Energy Usage and Building Efficiency

CSC 47400 - Visualization - Project Report

David Zhang and Hamim Choudhury

Abstract:

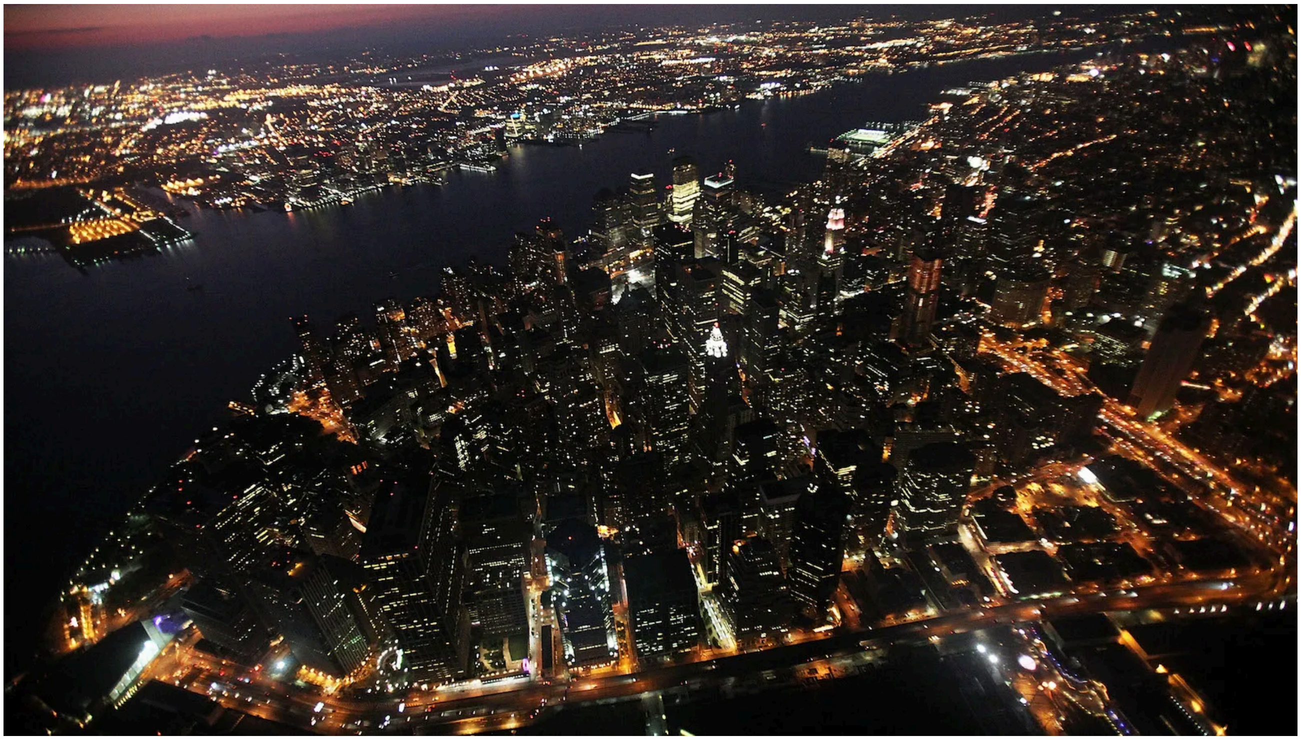
The goal of this project is to analyze and visualize building-level energy consumption trends across time, weather, occupancy, and building type. This is an important project because there are millions of buildings throughout New York City, coming in at different shapes and sizes. And in order for the building to run properly, it must use some form of energy. Energy consumption can be altered by weather, the size of the building, what the building is used for, and much more. It is important for people to know when they want to buy a commercial building or even a home in a certain location, what the energy consumption looks like in that specific area. You would want to pay less for your electricity bill, not more. It is important to make the right decision when choosing a building for both the buyer and the tenant.

We will be using two different datasets for this project, one from Kaggle and another from the NYC Open Data API. These datasets consist of thousands of buildings all throughout New York City, showcasing important attributes. Some of the most notable attributes are: Property GFA - Calculated (Buildings) (ft^2), Site Energy Use, Electricity Use, and ENERGY Star Score. There are more columns that we will be using to come to our conclusion. Our goal is to develop an interactive dashboard that enables users to explore these multivariate relationships through histogram plots, bar plots, line plots, and scatter plots.

Introduction:

Based on population, New York City is the largest city in the United States. It's also the most diverse with people from different backgrounds emigrating here and starting a life. With all of these people, they need to live somewhere. Renting an apartment or buying a house is the way to go. Other people live in shelters or hostels. Regardless, the majority of people live in a place that uses electricity and uses energy. These buildings can be of varying sizes. And the bigger the building, the more energy it uses.There is a difference between electricity and energy usage in the sense of this project. We will be focusing on energy usage, which refers to the total consumption of all forms of energy in a building. This includes electricity, gas, heating oil, steam, etc. Electricity usage just refers to a component of energy usage (like lighting and charging a device).

Energy usage varies, most obviously is the size of the building but there are also other factors such as weather, time, the amount of people in the building, emissions of carbon dioxide, natural gas usage, the type of building, and even the location. All of these factors have to come into consideration when calculating the energy usage for a building. This project is useful in the fact that there is an interactive dashboard where users will be able to explore energy usage throughout New York City. Users will be able to apply filters such as Borough, Property Type, size of the building, Energy Usage, and the ENERGY Star Score. On top of that, users will have access to different plots such as histogram, bar, and scatter to identify patterns, see where more energy is being used, and evaluate what really increases the energy usage and how different factors impact it. This project emphasizes energy management and making the right decision.



An overhead view of lower Manhattan with parts of New Jersey showing too

Background/Novelty:

This project that we chose is not something brand new, governmental agencies, universities, and energy companies all across the world focus on this. The United States has the Department of Energy, which manages the nuclear infrastructure and administers the country's energy policy. From their website, "People spend 90% of their time in buildings—in homes, offices, schools, hospitals, restaurants, stores, and elsewhere. Buildings provide shelter and safety. Buildings also use 74% of electricity in the United States and account for $370 billion in annual energy costs. Improving the energy efficiency of buildings is critical to lowering energy costs, strengthening resilience to extreme weather events, improving grid reliability, and making residential and commercial buildings more comfortable and healthier for all Americans." This is backed by research that they have done themselves by using statistical methods, surveys, and scientific research.

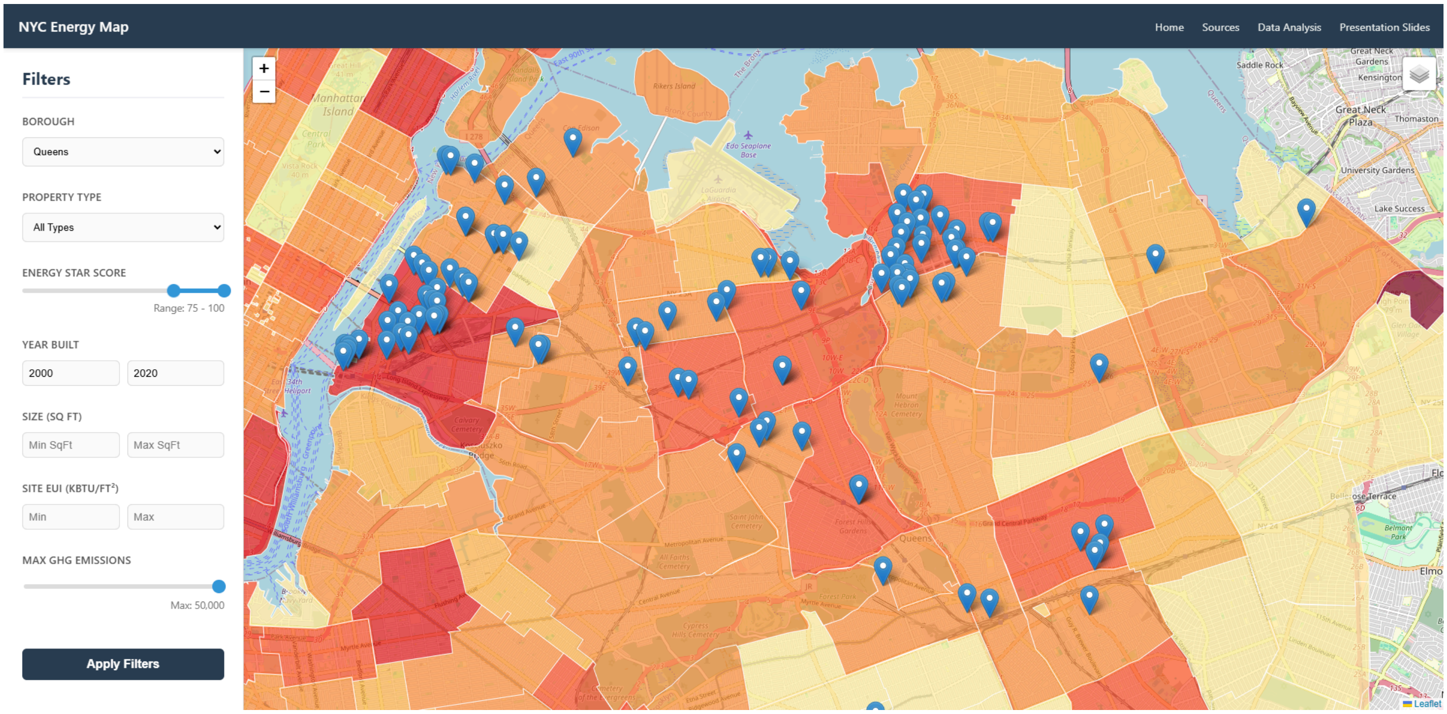
New York buildings are responsible for ⅓ of statewide greenhouse gas emissions. This is not good and the goal is to reduce emissions to make the air we breathe in better and have clean energy. Analysis of urban building energy data revealed correlations between energy usage and building characteristics such as type, size, and age, as well as external factors like weather and occupancy patterns. This is a major reason why New York is redesigning the way buildings are made and the type of energy being used. New York officials are hoping for better heat pumps for cooling and heating, achieve decarbonization, and even change the materials used during building construction. Unlike prior research that focuses on a lot of numbers and statistics or a single energy type, our approach is more so combining everything into one interactive, multivariate visualization dashboard to see the relationships and to drill down to individual buildings for detailed insights. Our project builds on this by providing interactive visualizations to explore multivariate relationships in building energy usage across millions of New York City buildings, enabling deeper insight into efficiency patterns and potential areas for improvement. People are better suited to understand the bigger picture when there's a visualization in front of them, that is why making a scatter plot and a histogram plot are main focuses. And it could not have come at a better time because the official NYC Energy & Water Performance map is down right now, https://energy.cusp.nyu.edu/#/ , users are unable to access this.

Data Description + Visualization Design:

The data for this project was provided by Kaggle and NYC Open Data API. The primary dataset is from NYC Open Data: "***NYC Building Energy and Water Data Disclosure for Local Law 84 2023 to Present (Data for Calendar Year 2022-Present)***". The great thing about this dataset is that it is very recent; it has 103,000 rows and 265 columns giving us an extremely indepth look into New York City buildings. Because of how vast this dataset is, there was cleaning and filtering done. After the preprocessing, there were a total of 99,526 rows and 21 columns. These columns were not randomly selected, we parsed through the dataset and chose relevant columns. Some of these columns are: "Property GFA - Calculated (Buildings) (ft²), Site Energy Use (kBtu), Weather Normalized Site Energy Use (kBtu), Electricity Use - Grid Purchase (kWh), etc. These attributes provide a multivariate view of building energy consumption, allowing us to explore relationships between building size, energy usage, and efficiency.

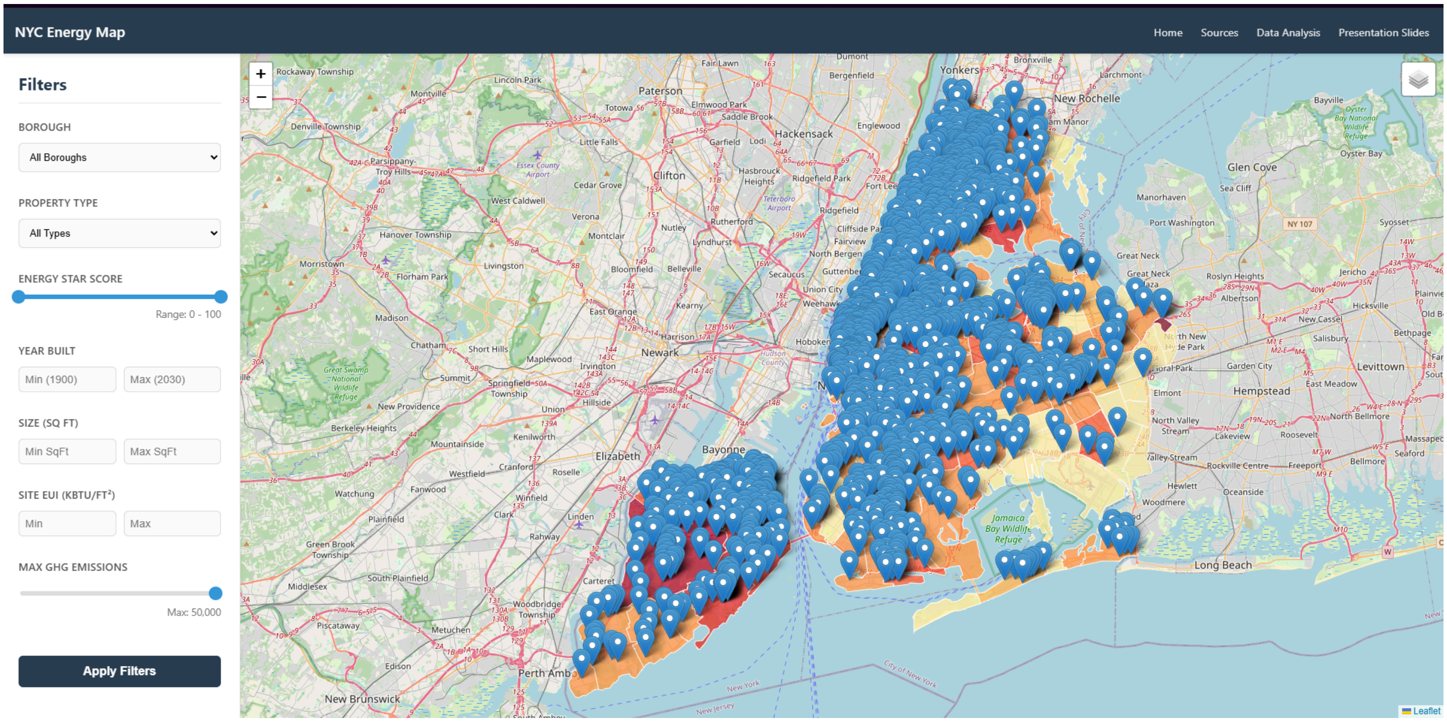
The primary goal of this project is allow users to be aware about patterns in energy usage and consumption across New York City. This project can be used by people who want to buy buildings, rent apartments, or even start a business somewhere; it truly is a multipurpose website. The key goals that we support is comparative analysis, users can compare energy efficiency across building types, sizes, and location. Another goal is that users can actually see how certain factors correlate with one another, the size of the building vs natural gas use for example; or even how weather affects the building energy consumption over time. An underlying factor that a lot of people may ignore or not think is important is the year the building was built. We incorporated this factor into the dashboard because older buildings may consume energy differently due to outdated heating and cooling systems or the use of certain construction materials, which can affect overall energy efficiency. Ultimately, the dashboard allows users to explore multivariate relationships in the data interactively, take a look at patterns that aren't fully clear, and make decisions regarding energy consumption in New York City.

To accomplish these goals, we designed an interactive dashboard that users can apply certain filters too. These filters include the size of the building, ENERGY star score, Max GHG Emissions, and more. With the interactive map, users can fully zoom in onto a certain neighborhood as well as look at certain buildings with the full metrics of ENERGY star score, Site EUI, and GHG Emissions.



A screenshot of the dashboard with filters applied. Queens, 75-100 ENERGY Star Score, and built from 2000-2020

On top of this, we also added data plots for added visualization and clarity. Some of the data plots include bar plots, histograms, scatter plots with density function, and line plots. The histogram plots are used to show the distribution of the ENERGY star score and the Site EUI by Borough. We chose a histogram plot for this because it allows users to see the range and look at any notable patterns or outliers. One of the main histogram plots is "***ENERGY Star Score and Relation to Buildings***". The ENERGY Score is a score from 1-100 that represents the worst performing building to the best performing. A score of 75+ means it is eligible to earn ENERGY Star certification. With the score, you can understand how your building's energy consumption measures up against similar buildings nationwide. And this chart showcases how many buildings in total and their score. Scatterplots are used to show relationships between two large variables. At times it can be hard to understand scatter plots, hence why a density function was added too.

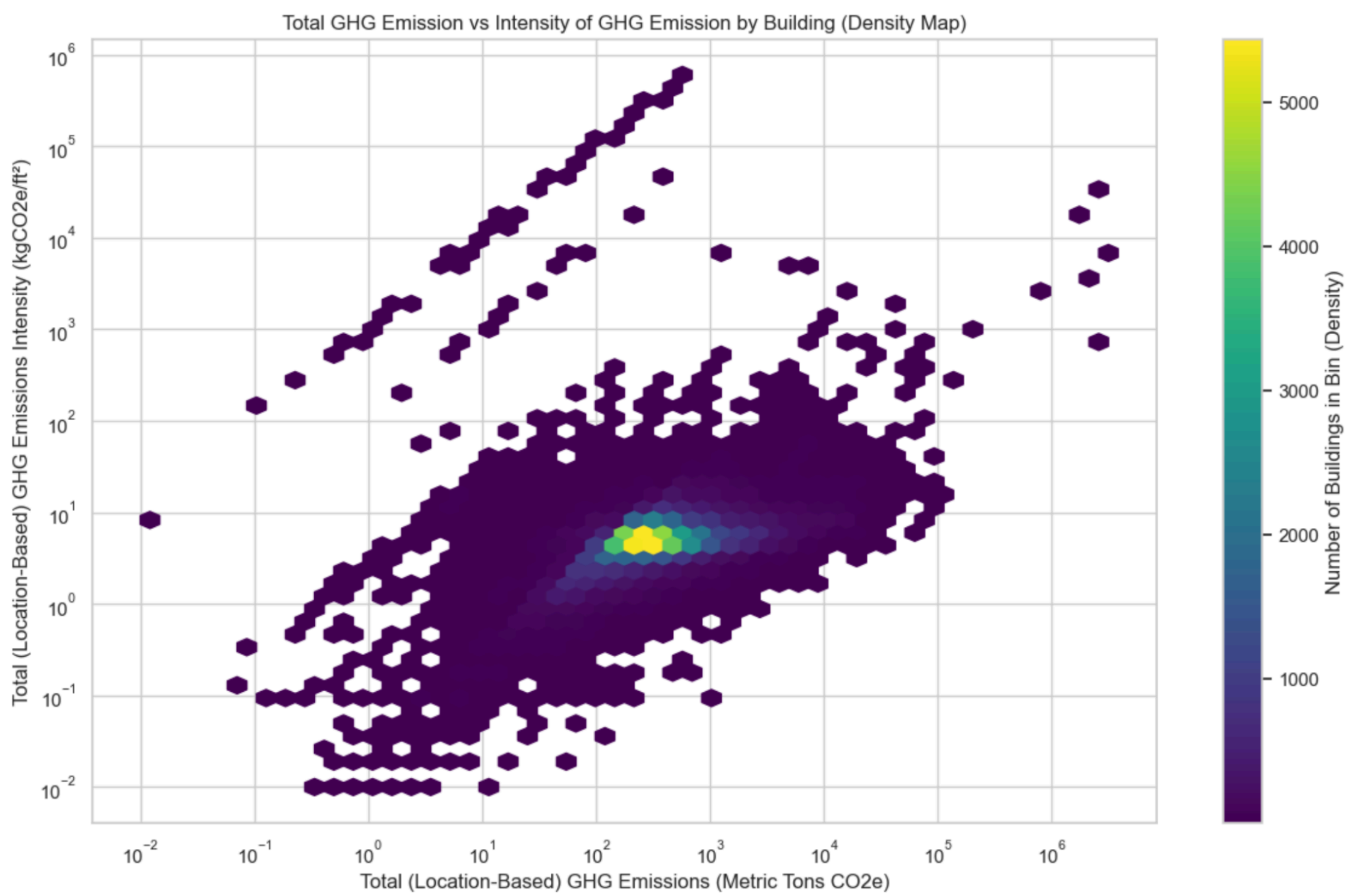


Our project website with the interactive dashboard showcasing energy usage and buildings throughout the 5 boroughs.

Results + Observation + Future Work + Conclusion:

Upon completing the coding process of the project as well as making the visualization plots, we came to multiple conclusions. The ENERGY star score histogram showed a wide spread across the dataset, with a large number of builds in the dataset performing in the accepted threshold of a score of 75 or more. This results in these buildings receiving the ENERGY Star Certification. However, there are still a fair amount of buildings where the ENERGY Star Score falls below this threshold. It shows that some buildings have a lot of work left to do in order to result in a clean energy consumption. Surprisingly, many of the buildings in the dataset have a score of 100, as shown in the histogram and the dashboard. The average Site EUI is calculated by a variety of different energy sources. A building's major energy usage comes from HVAC systems, lighting, and water heating. Combining all of these energy sources, site EUI does add up and become a large number. The bar chart for Site EUI by Borough shows that the five boroughs have a similar average energy use. While it appears close, Staten Island is the one that had a higher Site EUI average compared to all other boroughs. The reason for this is mainly because Staten Island is more of a residential borough, with not that many corporate offices or large buildings like Queens and Manhattan. So these standalone homes result in more energy per square feet for heating and cooling. And this shows how building factors can play a large role in determining energy usage.

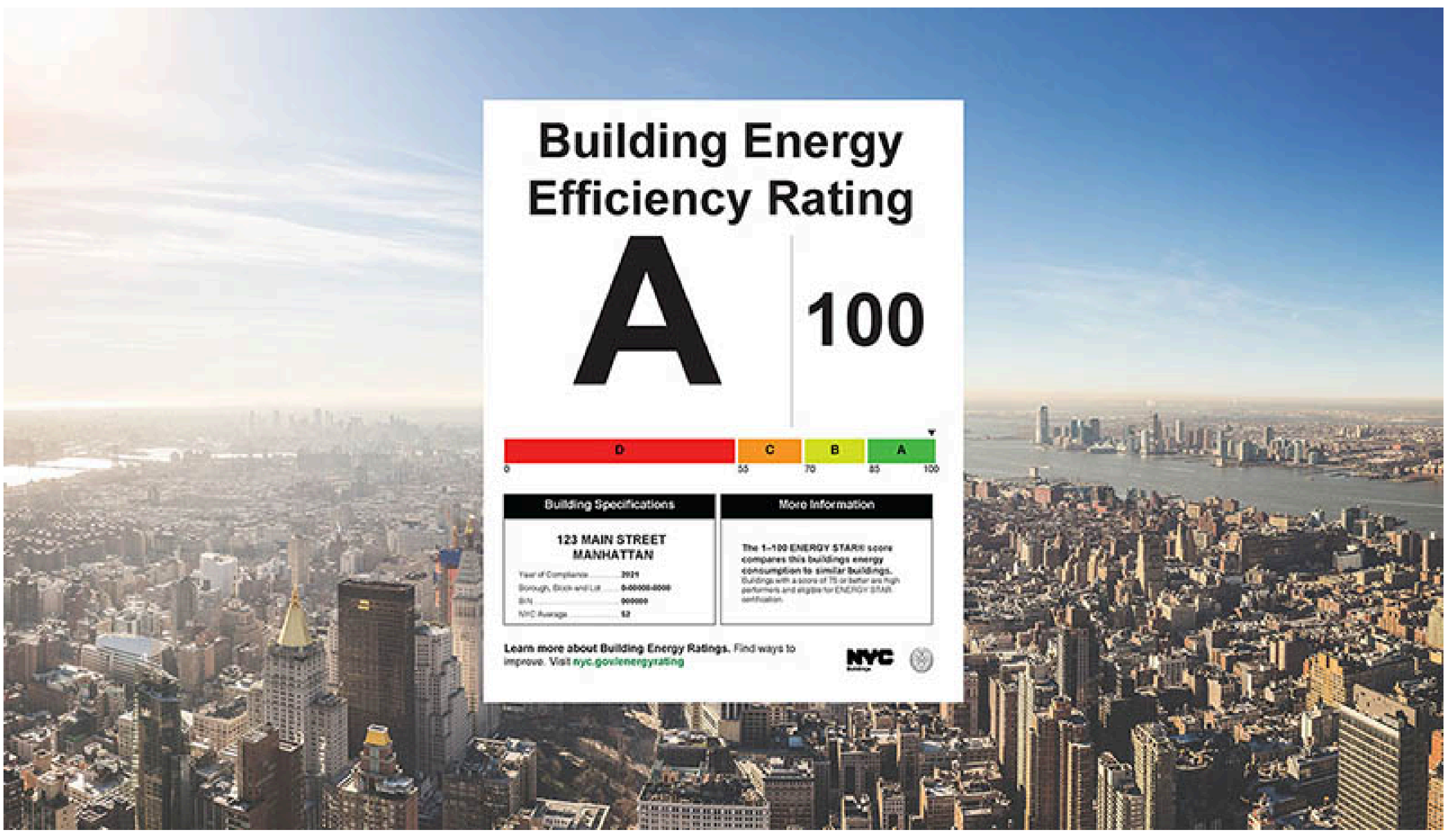
Another bar plot that I made was one that shows the top 10 Neighborhood Tabulation Areas (NTAs) in New York City with the highest total location-based GHG emissions (Metric Tons CO_2e). NTAs represent specific neighborhood areas identified by codes. This chart highlights which neighborhoods contribute the most to carbon emissions from buildings, helping identify areas with the largest environmental impact. Carbon emissions are bad for the environment because they blanket the Earth and trap the sun's heat. Two of the most important visualizations were scatter plots. Initially, the scatter plots were drawn normally, however, it was extremely unreadable with a lot of plots grouped together without the story being clear. To remedy this, a density function was also added to the scatter plot. The hexbin plots compare the size of a building (Gross Floor Area, ft^2) with its annual electricity consumption (kWh). Only buildings with positive values for both metrics are included. The log-log scale is used to visualize the wide range of values. The color intensity represents the number of buildings (density) falling into that hexagonal bin, making the distribution of building size and electricity use clearer. Buildings that emit more total greenhouse gases clearly have higher emission intensity, thus this scatterplot shows a positive relationship. Most buildings cluster in a midrange zone, with dense regions indicated by the people in the plot below. Like with most plots, there are outliers. In this case, there are a small number of buildings in the upper right corner. This means that they emit a lot of gases and have high emissions per square foot. Most likely, these are only large buildings like hospitals and data centers. The great thing about the scatter plot is that it highlights how New York City officials should consider large footprint buildings and high intensity buildings because too much greenhouse gas can cause the climate to rise and make a negative impact in the future.



A density function scatter plot, made with hexbins. This is a relation between the total GHG Emission and the intensity of the GHG Emission by building.

With the goal of this project completed, there are still a reasonable number of ways that this project can be improved. One really cool feature that can be implemented is temporal analytics, this is more richer data. This will allow users to view how individual building energy consumption and GHG emissions vary month by month. This would be a much richer histogram and bar plot because then it would be more in depth and allow to see long term trends, the before and after if a building makes drastic changes (changes to HVAC, construction on the building). A second great improvement would be incorporating more machine learning into the project, although machine learning was used to a fair extent, predictive machine learning, especially using linear regression, would be a great idea to potentially predict the ENERGY star score, expected consumption of energy, and expected emission of greenhouse gases. Lastly, a potential large improvement to the dataset could be to incorporate real time data. This will allow the website to be more of a monitoring tool rather than an exploration model. And it would be interesting to potentially allow homeowners and business owners to add their own property as well.

This project demonstrates the power of data visualization and machine learning to understand how energy usage correlates to building efficiency. With this project, we were able to analyze and visualize building-level energy consumption trends across time, weather, energy type, and building type. Using the interactive dashboard, this enables users to explore patterns that are not visible to the raw data alone. With the use of bar plots, scatter plots, line plots, and histograms we were able to clearly identify the patterns on how the size of the building, location of the building, and type of the building relate to the emissions, ENERGY star score, and site EUI. With all of these tools available, these insights till help people interested in buying their first house, renting an apartment, buying their own building, or starting somewhere new. Taking it a step further, this project can also help New York City policymakers and those who work with the energy department. Overall, the project definitely meets the goals of having multivariate relationships, having an interactive dashboard, and large scale data visualization. With more data and more future developments, this website can be a replacement to the current outdated and underperformed NYC Energy and Water Performance Map.



How a NYC Building Energy Efficiency looks like. This is the ENERGY star score officially.

Works Cited Page

1. "Energy Efficiency and Building Decarbonization." NYSERDA, www.nyserda.ny.gov/Impact-Energy-Efficiency-and-Building-Decarbonization

2. "Why Building Energy Codes? | Building Energy Codes Program." Www.energycodes.gov, www.energycodes.gov/why-building-energy-codes

3. "Buildings Energy Efficiency." Energy.gov, 19 Nov. 2024, www.energy.gov/topics/buildings-energy-efficiency

4. audreymcmanemin. "Enhancing Building Efficiency Modeling." Evolved-Energy, 6 Sept. 2024, www.evolved.energy/post/enhancingbuildingefficiencymodeling Accessed 10 Dec. 2025

5. "NYC Energy and Water Performance Map." Nyu.edu, 2021, energy.cusp.nyu.edu/#/

6. ashioyajotham. "NYC GEOJSON FILE." Kaggle.com, 2022, www.kaggle.com/datasets/ashioyajotham/nyc-geojson-file Accessed 10 Dec. 2025

7. "NYC Building Energy and Water Data Disclosure for Local Law 84 (2023-Present) | NYC Open Data." Data.cityofnewyork.us, data.cityofnewyork.us/Environment/NYC-Building-Energy-and-Water-Data-Disclosure-for-/5zyy-y8am/about_data