Using K-Means Clustering and Neural Net Analysis to Define and Predict Chicago Neighborhood Energy Consumption Trends.

Abstract

Energy markets worldwide are facing two primary issues: 1) meeting exponentially increasing demand and 2) satisfying external pressures to transition to clean energy sources. Before addressing these challenges, policymakers and marketmakers need access to analyses of current energy consumption levels to facilitate efficient planning and investment. This report studies electricity consumption patterns in the City of Chicago through historical socioeconomic data. Using an unsupervised method of analysis (K-Means Cluster) together with a supervised method (Neural Net), our report provides three classifications of Chicago neighborhoods and identifies the strongest predictors of electricity consumption in each. Using this analysis framework, major cities can effectively devise long-term strategies for meeting energy demand while planning for a sustainable future.

Introduction

Electricity demand is rising globally (EIA, 2019) and cities like Chicago are increasingly under pressure to meet it through sustainable means. In April of 2019, the City of Chicago committed to operating on 100% renewable energies by 2035 (City of Chicago, 2019). With this commitment comes a massive call for investment and planning, which first require a great deal of analysis into the consumption patterns of the City. Using historical economic, demographic, and energy usage data, this paper examines the current electricity consumption trends of Chicago neighborhoods in effort to better understand and predict the changing energy landscape.

There are many potential applications of this research. We plan to focus our efforts on setting the stage for two primary extensions:

1. Given that renewable energies often operate on an intermittent basis, matching their production to consumption can be a difficult task. Large scale battery storage technologies currently have very high costs and their widespread development is still years away. As such, this stage of the energy transition is primarily focused on incorporating renewables into existing portfolios of baseline electricity-generating sources, which requires supply and demand optimization. The first step to this matching process is understanding the systematic energy usage patterns of our grid.
2. For city officials and utilities on the supply side of the grid, accurately planning for future infrastructure developments is a key responsibility. It is crucial to have a detailed understanding of the problemed areas in order for this planning to be effective. Similarly, when looking for areas for improvement on the demand side,

energy efficiency programs should be focused on high-consumption customer groups.

In the following sections, we will discuss similar research completed in the energy field, provide a detailed description of our datasets, explain the analysis models we employed, and present the actionable insights that resulted. In whole, this report serves not only as a tool for the City of Chicago, but also as a framework for other major cities facing the need to understand and predict electricity consumption.

Literature Review

Understanding the electricity consumption growth trends of various urban and social classes has become a topic of interest to policy makers in recent years. In order to create and change energy policy, lawmakers must have the most current, holistic data available. Researchers have been able to use various data models to predict energy consumption trends on a national and local scale. Our team has incorporated various elements of such research, as described below, into our clustering and analysis model.

In “Forecasting Energy Demand in Jordan Using Artificial Neural Networks”, a study conducted by Bassam AbuAl-Foul, the author describes his attempt to forecast energy usage in Jordan using annual data from 1976-2008 (AbuAl-Foul 473). Jordan has suffered from energy scarcity in recent decades and relies on energy imports to meet its current demand. Mr. AbuAl-Foul applied artificial neural networks to the data available from 1976-2010 in an effort to help the government understand the rate of energy consumption growth that would take place in the coming years, specifically covering the period from 2010-2025. The author also used various socioeconomic indicators to aid in this model. Mr AbuAl-Foul concluded that “expected energy use will reach 8349, 9269, and 10189 Kt. (kilo tonnes) of oil equivalent in 2015, 2020, and 2025 respectively” (AbuAl-Foul 476). This conclusion aids the government of Jordan in creating a realistic energy consumption plan and aids the country’s lawmakers when creating and changing energy policy in the timeframe of the study. This research study also aids other researchers, such as ourselves, when creating an energy consumption forecasting model. The author’s inclusion of three socioeconomic indicators within his model may also help us determine which indicators are best to use in our model concerning energy consumption trends across the neighborhoods of Chicago.

In our initial discussion of this research, creating a model that included the energy consumption of buildings over 50,000 sq ft. was an important factor to consider. Our search for prior research incorporating such factors led us to an article entitled "Development and validation of regression models to predict monthly heating demand for residential buildings" by Tiberu Catalina, Joseph Virgone, and Eric Blanco. The French research study used regression models to predict monthly heating for various residential buildings (Catalina et al 1825). Their intent was to aid architects and design engineers in finding efficient energy construction designs and solutions. The model uses various design elements in residential buildings such as the building shape factor, window-floor area ratio and the climate index as a function of the sol-air temperature and heating set-point. During the validation stage, 270 scenarios were analysed for different inputs and the authors concluded that energy equations constructed for the model were able to predict heating demand for multiple building structures. While our analysis will not focus specifically on predicting the energy consumption of the buildings within Chicago, the study provides context for our research and offers a variable input that may be factored into future models and research.

When trying to define the data sets to be used in this research paper, our team experienced the natural limitations that come with government collection of energy consumption trends. Our original intent was to use data that reflected Chicago's current consumption patterns, ideally using data from 2015 to present. Unfortunately, the costs associated with surveying and collecting such data greatly limits the frequency in which data is collected and published, especially in a city as large as Chicago. This problem is highlighted and addressed in the study "Comparative Analysis for the Chicago Energy Retrofit Project: Project Report", a project conducted by the Argonne National Laboratory under the U.S. Department of Energy. In the study, researchers at Argonne developed and piloted "analytic methods that capture the energy performance of individual commercial buildings" (Guzowski et al vii). This was done through the creation of an energy calculator, *EECalc.* The researchers highlight the calculator's efficiency, use of observable data that may be collected with very little effort, and the limited expertise necessary from users to run correctly. Furthermore, the calculator was shown to predict energy savings and consumption by building "as well as professional auditors" Guzowski et al vii). This would allow cities like Chicago to publish energy consumption data regularly without incurring high costs. As it relates to this project, the *EECalc* was shown to be scalable to communities. Using the K-Means clustering to determine similarities among certain Chicago neighborhoods, our research could be combined with something similar to the energy calculator to predict energy consumption growth on a community basis among Chicago's neighborhoods.

As mentioned previously, this research team intended to incorporate various social and economic factors into the prediction of energy consumption trends among Chicago's 77 neighborhoods. Previous research related to this includes R.F. Hirsh and J.G. Koomey's "Electricity Consumption and Economic Growth: A New Relationship with Significant Consequences?". This article discusses the historical relationship between electricity consumption and economic growth using data from 1949-2010 (Hirsh and Koomey 73). Surprisingly, up until the 1970s, energy consumption grew steadily compared to GDP. By the 1990s, energy consumption was growing relatively slower than GDP, and since 2007, there has been limited correlation between the two factors. This conclusion predicts an increased scrutiny in future energy prediction models. It also verifies the need to include social demographic factors (such as education levels, racial information for social habits, and types of homes occupied) to create a holistic model of today's energy consumption habits across diverse communities. Similarly, the article helps to show that predictive models should be limited in range (5-10 years) as correlated factors may either emerge or become obsolete.

Finally, an application of energy consumption predictions to urban areas can be found in Thomas Fullerton and Adam Walke's "Empirical evidence regarding electricity consumption and urban economic growth", published in 2018. The two researchers compare the growth of energy consumption in El Paso, Texas with several markers of economic health, such as related stock prices and employment measures (Fullerton and Walke 1976). The authors acknowledge that national economic growth and related growth energy consumption has been documented previously, however, there has been very little research conducted regarding a correlation between electricity consumption and economic growth on a regional level, applied to an urban economy. While their conclusion presents the same findings as national research, this study highlights the importance and necessity for localized urban research for policy makers. Ideally, our study will capture the essence and premise of such a localized study, adding to the limited amount of research available.

Data Description

The primary dataset used for socioeconomic data is the Chicago Metropolitan Agency for Planning's (CMAP) '2017 Chicago Community Area Data Snapshots'. This dataset provides comprehensive data for all 77 of Chicago's community areas. The points we extracted for use in our research were the data entries for each community area's total population, population change from 2000-2017, population by age groups (under 19, 20-34, 35-49, 50-64, 65-74, 75-84, over 85), race (White, Black, Hispanic, Asian, other), employment/unemployment rate, level of education (Less than high school, high school, some college, associate's, bachelor's, graduate's), income levels (Less than 25k, 25-50k, 50-75k, 75-100k, 100-150k, greater than 150k), home value (less than 150k, 150-300k, 300-500k, greater than 500k), and primary industry. We included the granular and nominal values for the age, race, education, income, and home value categories. We extracted these points from the dataset because we believe they are the most distinct, useful, and identifiable socioeconomic traits of communities.

The dataset used to correlate energy consumption is 'Average Electricity Usage per Square Foot by Community Area', which contains data collected from the 2010 calendar year 'aggregated from ComEd and Peoples Natural Gas by Accenture'. This data was published on the City of Chicago data portal and comprises 88% of Chicago's electrical and gas usage for the year. This was the most comprehensive dataset we could find in relation to energy consumption in Chicago, and when we reached out wondering if there were any newer studies, we were told there were unlikely to be anymore in the near future because of how difficult it was to collect this data. The dataset itself simply consists of two columns, one being the community area, and the other being electricity usage per square foot measured in Kilowatt-hours. Because of this, we joined the community area columns of our datasets to aggregate them as one dataset.

Two datasets were used in early stages of testing our model but ultimately were left out of the final project. One was the Chicago energy benchmarking report which contains electricity usage for buildings larger than 50,000 square feet. It was left unused due to some inconsistencies in the dataset and we felt it was not nearly comprehensive enough as it only covered about 1% of Chicago's buildings. The other dataset had the count of new building permits each year for each community area. We ultimately decided that because it did not actually give a good idea of when construction was taking place, it did not reflect growth in infrastructure as we hoped it would.

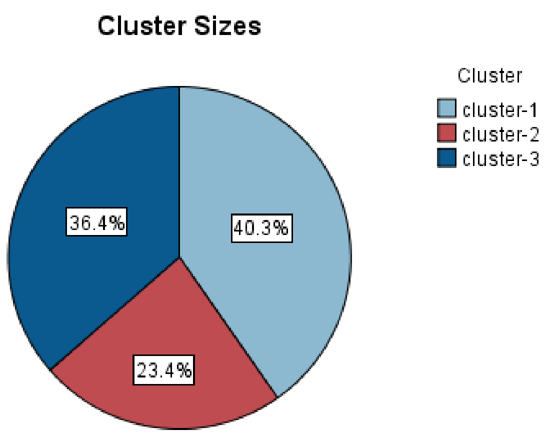
Model Explanation

We performed a preliminary cluster analysis on community area data in order to gain a more holistic view of the Chicago socio economic landscape. While this information did not directly relate to our goal of predicting electricity consumption, understanding the attributes of Chicago neighborhood categories was pivotal in interpreting the City's systematic energy trends at a deeper level. After testing various cluster analysis techniques, we found the K-Means method to be the most accurate given that its clusters exhibited the highest cohesion and separation ratings. Utilizing this clustering technique alongside our neural network analysis will provide us with a clear prediction and actionable results.

Additionally, utilizing a model that provided us with an actionable prediction while also allowing us to make iterable changes when incorporating more data was essential for this project. As such, the model we used was a neural network. Neural networks are extremely comprehensive, especially with data as sporadic as energy usage. Also, when considering the scope of this project, utilizing a neural network provided us the opportunity to incorporate as many inputs as necessary to generate an insightful prediction. Furthermore, as noted above in the Literature Review, neural networks have proven to be a useful technique for understanding complicated energy trends. They have shown positive results in predicting energy consumption levels in cities across the world, and we find that this methodology provided informative findings in Chicago as well.

Results

Based on the initial clustering of the data, each community area in Chicago fell into one of three distinct groups.



Size of Smallest Cluster	18 (23.4%)
Size of Largest Cluster	31 (40.3%)
Ratio of Sizes: Largest Cluster to Smallest Cluster	1.72

Cluster 1:

The first and largest cluster, consisting of 31 community areas, was characterized by being predominantly white. It also had low percentages of black or hispanic residents, and low unemployment rate.

The neural net was able to predict the electricity consumption of neighborhoods that belonged to this cluster with an accuracy of 61.4%. The two highest predictors were the percentage of the population that were between the ages of 75 and 84 or between the ages of 20 and 34, respectively. This indicates that the behavior of those age groups uses a particularly large or small amount of electricity. After age, the total population and median house value were the highest predictors of overall electricity consumption, for obvious reasons.

Cluster 2:

The second cluster only contained 18 community areas but was more varied. It had a high percentage of hispanic residents and low percentages of white or black residents. Its rate of unemployment also fell squarely in the middle between the other two clusters.

The neural net was only able to predict the electricity consumption of this cluster with 12.8% accuracy. We attribute this to the fact that this cluster was much more diverse than the other two, both in terms of racial and economic factors. Many neighborhoods that had high populations of other ethnicities were also grouped with this cluster, further complicating the inputs. Due to the low accuracy rate, we can assume that the results from this cluster are invalid and do not yield any significant insights.

Cluster 3:

The third and final cluster contained 28 community areas and was characterized by a high percentage of black residents and a low percentage of white or hispanic residents. This cluster was also distinct from the first, as it had a high rate of unemployment.

The neural net predicted the electricity consumption of neighborhoods in this cluster with 71.6% accuracy. Age was again very influential on the overall electricity consumption. The percentage of residents between the ages of 20 and 34 and residents between the ages of 65 and 74 were the two highest predictors, respectively. Total population estimate was the fourth highest predictor as well. It is worth noting that the percentage of the population that had a yearly income of $50K to $75K was the third highest predictor. This is significant because many residents most likely limit their electricity consumption due to financial factors.

Overall:

We analyzed each cluster individually and consider the insights gathered from Cluster 1 and Cluster 3 to be actionable and that the problems that arose when analyzing the second cluster do not affect the rest of our results.

Conclusion

In this study, applying an unsupervised method of analysis (K-Means clustering) allowed us to identify three racially-differentiated clusters among Chicago’s 77 neighborhoods. We then applied a neural net to determine key socioeconomic indicators related to the energy consumption within each of the three clusters. This allowed us to determine the leading indicators within each cluster to be used to predict the energy consumption and growth rate of the three clusters encompassing Chicago’s 77 neighborhoods. Such an ability to predict future electricity consumption may be useful when preparing for the rapidly rising electricity consumption patterns seen globally, especially given that governments may no longer rely solely on economic growth factors. Our research may also be used in creating energy consumption public policies and tailoring the policies to the three distinct neighborhood types.

The City of Chicago may also consider using such a model to match future energy consumption with renewable energy sources, meeting its goal to operate entirely on renewable sources of energy by 2035. When trying to meet this goal, the city may look to see which neighborhoods to focus their efforts on by consulting this model. After defining the areas to target, the city may try to target certain demographics, such as Gen Z or Millennials, based on the likelihood of participation in public initiatives. In searching for renewable energy solutions, policy makers may consult this model, capable of predicting future energy consumption in Chicago's distinct neighborhoods, to provide meaningful insights relevant to Chicago's transition to a reduced carbon future.

Limitations & Further Research

This analysis came with many data-related limitations. One prominent barrier was the scarcity of historical electricity consumption data. The most recent set of available data is from 2010 and lacks granular detail. Due to the difficult nature of the work, the City of Chicago has not endeavored to update the study since. This considerably restricted the application of socioeconomic data.

Additionally, the varied characteristics of the third and smallest cluster added complications to the overall analysis. The first two clusters were clearly defined in their characteristics and painted two polarizing ends of the spectrum. For both clusters, the neural net was applied effectively for our needs. The third cluster contained all of the records that fell between the extremes and exhibited an array of neighborhood archetypes. Because few characteristic categories in the third cluster maintained any kind of consistency, the neural network had a much harder time finding variables that accurately estimated electricity consumption.

From this point, further research is required to quantitatively predict future electricity consumption. We have found meaningful correlations between certain socioeconomic indicators and electricity consumption habits based on existing data, however, to translate these correlations into numeric future values requires both more modeling and more data. Ultimately, these studies will aid policy makers and market makers in the task matching renewable energy production cycles with municipal energy consumption patterns.

Aside from its assistance in addressing electricity consumption, our research also has applications in other areas of study. The clustering performed with the Chicago community areas provides for insightful discoveries. While our target has been energy, other targets could be implemented to find trends using the same demographic analysis model. This could benefit city infrastructure planning as a whole, allowing for more effective and efficient development projects.

Lastly, it is worth noting that this research is meant to be *replicated* in other cities, not simply taken as *fact*. Chicago neighborhoods are uniquely diverse on a socioeconomic level, and their energy consumption patterns are highly influenced by the City's distinctive climate. In applications elsewhere, the analysis methodology presented above is meant to provide a concise framework for studying energy consumption patterns in a specific geography.

References

AbuAl-Foul, Bassam M. "Forecasting Energy Demand in Jordan Using Artificial Neural Networks". *Topics in Middle Eastern and North African Economies*, vol. 14, 2012. https://ecommons.luc.edu/cgi/viewcontent.cgi?article=1169&context=meea. Accessed 12 February, 2020.

Catalina, Tiberiu, et al. "Development and validation of regression models to predict monthly heating demand for residential buildings." *Journal of Energy and Buildings*, vol. 40, no. 10, 2008, pp. 1825-1832.

Fullerton, Thomas M. and Walke, Adam G. "Empirical evidence regarding electricity consumption and urban economic growth." *Journal of Applied Economics*, vol. 51, no. 18, 2019, pp. 1977-1988.

Guzowski, Leah B., et al. *Comparative Analysis for the Chicago Energy Retrofit Project: Project Report.* Chicago: Argonne National Laboratory, 2014. Web. https://publications.anl.gov/anlpubs/2014/04/78280.pdf. Accessed 23 February, 2020.

Hirsh, Richard F. and Koomey, Jonathan G. "Electricity Consumption and Economic Growth: A New Relationship with Significant Consequences?" *The Electricity Journal*, vol. 28, no. 9, 2015, pp. 72-84.