Water Use in NYCHA Buildings

Anil Bhusal , Data Science Scholars
The City College of New York,
160 Convent Ave, New York, 10031

Abstract

In the light of looming future extreme droughts, it is necessary to keep track of water consumption for mega cities like New York (NYC), one of the most densely populated cities in the world. NYC uses nearly one billion gallons of water per day. Thus, it requires a careful analysis to supply the right amount of clean/pure water. We use NYCHA buildings (it is home to 1 in 16 New Yorkers in 2022) data for our analysis. They are easily accessible and public buildings which need more attention for conservation. We analyze water consumption and patterns of consuming water over nine years in each borough and buildings. We also work to find the high consumption areas and discover the relationship between NYCHA and its importance in controlling extreme consumption. Some of the significant methodologies are Data Preparation, Yearly analysis (it will give us overall scenario for each borough from 2013 to 2019), Seasonality analysis (it will inform us about the seasonal behavior of each boroughs), Extreme Percentile (it will give us idea about the extreme consumed areas and buildings that requires more attention for the conservation of the water), and Trend Analysis (which will give detail information about change in consumption). We found out that the most extreme consumption in Manhattan and Brooklyn is during winter months. Overall, water consumption is also increasing. As for Brooklyn, the zip code 11225, consumption is increasing rapidly, which requires more attention. Similarly, Zip code 10030 is consuming excess water in Manhattan. Also, it is important that people should be awarded for conservation of extreme consumption.

Introduction

Even though water is a basic need to survive, 785 million people lack drinking-water services globally(The Water Crisis), which is more than double the population of the United States. Also, a child dies of waterborne diseases every 15 seconds(Chambers). So, it is necessary to have a clear analysis of water use to ensure its safe and secure access. We choose New York City, which is densely populated and uses nearly one billion gallons of water each day(Chambers) as a case study. In New York City, we will use NYCHA buildings which are home to 1 in 16 New Yorkers in 2022(NYCHA Fact Sheet 2022), for our analysis. In NYCHA buildings, we will mainly analyze the pattern of its consumption in each borough over the nine years, find the high consumption buildings or areas for conservation of the extreme consumption, and discover the relationship between these buildings with New York City and globally.

Methods

- **Data Cleaning:**
  We first prepare data for the analysis. In the data preparation process, we clean duplicate data initially. After that, we split the data in terms of boroughs. In each borough, we find out the unique meter numbers of each building. After that, we set outliers to clean our data from the non-uniform dataset. We calculate the outliers using the formula:
  (x - mean) / standard deviation.
  Here, X represents the value of the consumption column and mean is the average value of the consumption column. Finally, we will get the outliers and we set normal values to be in the range of 10 deviation values from the above equation.
  After that, we also remove the insufficient data set from our original data to minimize the error in our analysis. During this insufficient data cleaning process, we set that the building should have at least 40 months of their data to be involved for the analysis over nine years. Also, we check if each building has all the data for nine years. If a building doesn't have data for nine years, we will randomly select the values for the missing month from its data set. For example, if the missing month of a building is February of any year, we choose the data for that date from the available January, February, and March datasets). Finally, we will have prepared a data set for our further analysis process.
- **Yearly Analysis:**
  After the data preparation process, we use these data sets for our different methods of analysis. The first method of analysis we performed is Yearly analysis, where we analyze the pattern of consumption between 2013 to 2019. We sum each month of each year among each building and calculate the overall yearly report.
- **Seasonality Analysis:**
  The second method of analysis is Seasonality analysis, where we will look over the average monthly behavior of the database. It will lead us to see how each borough is acting in each season. Their nature might differ in Fall, Spring, summer, and Winter. Hence, seasonality analysis will give us analysis in terms of season. We group each month and average their value to see the seasonality analysis.
- **Extreme Percentile:**
  The third method of analysis is the Extreme Percentile method, where we set our percentile to be 95% and find out the 95th percentile of each building. We use the below formula to calculate 95 percentile.
  N = 95100* total number values in the column
  From the above equation we will get the position of 95th extreme value from the sorted column. And any value greater than the value will be the extreme values.
  It will inform us maximum consumption is occurring in that specific month because most of the extreme consumption is happening in that period. we will discuss different factors of extreme consumption months like the temperature in that period and area of consumption. Also, we will visualize the occurrence of extreme building in each zip code. We calculate this using formula as,
  Probability: Total number of occurrence in extreme months / 12
  Finally, we will see how many times each building are occurring over each zip code.
- **Trend Analysis:**
  Finally, we will also analyze the pattern of each building whether their consumption is increasing or decreasing, or remaining constant. It will help us predict which places require more attention or which area is more responsible for high consumption.

Results

Manhattan

Brooklyn

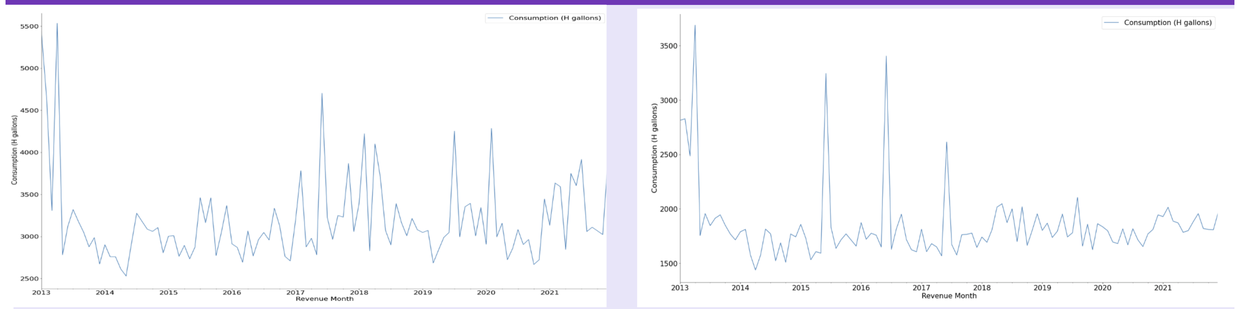

Figure 1: Yearly analysis of Consumption of water in Hundred gallons over nine years

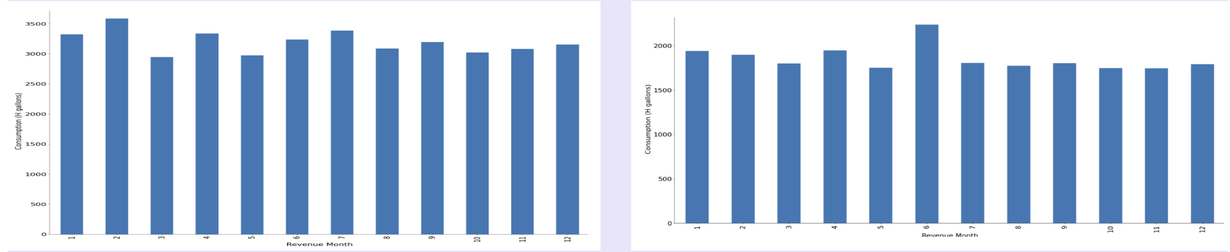

Figure 2: Seasonality analysis of Consumption of water

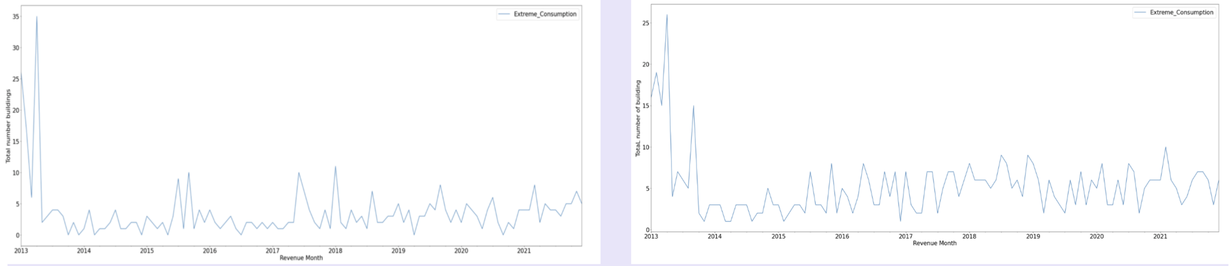

Figure 3: 95th Percentile

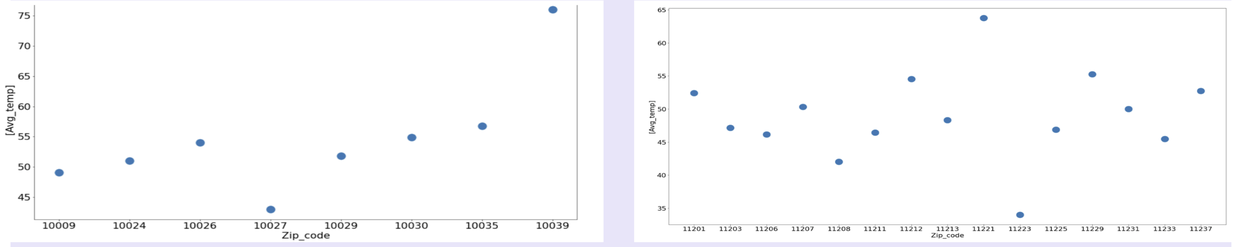

Figure 4: Average Temperature Analysis over Zip Code

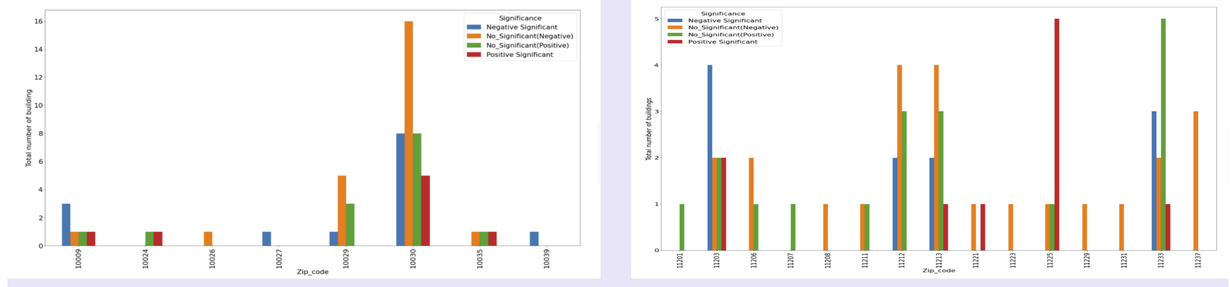

Figure 5: Trend Analysis

Borough Comparison

We are comparing only two boroughs of NYC. Comparing the Yearly analysis of Manhattan and Brooklyn, we found that both boroughs were high during early 2013. Brooklyn had few repetitive high consumptions in mid-2015, 2016, and 2017. Manhattan's overall consumption has increased since 2017, and also Brooklyn's Consumption slightly increased since around 2018. We can say that overall consumption of water is increasing after a certain period of time and it is remaining constant with few irregular high and low consumptions. Comparing the Seasonality analysis, By Looking at the seasonality figure for Manhattan in Figure 2, there is a high consumption in February and lowest in March. Overall their consumption is close to each month. For Brookllyn, June has the highest consumption. But looking at overall or the rest of the months, they are very close to each other as well. Thus, we can say that their water consumption does not gradually change over the season. After Comparing the extreme percentile of Manhattan and Brooklyn. We found out the top 12 months from each of the boroughs. Early 2013 has had similar extreme months for both, and the rest of the extreme months differ. For both of the boroughs April 2013 is an extreme month. During the extreme month only one building from Brooklyn has occurred for all twelve extreme peaks. In Manhattan none of the buildings appeared to be for all top twelve peaks. only three buildings in Manhattan have occurred rapidly for a maximum of 5 occurrences during these top twelve peaks. Similarly for brooklyn three buildings are occurring for 5 times of extreme peaks. Comparing the overall extreme occurrence, Manhattan buildings are occurring three times in top extreme months whereas Brooklyn buildings mostly occur at twice. Comparing the total number of buildings with zip codes in each borough, we found that 10030 is a densely populated area in Manhattan, and zip code 11233 is a densely populated zip code in Brooklyn. Brooklyn has the most buildings among all the boroughs. Also, there are more zip codes where people reside in Brooklyn. But, looking at Manhattan, most of the buildings are in the same Zip Code 10030. There is more spread out in extreme consumption in each zip code in Brooklyn. Comparing the Temperature of Manhattan and Brooklyn, we found that most buildings have lower temperatures in high consumption months. It means people use more water in cold weather, which proves our initial hypothesis. Comparing the trend analysis of each building, we calculate the Significance of each building in each zip code. Red, blue, green, and orange represent the Positive significance, Negative Significance, Positive non-significance, and Negative non-Significance, respectively. We need to be more focused on Positive Significance, because they are rapidly increasing over time. For Manhattan, ZipCode 10030 has the highest significant value, Similarly, For Brooklyn, Zip Code 11225 has the highest Positive Significance. Also, 11203 has the second highest number of positive significance buildings. They need more attention for high consumption of water. Hence, we should keep an eye on these extreme zip codes for conservation of extreme consumption of water.

Conclusion

We found out that the most extreme consumption in Manhattan and Brooklyn is during cold weather. Also, the ratio of consumption of water is increasing periodically. Zip code 10030 is consuming excess water in Manhattan. Among them, five buildings are contributing to extreme consumption. For Brooklyn, zip code 11225, consumption is increasing rapidly, which requires more attention for the conservation of extreme water. The Brooklyn data set is more distributed compared to the Manhattan data set. In future, using alternative methods for heating systems and more awareness programs for water conservation should be conducted timely. Identifying the problems of each high consumed area around the world, and solving them to save water from high consumption will make water easily accessible for all the people around the world. It will finally lead us to a water problem free world. Hence, we should control the excess use of water all around.

Refrences

- Chambers, Aaron. “How Much Water Does NYC Use Daily?” Hudson Reed, 23 Apr. 2018, usa.hudsonreed.com/info/blog/the-truth-about-new-yorks-water-usage/#:%7E:text=New%20York%20City%20uses%20nearly,all%20of%20the%20facts%20below%E2%80%A6
- “NYCHA Fact Sheet 2022.” NYCHA, 2022, www1.nyc.gov/assets/nycha/downloads/pdf/NYCHA_Fact_Sheet_2022.pdf.
- “Water Consumption And Cost (2013 - Feb 2022) | NYC Open Data.” NYC Open Data, 13 May 2022, data.cityofnewyork.us/Housing-Development/Water-Consumption-And-Cost-2013-Feb-2022-/66be-66yr.
- “Water Consumption in the City of New York | State of New York.” Open NYC, data.ny.gov/widgets/ia2d-e54m. Accessed 5 Aug. 2022.
- “The Water Crisis.” The Last Well, thelastwell.org/the-water-crisis/?gclid=Cj0KCQjwuaiXBhCCARIsAKZLt3nSmGXI4cVjFalKOSGB8GU2eZ6-6vhIZeG34X4PuRPG8saZ5o1CrVgaAjG2EALw_wcB. Accessed 3 Aug. 2022.

Acknowledgements

This research was conducted online during Summer 2022. I thank Dr. Naresh Devineni for giving me an opportunity to work in this research and guiding in every step. I also thank the Data Science Scholars Program and Dr. Claude Brathwaite for providing a great platform to work on this project.