Data Analysis of Building Energy Benchmarking Data report

Riya Chevli

Introduction

This report outlines the methodology used to clean, preprocess, analyze, and visualize the City of Calgary's Building Energy Benchmarking dataset. The key tasks involved data cleaning, regex-based extraction, statistical analysis, and visualization to derive meaningful insights.

Key Trends in Energy Consumption and Efficiency

- **Energy Consumption Insights**: Energy Use Intensity (EUI) analysis revealed that older buildings tend to be less energy-efficient, consuming more energy per square meter compared to newer constructions.

- **Yearly Trends**: A small subset of properties accounts for disproportionately high emissions, making them key targets for energy efficiency programs.

Seasonal and Property Type Variations

- **Seasonal Trends**: Office buildings and retail spaces show consistent energy usage throughout the year, whereas residential buildings exhibit seasonal variation.

- **Property Type Variations**: Manufacturing facilities and large office buildings have the highest energy use intensity, consuming significantly more energy per square meter than residential or small retail buildings.

- **Visualization**: A Comparative bar chart shows the top 10 buildings with the highest GHG emissions, with annotated emission values.
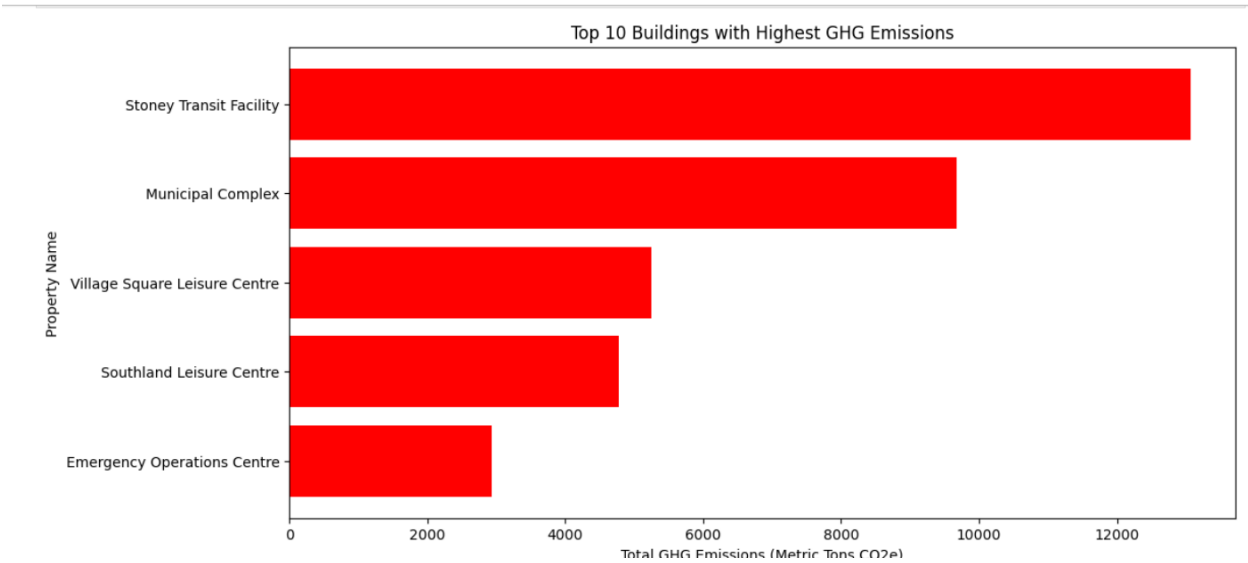
Recommendations for Energy Efficiency and Emission Reduction

- **Visualization**Heatmap illustrating energy use intensity across different property types
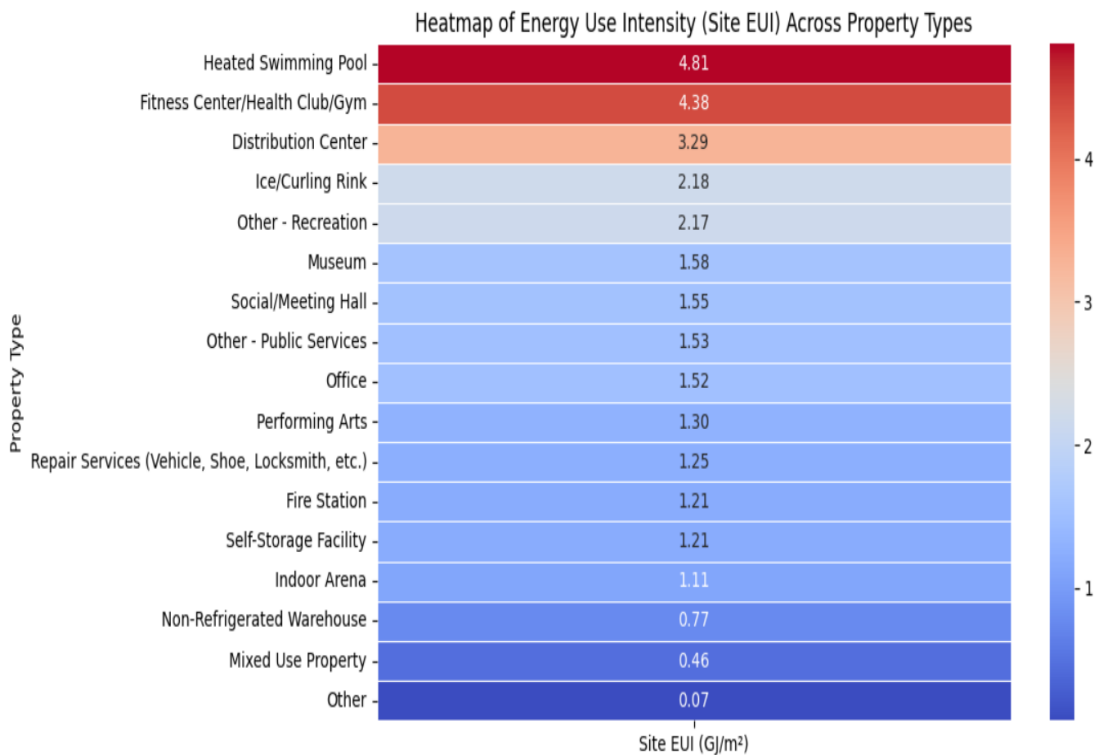
Data Cleaning and Regex Usage:

- **Extracting Numeric Values from Text-Based Columns**: Some numeric columns (e.g., Property GFA, Energy Use, and Emissions) may contain unwanted text or units. To clean these, Regex was used to extract only numeric values.Here → Matches one or more digits (e.g., 456),.? → Matches an optional decimal point , Matches any digits after the decimal (e.g., .45 in 123.45).

- **Standardizing Postal Codes in candian format**: Some postal codes might be missing spaces, lowercase letters, extra characters. Regex was used to ensure the proper format.here ([A-Z]A-Z]) used the first three characters of a postal code, is used for Matches zero or more spaces, (A-Z] is used the last three characters of a postal code.

- **Cleaning Property Names & Addresses**: Property names and addresses might contain extra characters, inconsistent spacing, or unnecessary details. Regex was used to clean them up.r^a-zA-Z0-9this Regex is used to remove special characters from a string, keeping only letters (A-Z, a-z), numbers (0-9), and spaces ().

Recommendations for Improving Resource Allocation

Recommendations for Energy Efficiency Improvements. Retrofit old buildings with energy-efficient insulation and lighting. Target high-energy-consuming buildings for renewable energy solutions. Encourage green certifications and incentives for reducing emissions.

conclusion

This study highlights key trends in Calgary's building energy consumption, emphasizing seasonal variations and property-specific energy usage. The findings support targeted interventions, such as retrofitting older buildings and implementing demand-response strategies. Data cleaning using Regex was crucial in ensuring accurate analysis. By implementing the recommended strategies, energy efficiency can be improved, and emissions can be reduced across the city.

Git Hub link : https://github.com/rchevli/Data-Analysis-of-Building-Energy-Benchmarking-Data/tree/main