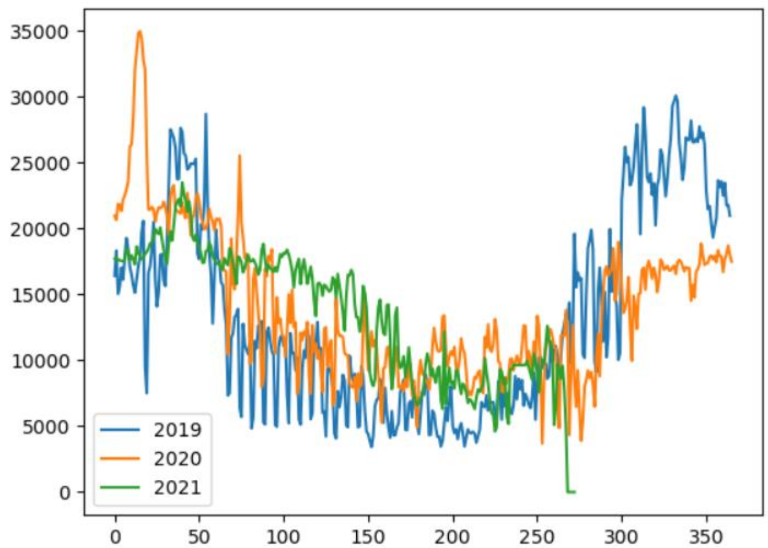


The above graph has been plotted between daily electricity consumption and the number of days in 3 years. Here we can see that during the year 2020 there has been a noticeable drop in electricity consumption for the Engineering building and even the consumption fluctuations were far less than 2019. Similarly, we can see the same trend of other types of energy for different buildings.



Looking at Aurora Hall, a residential building, in the middle of the year when the students migrated to their homes, we can see that there has been a considerable dip in electricity consumption. Further we can also see that in the middle of 2021 when students began to come to university, the consumption of electricity went up the charts again.

Now, looking at an office-type building on campus, The Admin building, the drops were not significant here.



This may be because during COVID lockdown the college office was open to cater for the students and professors' needs.

Q5. Has the efficiency of heating energy usage on campus buildings improved from 2018 to 2022?

When comparing the efficiency of heating energy use in the same building at different times, it is not accurate to compare the total amount of energy used. This is because external factors, such as the weather, can significantly impact the energy required to heat a building. For example, a building may use more energy to heat itself on a frigid day than it would on a mild day. To accurately compare the efficiency of heating energy use at different times, it is necessary to normalize the data by considering the weather conditions. This is typically done by using a weather normalization method.

Weather Normalization for Energy Consumption

Weather normalization is a technique used in building energy efficiency to account for variations in weather when comparing the energy consumption of different buildings or over different periods. Weather normalization aims to isolate weather's effects on energy consumption and adjust for these effects so that the energy consumption data is more accurately comparable by using heating degree days to normalize the energy consumption data. This analysis used HDD to normalize the heating energy consumption data for all six campus buildings. This allowed us to compare the energy efficiency of the building for different years and identify potential areas for improvement. By accounting for differences in weather, we can more accurately determine the actual efficiency of the building and identify potential areas for improvement.

Calculate the Degree Days

Heating degree days (HDD) are a measure of how cold the weather is during a specific period relative to a base temperature (WatchWire, 2018). For example, a building that uses more energy on a cold day with many heating degree days can be compared to a building that uses less energy on a warm day with fewer heating degree days. This allows for a fair comparison of energy consumption across different weather conditions.

The average daily temperature is subtracted from a base temperature (18 degrees Celsius in Canada) to calculate HDD. The result is zero if the average daily temperature is above the base temperature. If it is below the base temperature, the result is the number of degrees below the base temperature. HDDs are typically summed up over a specific time period to determine the total heating energy needed for a building. In our analysis, we used HDD to normalize the heating energy consumption data for our study period. This allowed us to compare the energy consumption of different buildings and identify opportunities for improving energy efficiency. To do this, we calculated the HDD for each day in our study period and summed them up to get the total HDD for the period.

```
create_statement1 = '''
CREATE TABLE IF NOT EXISTS xiaoyu18 AS (
select Station_Name, Date,Air_Temp_Avg, Air_Temp_Min, Air_Temp_Max,  abs(18- s1.Air_Temp_Avg) as Heating_Degree_Days, 0 as c,  Relative_Humidity_Avg, Wind_Speed_10m_Avg
FROM station1 s1
Where s1.Air_Temp_Avg <= 18
);
'''


create_statement2 = '''
CREATE TABLE IF NOT EXISTS dayu18 AS (
select Station_Name, Date,Air_Temp_Avg, Air_Temp_Min, Air_Temp_Max, 0 as Heating_Degree_Days  , (s1.Air_Temp_Avg- 18) as Cooling_Degree_Days , Relative_Humidity_Avg, Wind_Speed_10m_Avg
FROM station1 s1
Where s1.Air_Temp_Avg > 18
);
'''
create_statement3 = '''
CREATE TABLE IF NOT EXISTS finaldd AS (
select * from xiaoyu18 union all select * from dayu18
order by `Date`);
'''
create_statement4 = '''
CREATE TABLE IF NOT EXISTS yearly AS (
SELECT Station_Name, year(`Date`) as Year,avg(Air_Temp_Avg) Annual_Air_Temp, sum(Heating_Degree_Days) as Annual_Heating_Degree_Days , sum(c) as Annual_Cooling_Degree_Days
From finaldd
Group by year(`Date`));
'''
create_statement5 = '''
CREATE TABLE IF NOT EXISTS monthly AS (
SELECT Station_Name, year(`Date`) as Year ,month(`Date`) as Month,avg(Air_Temp_Avg)  Monthly_Air_Temp, sum(Heating_Degree_Days) as Monthly_Heating_Degree_Days , sum(c) as  Monthly_Coolin
From finaldd
Group by year(`Date`), month(`Date`));
'''
```

```
statement = '''
select Station_Name, Date,Heating_Degree_Days  FROM finaldd w
;
'''


df_b1energynormalized = pd.read_sql_query(statement, engine)
display(df_b1energynormalized.head())
```

	Station_Name	Date	Heating_Degree_Days
0	Calgary Int'L CS	2018-01-01	36.7
1	Calgary Int'L CS	2018-01-02	23.8
2	Calgary Int'L CS	2018-01-03	21.7
3	Calgary Int'L CS	2018-01-04	24.4
4	Calgary Int'L CS	2018-01-05	19.6

Calculation of Normalized Energy Consumption using Weather Normalization Method

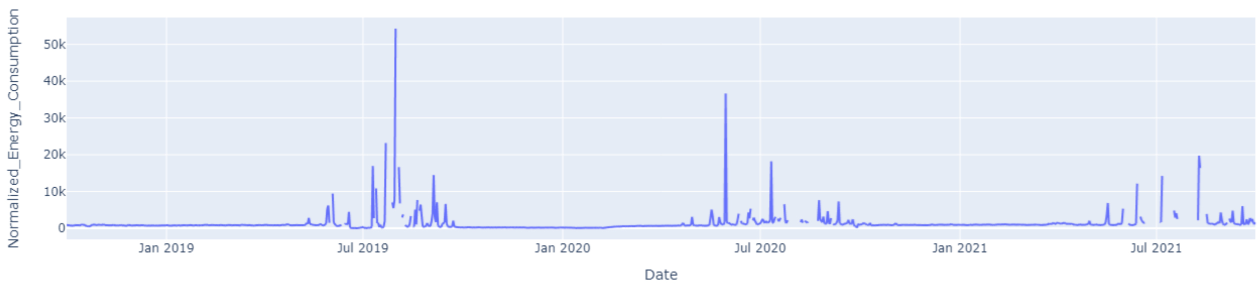
After calculating the heating degree days (HDD) for each day in the period this study covers, we used the daily HDD values to calculate the weather normalized heating energy consumption for six buildings on campus. This allowed us to compare the energy efficiency of the buildings and identify potential areas for improvement. By dividing the total heating energy consumption for each building by the total HDD for the period, we can obtain the weather normalized heating energy consumption for each building. This allowed us to compare the energy efficiency of the buildings on an equal footing since the effects of weather on energy consumption were accounted for in the normalization process. By comparing the weather normalized energy consumption of the buildings, including the same building for different years, we can identify which buildings were more efficient and which had the potential for improvement.

```
create_statement3 = '''
select b1.Date, b1.Heating_kWh, w.Air_Temp_Avg, w.Heating_Degree_Days
, (b1.Heating_kWh/ nullif(w.Heating_Degree_Days, 0)) as Normalized_Energy_Consumption
FROM b1 INNER JOIN finaldd w
ON b1.Date = w.Date
where month(b1.Date) between 10 and 12 or month(b1.Date) between 1 and 5
;
'''


df_b1energynormalized = pd.read_sql_query(create_statement3, engine)
display(df_b1energynormalized.head())
```

	Date	Heating_kWh	Air_Temp_Avg	Heating_Degree_Days	Normalized_Energy_Consumption
0	2018-10-01	16244.45744	-1.2	19.2	846.065492
1	2018-10-02	17772.23644	-2.9	20.9	850.346241
2	2018-10-03	16852.79126	-3.3	21.3	791.210857
3	2018-10-04	15450.01236	-2.2	20.2	764.852097
4	2018-10-05	13422.23296	-0.9	18.9	710.171056

Heating Energy Efficiency Analysis on Campus Building

In this part of the analysis, we used a line plot to visualize the daily normalized energy consumption data for six buildings on campus from October 2018 to September 2021. However, we observed that the plot contained many outliers, which made it challenging to analyze the trend. After further investigation, we discovered that the outliers were concentrated in the data from June to September, which corresponded to the summer months. This is because the heating degree days (HDD) during the summer are typically less than one, which means that dividing the energy consumption by a value less than one will result in a large number. Therefore, we decided to focus on the winter data for our analysis from October to May. This allowed us to avoid the effects of the summer months and more accurately analyze the trend in the heating energy consumption of the buildings.



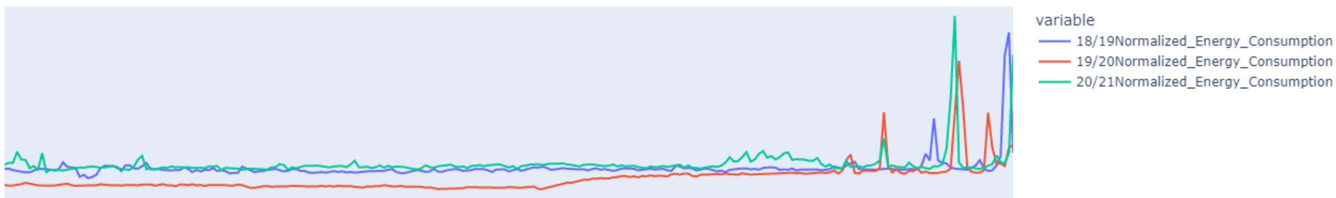
To analyze the trend of heating energy efficiency in campus buildings, we have plotted three plots for each group of buildings.

The first plot is a line graph that shows the Normalized Energy consumption of Building 1 for three winter seasons, from October to May. The plot includes three lines, one for each winter season: 18/19, 19/20, and 20/21. The second plot is a bar graph that shows the monthly Normalized Energy consumption values for the three winter seasons. The third plot is a bar graph of the annual winter Normalized Energy consumption for 2018, 2019, and 2020.

For each building, the SQL query for data wrangling for each group of plots is similar, joined with different building tables. So, we only show the data wrangling for plotting Building 1 as an example.

```
create_statement1 = '''
select (b1.Heating_kWh/ nullif(w.Heating_Degree_Days, 0)) as '18/19Normalized_Energy_Consumption'
FROM b1 INNER JOIN (select Date, Heating_Degree_Days from finaldd where Date between '2018-10-01' and '2019-05-31' ) as w
ON b1.Date = w.Date
where month(w.Date) between 10 and 12 or month(w.Date) between 1 and 5
and (b1.Heating_kWh/ nullif(w.Heating_Degree_Days, 0)) < 20000
;
'''
create_statement2 = '''
select (b1.Heating_kWh/ nullif(w.Heating_Degree_Days, 0)) as '19/20Normalized_Energy_Consumption'
FROM b1 INNER JOIN (select Date, Heating_Degree_Days from finaldd where Date between '2019-10-01' and '2020-05-31' ) as w
ON b1.Date = w.Date
where month(w.Date) between 10 and 12 or month(w.Date) between 1 and 5
and (b1.Heating_kWh/ nullif(w.Heating_Degree_Days, 0)) < 20000

;
'''
create_statement3 = '''
select (b1.Heating_kWh/ nullif(w.Heating_Degree_Days, 0)) as '20/21Normalized_Energy_Consumption'
FROM b1 INNER JOIN (select Date, Heating_Degree_Days from finaldd where Date between '2020-10-01' and '2021-05-31' ) as w
ON b1.Date = w.Date
where month(w.Date) between 10 and 12 or month(w.Date) between 1 and 5
and (b1.Heating_kWh/ nullif(w.Heating_Degree_Days, 0)) < 20000
;
'''
df1 = pd.read_sql_query(create_statement1, engine)
df2 = pd.read_sql_query(create_statement2, engine)
df3 = pd.read_sql_query(create_statement3, engine)
dfa = pd.concat([df1, df2], axis=1)
dfb = pd.concat([dfa, df3], axis=1)
fig = px.line(dfb, y=dfb.columns, title='Building 1 Winter Daily Normalized Heating Energy Consumption')
fig.update_xaxes(visible=False)
fig.update_yaxes(visible=False)
fig.show()
```

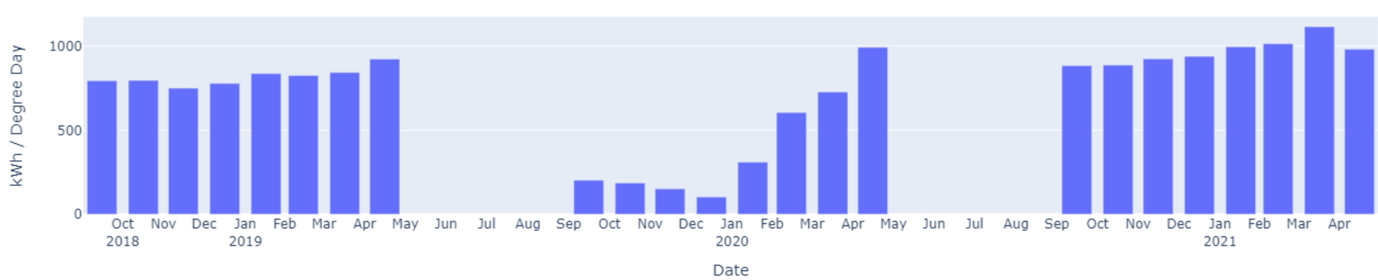
Building 1 Winter Daily Normalized Heating Energy Consumption




```
create_statement3 = '''
select DATE_FORMAT(w.Date, '%Y-%m')  AS Date ,
avg(Air_Temp_Avg) as Air_Temp_Avg , sum(Heating_Degree_Days) as Heating_Degree_Days, sum(Heating_kWh) as Heating_kWh,
(sum(Heating_kWh)/sum(Heating_Degree_Days)) as Normalized_Energy_Consumption
FROM b1 INNER JOIN finaldd w
ON b1.Date = w.Date
where month(b1.Date) between 10 and 12 or month(b1.Date) between 1 and 5
Group by year(w.`Date`), month(w.`Date`)
;
'''
df_b1nem = pd.read_sql_query(create_statement3, engine)
display(df_b1nem.head())
fig = px.bar(df_b1nem, x='Date', y="Normalized_Energy_Consumption",hover_data={"Date": "|%B %d, %Y"},
                                                     labels={"Normalized_Energy_Consumption": "kWh / Degree Day",},
  title='Building 1 October to May Normalized Heating Energy Consumption From 2018 to 2021')
fig.update_xaxes(
    dtick="M1",
    tickformat="%b\n%Y",
    ticklabelmode="period")
fig.show()
```

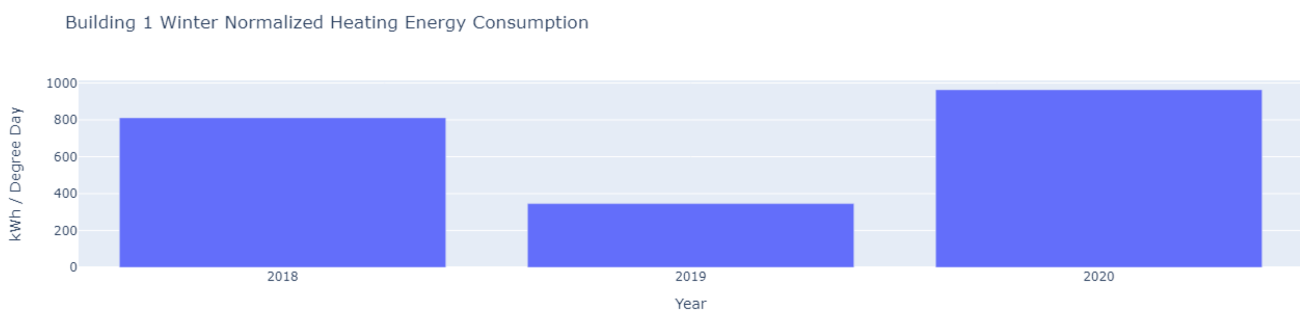
	Date	Air_Temp_Avg	Heating_Degree_Days	Heating_kWh	Normalized_Energy_Consumption
0	2018-10	3.593548	446.6	354675.28374	794.167675
1	2018-11	-0.656667	559.7	445608.68982	796.156316
2	2018-12	-3.516129	667.0	500208.73350	749.938131
3	2019-01	-3.938710	680.1	529461.53468	778.505418
4	2019-02	-18.525000	1022.7	855939.57364	836.941013

Building 1 October to May Normalized Heating Energy Consumption From 2018 to 2021

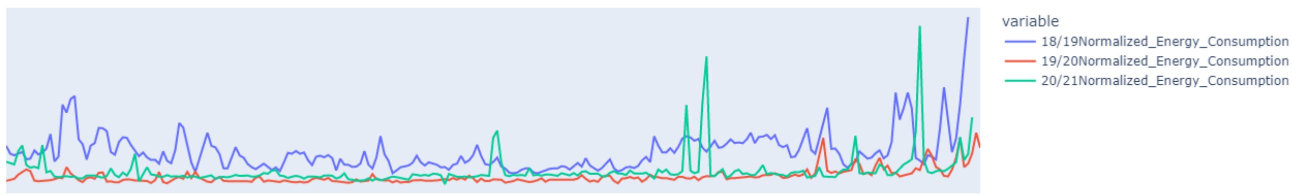


```python
create_statement3 = '''
select '2018'  AS Date , avg(Air_Temp_Avg) as Air_Temp_Avg , sum(Heating_Degree_Days) as Heating_Degree_Days,
sum(Heating_kWh) as Heating_kWh, (sum(Heating_kWh)/sum(Heating_Degree_Days)) as Normalized_Energy_Consumption
FROM b1 INNER JOIN finaldd w ON b1.Date = w.Date
Where w.Date between '2018-10-01' and '2019-05-31'
union
select '2019'  AS Date , avg(Air_Temp_Avg) as Air_Temp_Avg , sum(Heating_Degree_Days) as Heating_Degree_Days,
sum(Heating_kWh) as Heating_kWh, (sum(Heating_kWh)/sum(Heating_Degree_Days)) as Normalized_Energy_Consumption
FROM b1 INNER JOIN finaldd w ON b1.Date = w.Date
Where w.Date between '2019-10-01' and '2020-05-31'
union
select '2020'  AS Date , avg(Air_Temp_Avg) as Air_Temp_Avg , sum(Heating_Degree_Days) as Heating_Degree_Days,
sum(Heating_kWh) as Heating_kWh, (sum(Heating_kWh)/sum(Heating_Degree_Days)) as Normalized_Energy_Consumption
FROM b1 INNER JOIN finaldd w ON b1.Date = w.Date
Where w.Date between '2020-10-01' and '2021-05-31'
;
'''
df_b1nem = pd.read_sql_query(create_statement3, engine)
display(df_b1nem)
fig = px.bar(df_b1nem, x='Date', y="Normalized_Energy_Consumption",hover_data={"Date": "|%B %d, %Y"},
labels={                   "Normalized_Energy_Consumption": "kWh / Degree Day",     "Date": "Year",
             ,title='Building 1 Winter Normalized Heating Energy Consumption')
fig.update_xaxes(
    dtick="M1",
    tickformat="%b\n%Y",
    ticklabelmode="period")
fig.show()
```

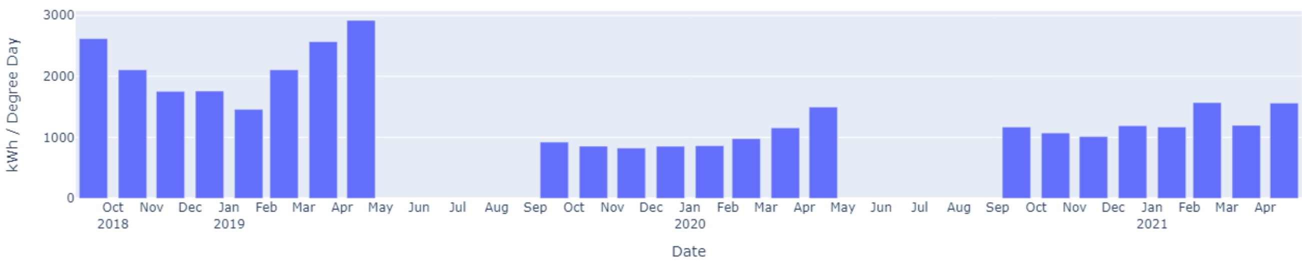
	Date	Air_Temp_Avg	Heating_Degree_Days	Heating_kWh	Normalized_Energy_Consumption
0	2018	-1.395885	4713.2	3.824050e+06	811.349037
1	2019	-1.283197	4705.1	1.627922e+06	345.991001
2	2020	-0.218107	4427.0	4.264969e+06	963.399269



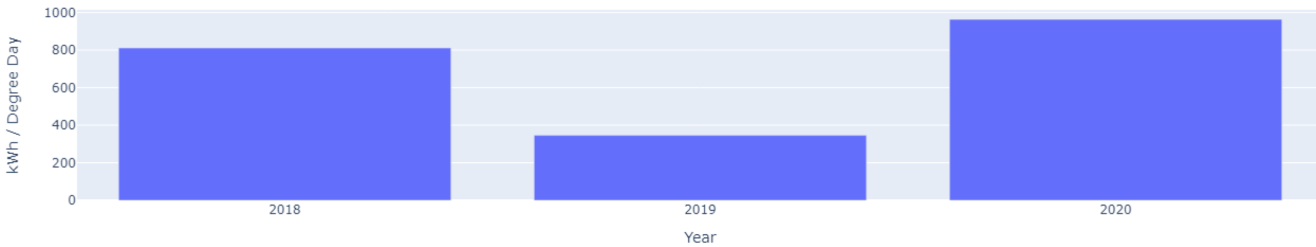
Building 2 Winter Daily Normalized Heating Energy Consumption



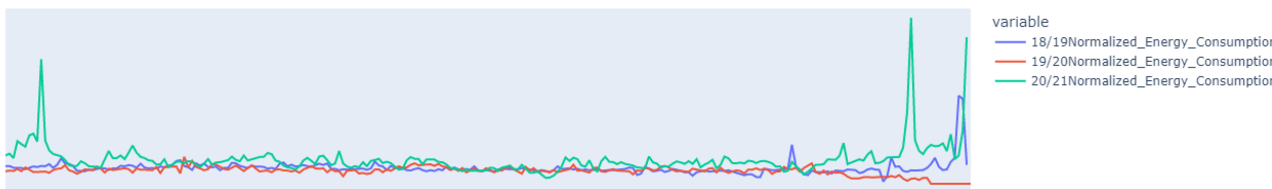
Building 2 October to May Normalized Heating Energy Consumption From 2018 to 2021



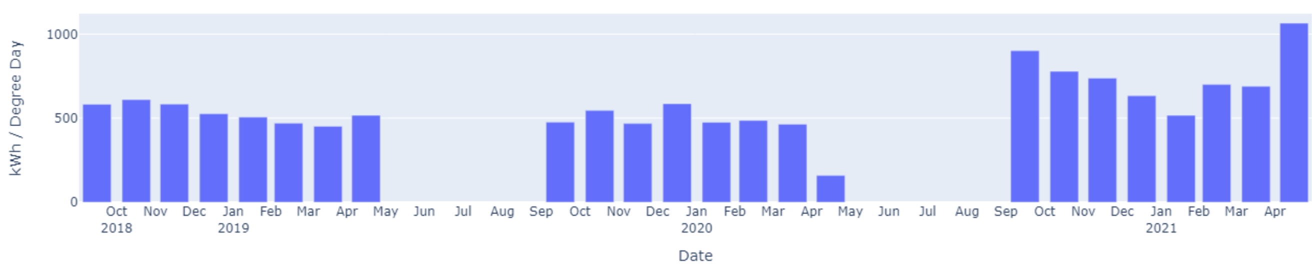
Building 2 Winter Normalized Heating Energy Consumption



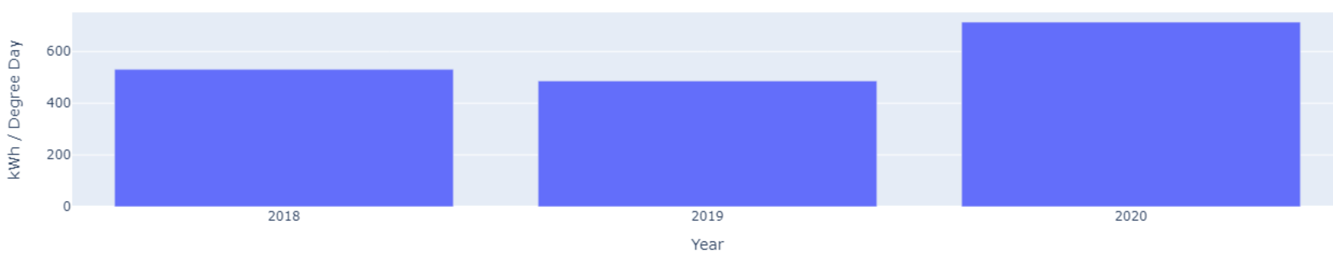
Building 3 Winter Daily Normalized Heating Energy Consumption



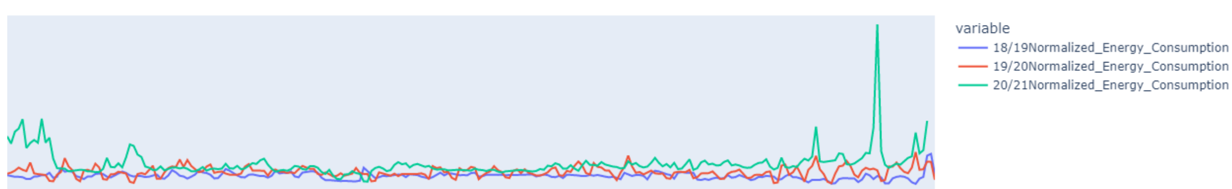
Building 3 October to May Normalized Heating Energy Consumption From 2018 to 2021

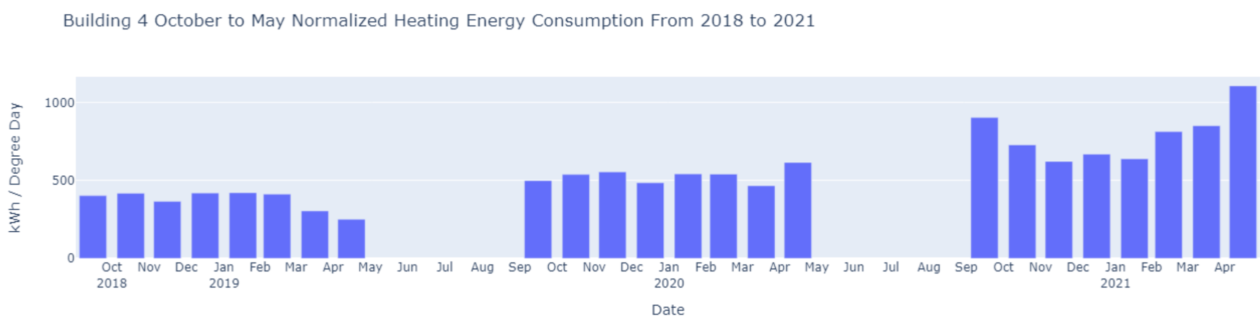


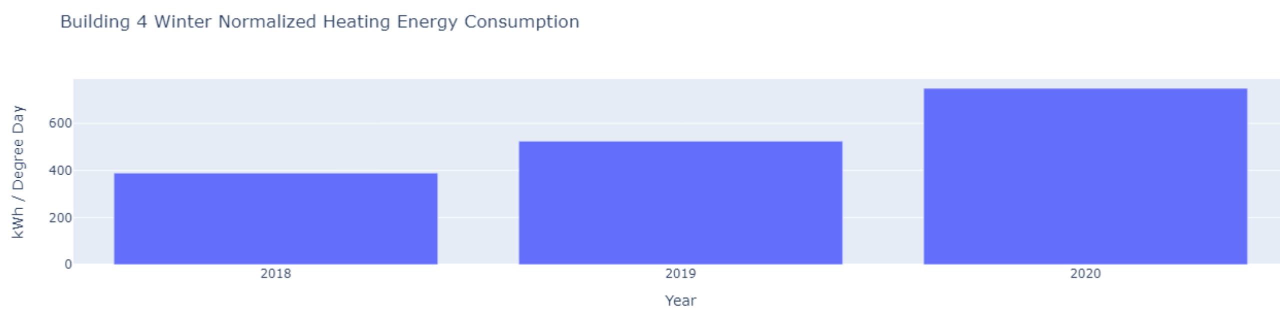
Building 3 Winter Normalized Heating Energy Consumption

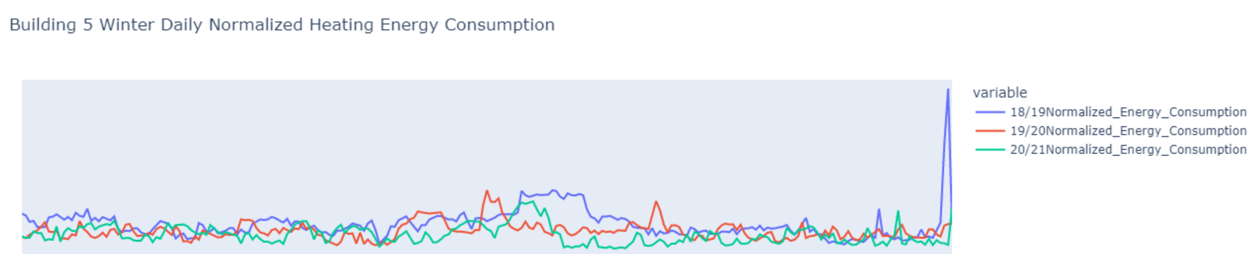


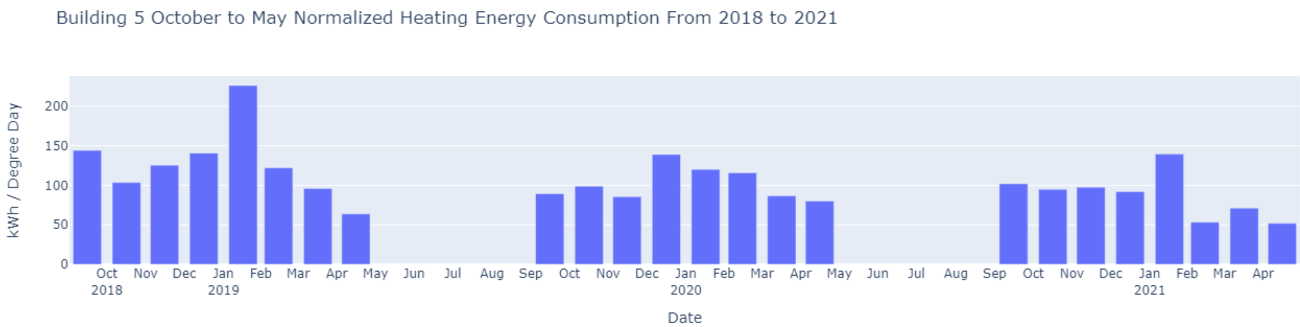
Building 4 Winter Daily Normalized Heating Energy Consumption



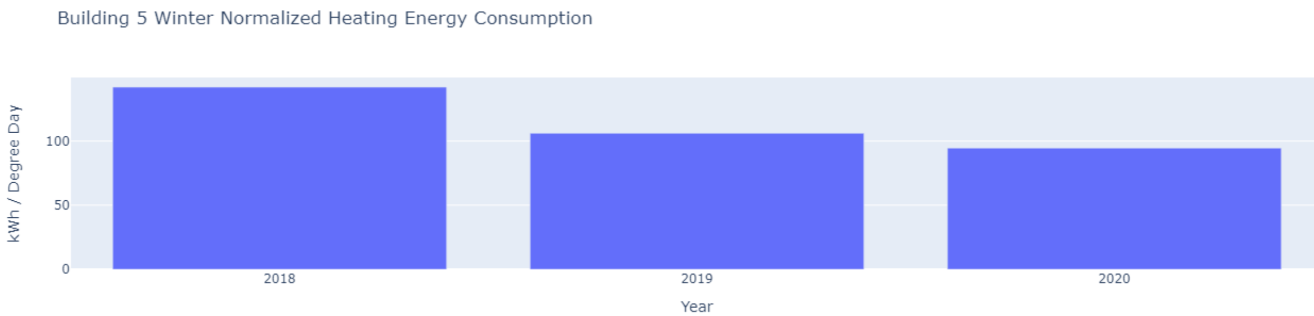

Building 4 October to May Normalized Heating Energy Consumption From 2018 to 2021
kWh / Degree Day
1000
500
0
Oct
2018
Nov
Dec
Jan
2019
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Jan
2020
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Jan
2021
Feb
Mar
Apr
Date


Building 4 Winter Normalized Heating Energy Consumption
kWh / Degree Day
600
400
200
0
2018
2019
2020
Year

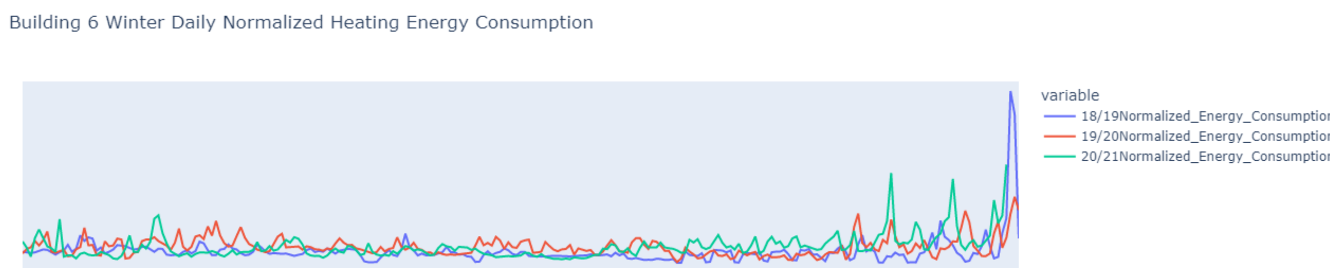

Building 5 Winter Daily Normalized Heating Energy Consumption
variable
18/19Normalized_Energy_Consumption
19/20Normalized_Energy_Consumption
20/21Normalized_Energy_Consumption



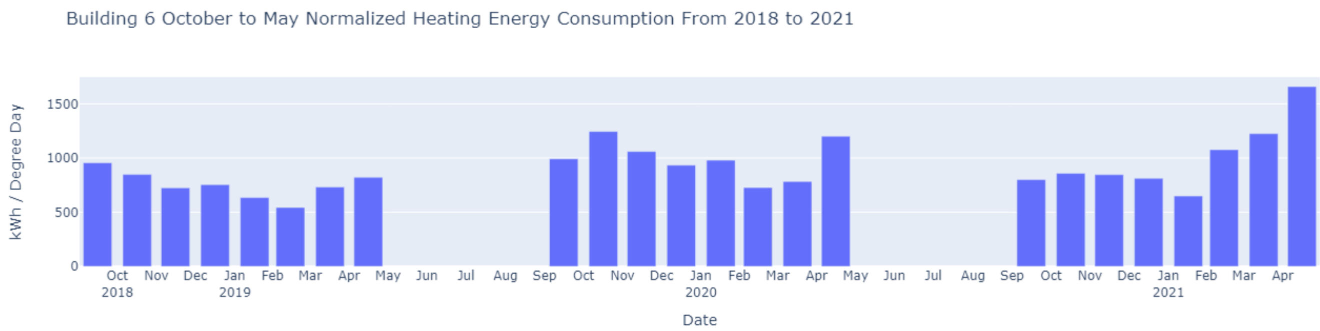
Building 5 October to May Normalized Heating Energy Consumption From 2018 to 2021
kWh / Degree Day
0
50
100
150
200
Oct
2018
Nov
Dec
Jan
2019
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Jan
2020
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Jan
2021
Feb
Mar
Apr
Date

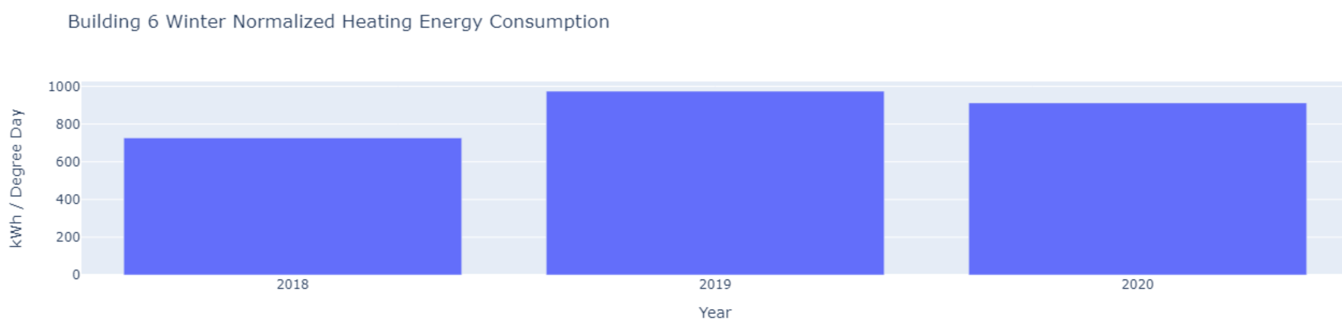


Building 5 Winter Normalized Heating Energy Consumption
kWh / Degree Day
0
50
100
2018
2019
2020
Year



Building 6 Winter Daily Normalized Heating Energy Consumption
variable
18/19Normalized_Energy_Consumption
19/20Normalized_Energy_Consumption
20/21Normalized_Energy_Consumption





Analysis of the plot indicates that, apart from building b5, there has not been an improvement in the efficiency of heating energy consumption in campus buildings from 2018 to 2021.

From 2018 to 2019, the energy consumption efficiency for heating in most campus buildings increased. Particularly, b1 and b2 demonstrate a notable improvement in efficiency.

From 2019 to 2020, most campus buildings exhibited decreased or sustained efficiency in their utilization of heating energy.

The efficiency of heating energy consumption in building 5 is increasing steadily, whereas the efficiency of heating energy consumption in building 4 is declining annually.

Comparing the efficiency of heating energy usage in different buildings

Since each building has a different floor area, we need to calculate the energy use intensity of each building to compare the heating energy use efficiency of different buildings. By dividing the weather normalized heating energy usage of a building by its corresponding total property gross floor area, we can derive the weather normalized heating energy usage and thus compare the energy efficiency between buildings.

By checking the records in Archives & Special Collections at the University of Calgary, we obtained the floor area records of these six buildings. Gross Floor Area (GFA) of Buildings: Building 1 has 24531 m^2. Building 2 has 54733 m^2. Building 3 has 22233.92 m^2. Building 4 has 17704.82 m^2. Building 5 has 11315.59 m^2. Building 6 has 11904 m^2.

Calculate the Weather Normalized Heating EUI

```
# initialize list of lists
data = [['b1', 24531], ['b1', 54733], ['b3', 22233.92], ['b4', 17704.82], ['b5', 11315.59], ['b6', 11904]]
# Create the pandas DataFrame
df_gfa = pd.DataFrame(data, columns=['Building', 'GFA'])
# print dataframe.
df_gfa
#https://asc.ucalgary.ca/building/arts-education/
df_gfa.to_sql('gfa', engine )
```

	Building	GFA
0	b1	24531.00
1	b2	54733.00
2	b3	22233.92
3	b4	17704.82
4	b5	11315.59
5	b6	11904.00

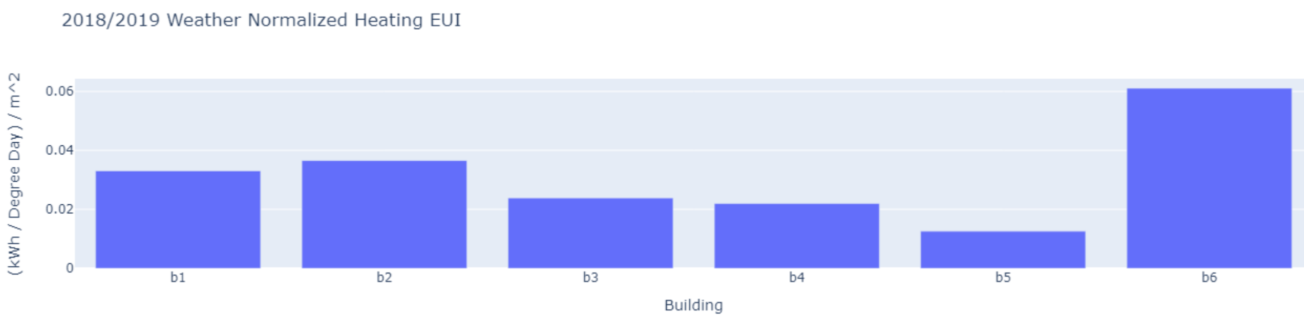
```
create_statement3 = '''
select  t1.Building, t1.Normalized_Energy_Consumption AS '18/19Normalized_Energy_Consumption',
t2.Normalized_Energy_Consumption AS '19/20Normalized_Energy_Consumption',
t3.Normalized_Energy_Consumption AS '20/21Normalized_Energy_Consumption'
from c1819w t1 inner join c1920w t2 on t1.Building = t2.Building
inner join c2021w t3 on t1.Building = t3.Building;
'''
Percentage_Change = pd.read_sql_query(create_statement3, engine)
display(Percentage_Change)
```

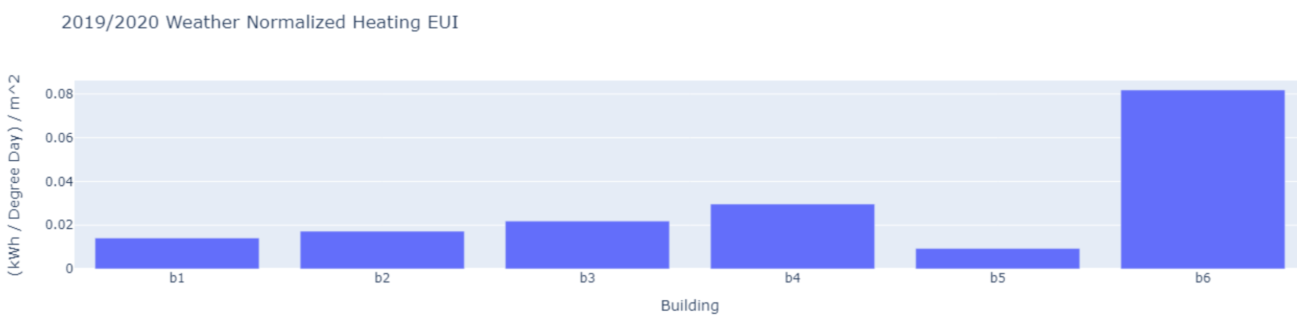
	Building	18/19Normalized_Energy_Consumption	19/20Normalized_Energy_Consumption	20/21Normalized_Energy_Consumption
0	b1	811.349037	345.991001	963.399269
1	b2	2001.753773	942.181107	1210.162728
2	b3	530.932464	486.194448	713.738304
3	b4	389.293502	524.500841	749.543221
4	b5	142.071554	106.064072	94.593856
5	b6	727.217641	974.388501	912.397512

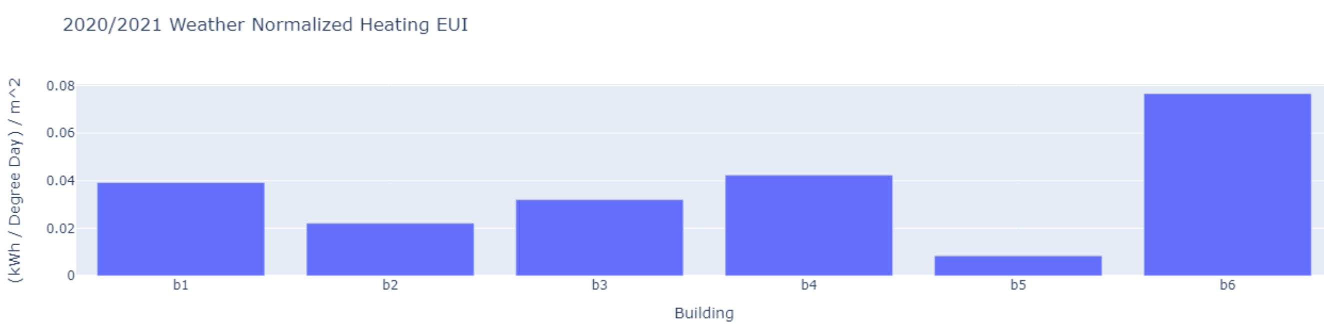
```python
statement = '''
SELECT t1.Building, GFA, (`18/19Normalized_Energy_Consumption`/GFA) AS '18/19 Weather Normalized Heating EUI',
(`19/20Normalized_Energy_Consumption`/GFA) AS '19/20 Weather Normalized Heating EUI',
(`20/21Normalized_Energy_Consumption`/GFA) AS '20/21 Weather Normalized Heating EUI'
from gfa t1
left JOIN (
select  t1.Building as Building, t1.Normalized_Energy_Consumption AS '18/19Normalized_Energy_Consumption',
t2.Normalized_Energy_Consumption AS '19/20Normalized_Energy_Consumption',
t3.Normalized_Energy_Consumption AS '20/21Normalized_Energy_Consumption',
ROUND((t2.Normalized_Energy_Consumption- t1.Normalized_Energy_Consumption)/t1.Normalized_Energy_Consumption * 100,2) as 'Percentage Change 18 to 19',
ROUND((t3.Normalized_Energy_Consumption - t2.Normalized_Energy_Consumption)/t2.Normalized_Energy_Consumption*100,2) as 'Percentage Change 19 to 20'
from c1819w t1 inner join c1920w t2 on t1.Building = t2.Building
inner join c2021w t3 on t1.Building = t3.Building) AS t2 ON t1.Building = t2.Building
;'''

df_wneui = pd.read_sql_query(statement, engine)
display(df_wneui)
```

	Building	GFA	18/19 Weather Normalized Heating EUI	19/20 Weather Normalized Heating EUI	20/21 Weather Normalized Heating EUI
0	b1	24531.00	0.033074	0.014104	0.039273
1	b2	54733.00	0.036573	0.017214	0.022110
2	b3	22233.92	0.023879	0.021867	0.032101
3	b4	17704.82	0.021988	0.029625	0.042336
4	b5	11315.59	0.012555	0.009373	0.008360
5	b6	11904.00	0.061090	0.081854	0.076646

Comparative analysis of the heating energy usage efficiency of campus buildings







Based on the plots, the efficiency of heating energy use in the buildings varies over time. In the 2018/2019 heating season, building 6 had the least efficient heating energy usage, followed by buildings 2 and 1. Buildings 3, 4, and 5 had the most efficient heating energy usage.

In the 2019/2020 heating season, the rankings changed slightly. Building 6 still had the least efficient heating energy usage, but buildings 4 and 3 had the highest usage. Buildings 2 and 1 had relatively efficient heating energy usage, and building 5 continued to have the most efficient heating energy usage.

In the 2020/2021 heating season, the rankings changed again. Building 6 still had the least efficient heating energy usage but building 4 had the subsequent highest usage. Buildings 1 and 3 had relatively efficient heating energy usage, and buildings 2 and 5 had the most efficient energy.

Overall, building 6 consistently has the least efficient heating energy usage, while building 5 consistently has the most efficient heating energy usage. It is also worth noting that the rankings of the other buildings changed over time, indicating that the efficiency of heating energy use in these buildings may be influenced by factors other than weather normalization.

Discussion

All of us have individually contributed to this project and have had different learning outcomes.

Zheyu Song: In this project, I gained a deeper understanding of using a weather normalization approach to analyze energy efficiency data. This involved using SQL to calculate Heating Degree Days from the daily average temperature data, which allowed me to normalize the heating energy data based on weather conditions. This approach allowed me to understand better the relationship between weather conditions and heating energy usage and provided me with valuable insights that I can use to analyze the efficiency of heating systems. In addition to learning about weather normalization, I also gained experience using SQL to work with large datasets and extract relevant information for my analysis. This skill will be helpful in many different contexts, as SQL is a widely used and powerful tool for working with data. Furthermore, I learned how to properly look up information and data and write references for it. Overall, this project has provided me with a range of valuable skills and knowledge that will be useful in my future endeavors.

In the future, I would like to extend this project by obtaining more years of data from the university. This will allow me to conduct a longer-term analysis of the campus's energy efficiency and identify trends and patterns over time. Additionally, I plan to use the weather normalization methods I have learned to analyze the energy efficiency of other types of buildings, such as public and government buildings. This will provide valuable insights into the energy efficiency of these buildings and help me better understand how they compare to the campus buildings. Overall, I plan to continue working on this project and use the data and analysis techniques I have learned to improve our understanding of energy efficiency and its environmental impact.

Khushi: My dataset was Building Energy Benchmarking, and I worked extensively on it. I cleaned the dataset and concluded substantial information by running queries on it. I have described my work on this dataset in my Individual Milestone and the Report. After this, I joined my dataset with the Weather dataset to find more energy related information using queries.

The most important thing I learnt here was the cumulation of Python and SQL. There are instances in my milestone where I found it easier to do something (like join two data frames) using Pandas rather than SQL. It helped me explore and thoroughly understand both these methods to make informed decisions about my choice. Furthermore, this project made me comfortable working with SQL and python simultaneously and made the assignments an easy task. Overall, this was a great learning experience, and I had a wonderful time working with my teammates.

Jannatul: I worked with dataset 2 (corporate energy consumption), this dataset was cleaner, so I moved forward to the next steps, I converted the month notation in month column for the further SQL queries to join with dataset 4. I explored the dataset with SQL queries to observe the average, total, minimum, maximum natural gas and electricity consumption. I have also gained insight on joining tables in SQL and then ran some more queries. I concluded that the corporate buildings of Calgary consume more energy from natural gas than electricity. I have also calculated normalized energy for three different years from 2019 to 2021 to observe the impact of Covid and compare that with pre-covid and post-covid years. But I inferred from the result that the normalized energy consumption in average is increasing over the years for corporate buildings which is undesirable, as increasing normalized energy is an indication of lower energy efficiency. In future I would like to collaborate on this project with air quality data of Calgary and observe the impact of using Natural gas on air pollution. It was a pleasant experience for me to work on this project and get a chance to apply the SQL knowledge along with learning in the course. Moreover, all my team members were cooperative, that made this group project even more meaningful and interesting.

Shashank: I worked on the university energy consumption data. It was interesting to look at the university's energy consumption. Wrangling the data with SQL was something new to me. Though I was not able to use many complex queries in my part of the research, I tried to implement what I learned from various sources. I struggled with it at first but was soon able to get the hang of it. Talking about the data, if we look very closely the data responded to weekends as well. The total energy dipped on weekends. Working with institutional data helped me get an insight into how a university of this size consumes energy. Learning why one building was consuming more energy than the other, why one type of building was consuming a specific kind of energy the most. Studying the impact of COVID was the most interesting part. We saw how the energy consumption dipped and even the variation of energy reduced.

Conclusion

Our analysis of the heating energy usage efficiency of six buildings on the University of Calgary campus showed that none of the buildings, except for building five, exhibited an increasing trend in efficiency. From 2019 to 2020, we observed that most buildings showed decreased or sustained efficiency in their heating energy use. Additionally, we found that building six consistently had the least efficient heating energy usage, while building five consistently had the most efficient heating energy usage. However, it should be noted that we were only able to analyze trends over three years due to data limitations. This prevents us from accurately analyzing the heating energy trends of campus buildings over a more extended time period. Despite this limitation, our findings provide valuable insights into the current state of heating energy use efficiency on campus and can inform future efforts to improve sustainability and reduce carbon emissions.

We can say that we concluded the following:

- Overall energy use has remained the same throughout the three years. There was a dip in the energy use in the corporate buildings and Campus buildings during 2020, and we can assume this was due to the COVID-19 pandemic.
- Normalized energy use has also remained the same for the three years with a dip in 2020, due to the pandemic.
- We can safely say that there is a direct correlation between Temperature and Energy consumption. There may be a dependency on other factors but no direct correlation.
- There was a significant drop in the Energy use in Campus buildings from the year 2019 to 2020, and then an increase from 2020 to 2021. This is again due to the pandemic situation. All buildings were affected by this except the Office/Admin buildings which we can assume were open partially.
- Most buildings in the City of Calgary are majorly dependent on Natural Gas as their main source of energy, followed by electricity.

References

Canada Energy Regulator-Energy in Canada [Online]. Available at: https://www.cer-rec.gc.ca/en/about/publications-reports/annual-report/2018/energy-in-canada.html#:~:text=Canada%20is%20currently%20ranked%20the,and%20future%20needs%20of%20Canadians (Accessed November 6, 2022)

Sustainable Building Partnership Program, City of Calgary [Online]. Available at: https://www.calgary.ca/development/sustainable-building-partnership.html (Accessed November 6, 2022)

Canada Energy Regulator - Provincial and Territorial Energy Profiles – Alberta [Online]. Available at: https://www.cer-rec.gc.ca/en/data-analysis/energy-markets/provincial-territorial-energy-profiles/provincial-territorial-energy-profiles-alberta.html (Accessed November 6, 2022)

Corporate energy consumption - City of Calgary [Online]. Available at: https://data.calgary.ca/Environment/Corporate-Energy-Consumption/crbp-innf (Accessed November 1, 2022)

Building Energy Benchmarking - City of Calgary [Online]. Available at: https://data.calgary.ca/Environment/Building-Energy-Benchmarking-City-of-Calgary/8twd-upbv (Accessed November 1, 2022)

University of Calgary data - The data for this project was retrieved through the Office of Sustainability Campus as a Learning Lab initiative *

Alberta Agriculture, Forestry and Rural Economic Development, Alberta Climate Information Service. (2022) Current and Historical Alberta Weather Station Data [Online]. Available at: https://acis.alberta.ca/acis/weather-data-viewer.jsp (Accessed November 5, 2022)

University of Calgary News, UCalgary climbs to top 3% of Times Higher Education global sustainability ranking [Online]. Available at: https://ucalgary.ca/news/ucalgary-climbs-top-3-times-higher-education-global-sustainability-ranking (Accessed November 5, 2022)

University of Calgary Utilites, Thermal Comfort [Online]. Available at: https://www.ucalgary.ca/facilities/thermal-comfort (Accessed November 5, 2022)

2022 Sustainability Report - University of Calgary [Online]. Available at: https://www.ucalgary.ca/sustainability/strategy/2022-sustainability-report/2022-sustainability-report-climate-change-and (Accessed November 1, 2022)

Degree Days - Energy Lens [Online]. Available at: https://www.energylens.com/articles/degree-days (Accessed November 1, 2022)

Evans J, SQL queries don't start with SELECT [Online]. Available At: https://jvns.ca/blog/2019/10/03/sql-queries-don-t-start-with-select/ (Accessed November 4, 2022)

How to join tables using SQL to combine datasets [Online]. Available At: https://towardsdatascience.com/combining-tables-using-sql-f6ddfd5f8b6f (Accessed November 5, 2022)

SQL: Combining the AND and OR Conditions, SQL: Combining the and/ or conditions. Available at: https://www.techonthenet.com/sql/and_or.php (Accessed: November 25, 2022).

This article is written by plus2net.com team.https://www.plus2net.com (2000) Sum of values of columns in mysql table with group by, in and case, Sum sql for data in multiple columns and across rows with Total & Percentage. plus2net.com. Available at: https://www.plus2net.com/sql_tutorial/sql_sum-multiple.php (Accessed: November 25, 2022).

Heating Degree Day (HDD): Definition and How To Calculate. Investopedia, [Online]. Available At: https://www.investopedia.com/terms/h/heatingdegreeday.asp (Accessed December 11, 2022)

Heating and Cooling Degree Days: What Are They and Why Do They Matter? [Online]. Available At: https://watchwire.ai/heating-cooling-degree-days-explained/ (Accessed December 12, 2022)

University of Calgary Archives & Special Collections, Built History of the University of Calgary [Online]. Available at: https://asc.ucalgary.ca/buildings/ (Accessed December 12, 2022)

How to Calculate Weather-Normalized Energy Consumption [Online]. Available at: https://medium.com/builtrix/how-to-calculate-weather-normalized-energy-consumption-55acacad50ea (Accessed December 13, 2022)