Part-5 5.1 Summary Report • Write a structured report (300-500 words) covering: o Key trends in energy consumption and efficiency. o Seasonal and property type variations. o Recommendations for improving energy efficiency and reducing emissions. • Include supporting visualizations with clear titles, labels, and legends. • Submit the GitHub repository link in the report on D2L along with the Jupyter Notebook. • Highlight in the report where Regex was used for data cleaning and extraction.

==> Key trends in energy consumption and efficiency. Greenhouse Gas (GHG) Emissions: Properties that consume greater amounts of energy tend to emit GHG at a higher rate. A particularly strong correlation was found between the property size (GFA) and total emissions. Also, Residential buildings tend to use the more enrgy along with the commercial buildings as evident in the below barchart.

Publoc spaces like calgary public and fitness centres have considerably higher usage.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
import re
data=pd.read_csv("Building_Energy_Benchmarking.csv")
df=pd.DataFrame(data)


# Convert 'Total GHG Emissions (Metric Tons CO2e)' column to numeric
df['Total GHG Emissions (Metric Tons CO2e)'] = pd.to_numeric(df['Total GHG Emissions (Metric Tons CO2e)
plt.figure(figsize=(12,6))
x_labels = df['Property Name'].head(10)
y_values = df['Total GHG Emissions (Metric Tons CO2e)'].head(10)

plt.bar(x_labels, y_values)
plt.xticks(rotation=45, ha='right')
plt.title('Top 10 Buildings with Highest GHG Emissions')
plt.xlabel('Property Name')
plt.ylabel('Total GHG Emissions (Metric Tons CO2e)')

# Annotate bars with emission values using plt.text()
for i, value in enumerate(y_values):
    plt.text(i, value + (value * 0.01), f'{value:.1f}', ha='center', fontsize=10, color='black')
plt.show()
```

Also, there has been a gradual increase in the usage after 2021, and it keeps increasing yearly.

```
# Converting 'Year Ending' to numeric
df['Year Ending'] = pd.to_numeric(df['Year Ending'], errors='coerce')

# Converting 'Site Energy Use (GJ)' to numeric, handling errors
df['Site Energy Use (GJ)'] = pd.to_numeric(df['Site Energy Use (GJ)'], errors='coerce')

# Grouping by 'Year Ending' and calculate the mean of 'Site Energy Use (GJ)'
Trends = df.groupby('Year Ending')['Site Energy Use (GJ)'].mean()

# Plotting
plt.figure(figsize=(10, 5))
sns.lineplot(x=Trends.index.astype(int), y=Trends.values, marker='o', linestyle='-', color='red')
```

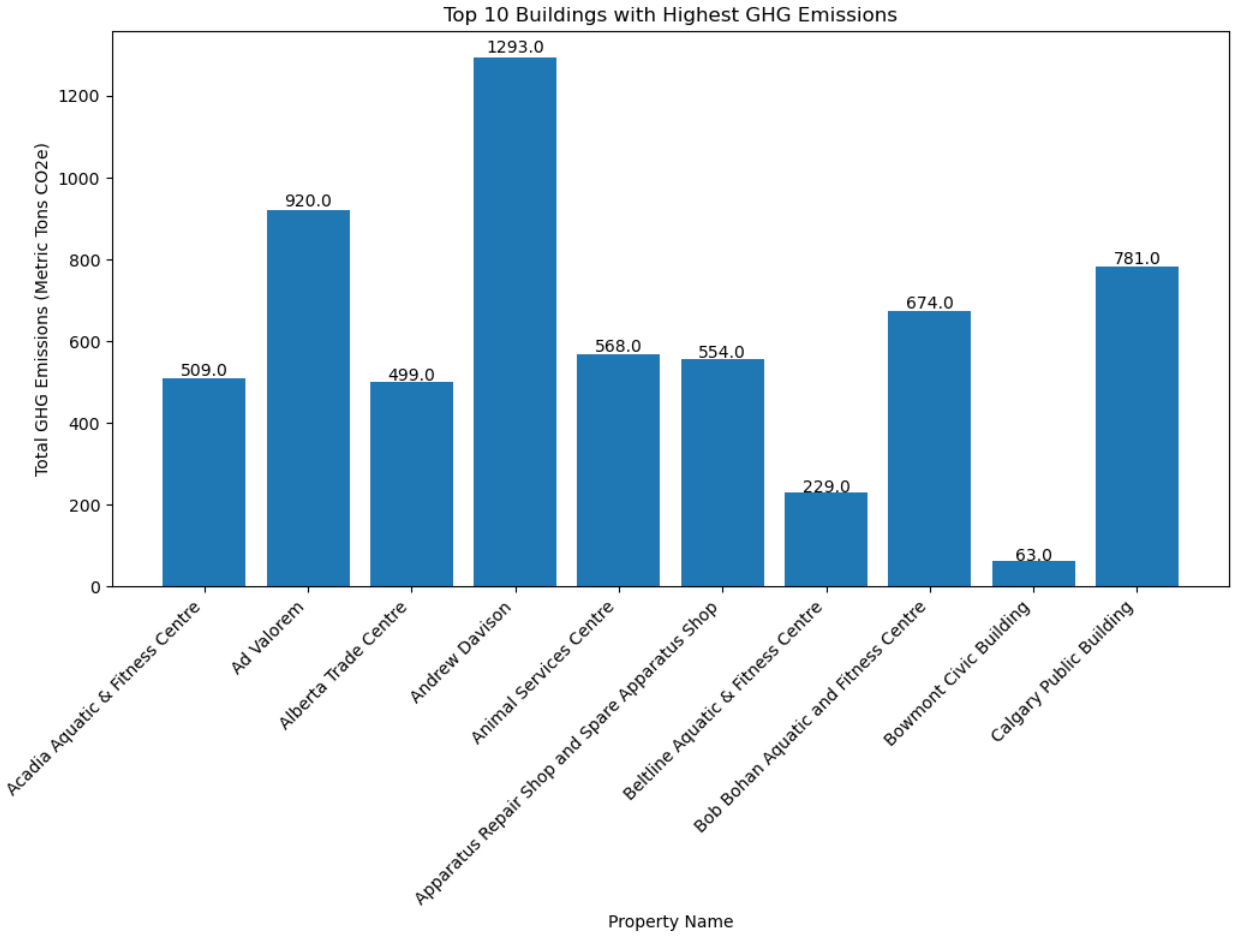
Top 10 Buildings with Highest GHG Emissions
Total GHG Emissions (Metric Tons CO2e)
0
200
400
600
800
1000
1200
509.0
920.0
499.0
1293.0
568.0
554.0
229.0
674.0
63.0
781.0
Acadia Aquatic & Fitness Centre
Ad Valorem
Alberta Trade Centre
Andrew Davison
Animal Services Centre
Apparatus Repair Shop and Spare Apparatus Shop
Beltline Aquatic & Fitness Centre
Bob Bohan Aquatic and Fitness Centre
Bowmont Civic Building
Calgary Public Building
Property Name

Figure 1: png

```
plt.title("Yearly Trend of Average Site Energy Use Intensity (EUI)", fontsize=14)
plt.xlabel("Year", fontsize=14)
plt.ylabel("Site Energy Use (GJ)", fontsize=10)
plt.grid(True)
plt.xticks(Trends.index.astype(int))

# Annotating data points
for year in Trends.index:
    plt.text(int(year), Trends[year], f"{Trends[year]:.2f}", fontsize=12, ha='right', color='black')

plt.show()
```

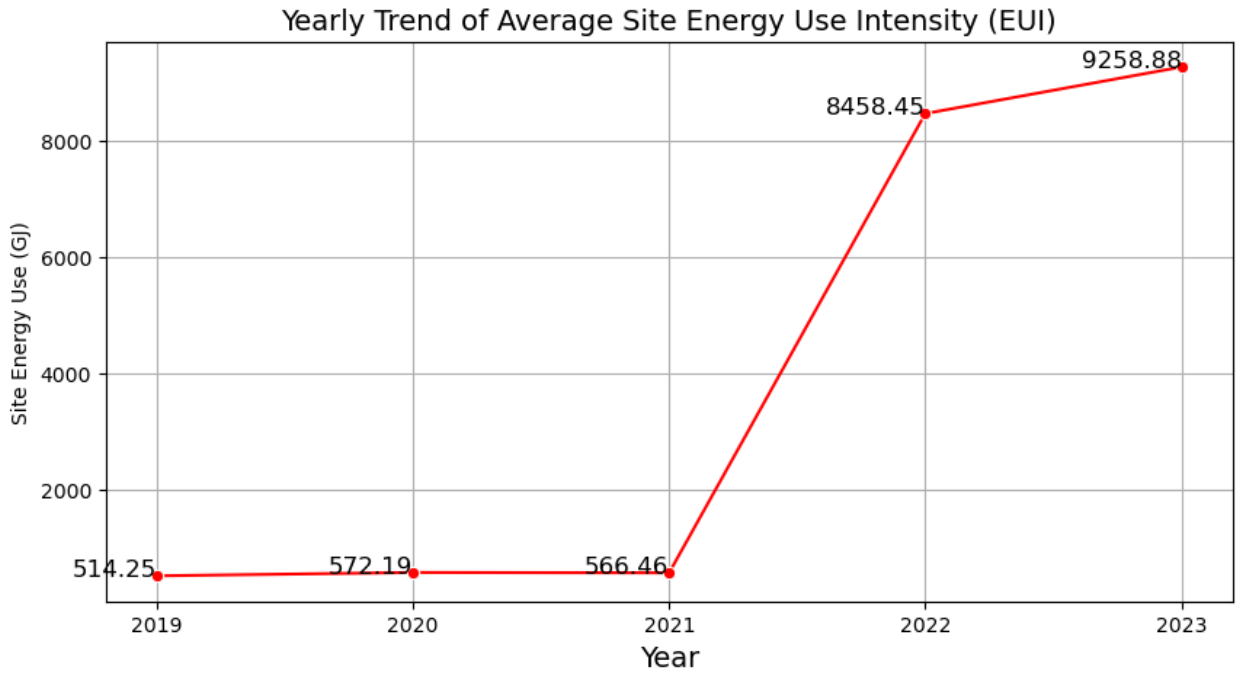

Figure 2: png

==> Seasonal and property type variations. We plotted the box plots based on the property type and found that the industrial and public usage buildings usese more energy the others. Additioanlly there are many outliers that indicates the inconsistency in the pattern of usage.

We should also consider the usage season wise as in winters, heating systems such as boilers and furnaces will require more gas usage, which in turn increasing the electricity consumption. Seasonal fluctuation will be less in the commercial buildings such as hospitals and offices however property such as School will use less energy in summer.

If we include the month wise data in the given data, we might shed more light on the seasonal usage of energy

```
plt.figure(figsize=(12, 8))
sns.boxplot(data=df, x='Primary Property Type - Self Selected', y='Site EUI (GJ/m²)')
plt.xticks(rotation=45, ha='right')
plt.title("Energy Use Intensity by Property Type")
plt.xlabel("Property Type")
```

```
plt.ylabel("Site EUI (GJ/m²)")
plt.show()
```

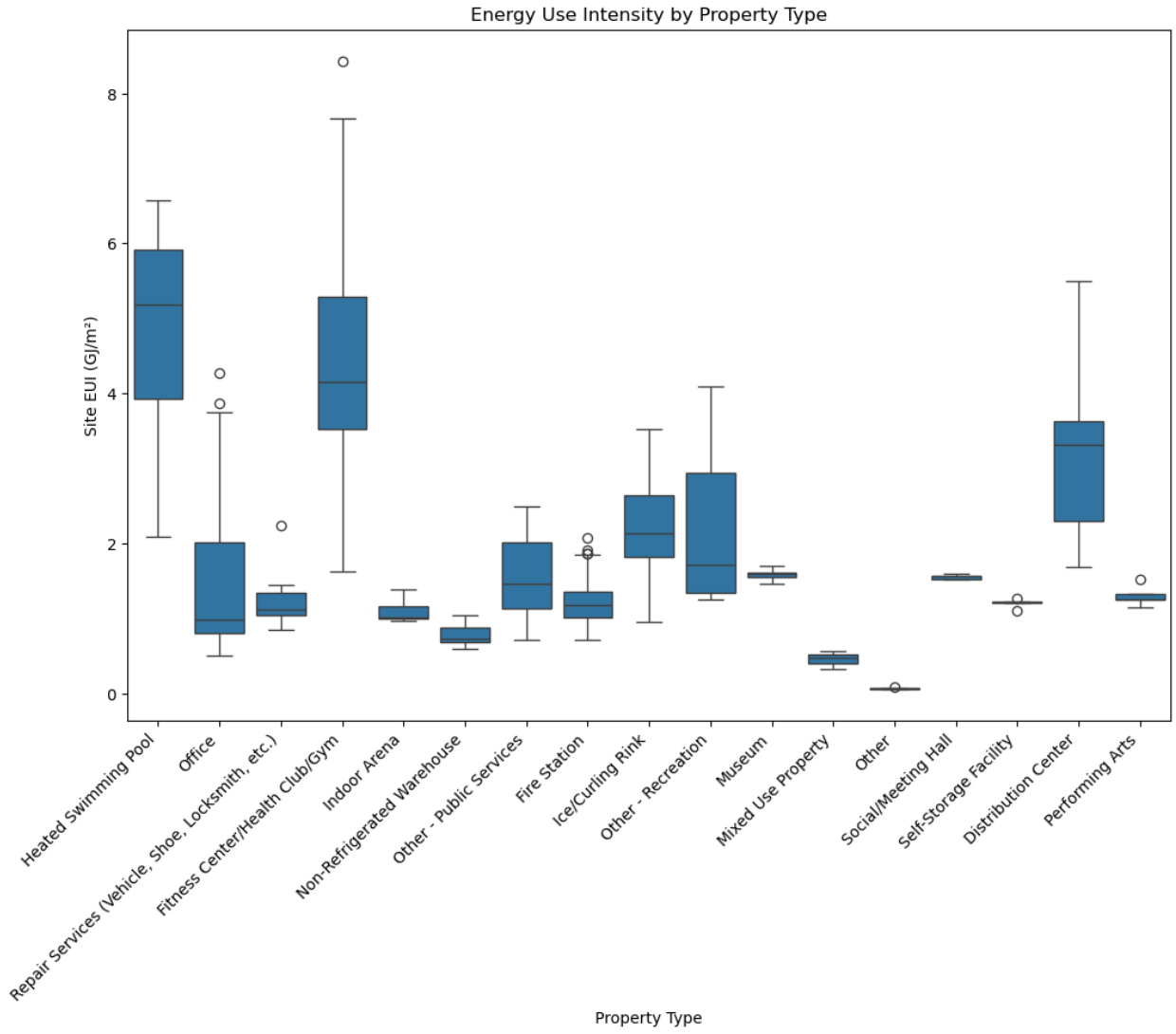

Figure 3: png

==> Recommendations for improving energy efficiency and reducing emissions

1. Monitoring performance and settig goals - As we have the past data we can set the usage goals for future and track for future. We can come up with the real time tracking system which will allow us to monitor day wise usage and help in taking measure based on the high usage areas and requirements. It is extremely important to regularly monitor this data and report to highlight the areas which need more attention
2. Should increase the usage if renewable enrgy sources - Installing Solar panels and geao thermal heating and cooling systems could be th initial step to move further towards more sustainable environment. We should also think about installing more energy efficient equipments.
3. Optimising Water and Electricity usage - Specifically in the residential buildings ,we should have a real time sytem to track these usage. Installing low-flow water fixtures to reduce water heating demand. Also, raiun water harvesting is a good way of reducing the water consumption.

=== Regex Use in the data

I have used regex in the data cleaning as mentioned in the instructions. For extracting numric values from the text based column, to remove spaces and commas from the data using re.sub. Created a numerical pattern to match and extract from the clean data using re.search.

For standardarising the postal codes based on the canadian posta code format i have used re.match to match with the format. Also used re.sub again to remove any special characters and signs from the data.

Analyze the relationship between building age and energy efficiency.

To analyse the data we plotted the scatter plot,using the correlation matrix which suggest that the building age has a weaker effect on the energy consumption. As the other factors , such the maintainance of the buliding , equipment used, renovations, upgrades should be considered for braoder influence.

```
import datetime
#defining teh column and converting in the datetime format
if 'Year Built' in df.columns:
    year = datetime.datetime.now().year
    df['Building Age'] = year - df['Year Built']

    # Removing invalid values
    df = df[df['Building Age'] > 0]

    # Computing correlation between Building Age and Site EUI
    corr = df[['Building Age', 'Site EUI (GJ/m²)']].corr()
    print("Correlation Matrix:\n", corr)

    # Scatter plot: Building Age vs. Site EUI
    plt.figure(figsize=(10, 6))
    sns.scatterplot(x=df['Building Age'], y=df['Site EUI (GJ/m²)'], alpha=0.5)
    plt.title("Building Age vs. Energy Efficiency (Site EUI)")
    plt.xlabel("Building Age Group")
    plt.ylabel("Site EUI (GJ/m²)")
    plt.show()

else:
    print("Year Built column is missing in the dataset.")
```

```
Correlation Matrix:
                    Building Age  Site EUI (GJ/m²)
Building Age            1.000000          0.170513
Site EUI (GJ/m²)        0.170513          1.000000
```

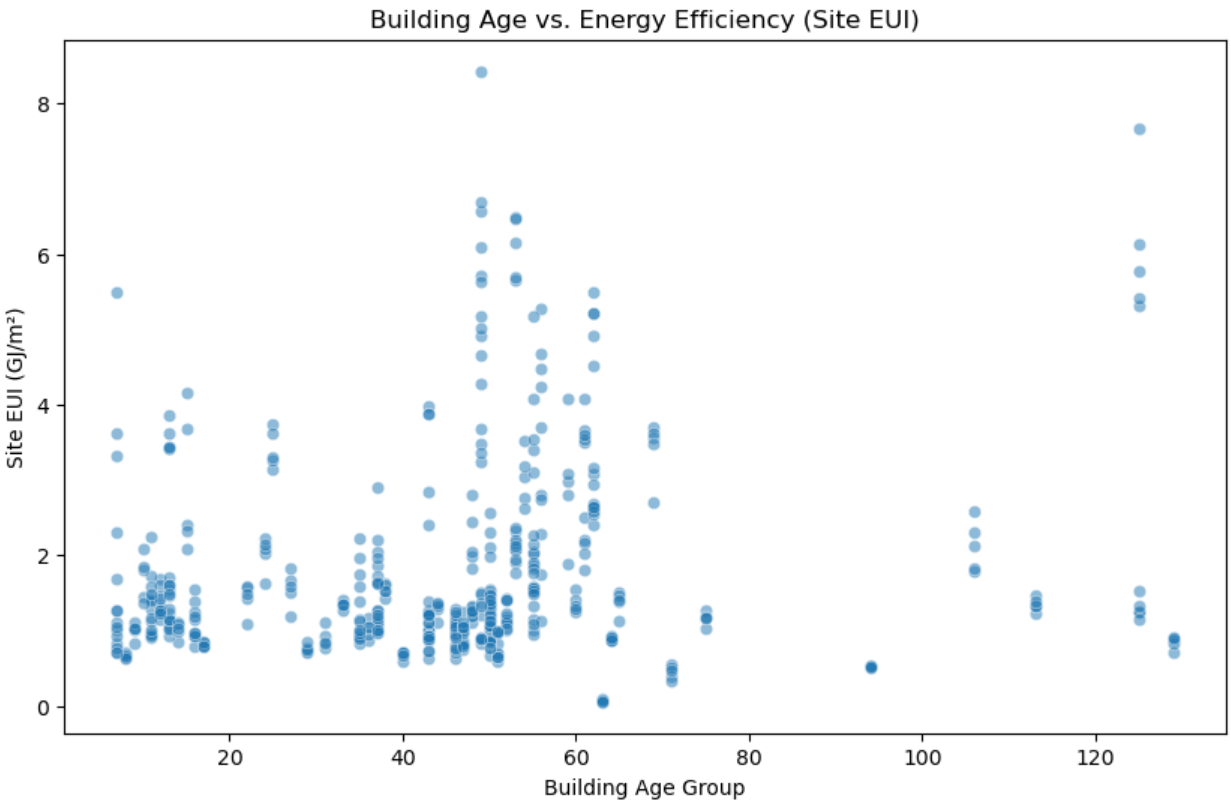
Building Age vs. Energy Efficiency (Site EUI)
Site EUI (GJ/m²)
Building Age Group
0
2
4
6
8
20
40
60
80
100
120

Figure 4: png