The Electricity Journal 32 (2019) 106643

Contents lists available at ScienceDirect

The Electricity Journal

journal homepage: www.elsevier.com/locate/tej

Six unique load shapes: A segmentation analysis of Illinois residential electricity consumers

Jeff Zethmayr*, Ramandeep Singh Makhija

Citizens Utility Board, 309 W. Washington St. Suite 800, Chicago, IL, 60606 United States

ARTICLE INFO

Keywords:
Energy efficiency
State policy
Electric utility
Rate design
Customer segmentation
Advanced Metering Infrastructure

ABSTRACT

Many electricity policies have disparate impacts on consumers with different load shapes. This study applies k-means clustering to 2.5 million Illinois customers, and matches six resulting clusters with demographics. Flatter load shapes were more likely in urban and low-income areas, with high-volume, peak usage more likely in high-income/suburban areas. This highlights potential for grid cost reduction through DR programs targeting suburban areas, and illustrates potential cross-subsidization intrinsic to common electric rate designs.

1. Introduction

Discussions of electric utility regulation have long recognized the impact of peak load on total system costs. According to Faruqui et al., "even a 5% reduction in peak demand in the United States could lower consumer energy costs by at least $3 billion a year."[1]

For this reason, economists have argued for the creation of time-variant rates for the various components of electric service, as an encouragement for consumers to manage their usage to lower their contribution to peak system load and improve system load shape. In a similar vein, demand response programs are used to induce lower usage of electricity at times of peak energy demand, and many utilities have initiated energy efficiency programs to provide incentives to consumers to install energy saving devices and other home upgrades to reduce overall energy consumption.

To support these initiatives and enhance service reliability, many utilities have upgraded their metering systems with the implementation of Advanced Metering Infrastructure (AMI), including smart meters.[2] AMI allows the utility to collect customer usage volumes in 15 min (or less) increments throughout the day rather than on a monthly basis.

This provides a wealth of information into the way individual customers use electricity and their actual contribution to the system's peak load. It also allows for customer segmentation techniques that were previously unavailable to researchers. Attaining a better understanding of the differences in how consumers actually use electricity is important for several reasons.[3]

First, it can help improve energy efficiency, demand response, and time-variant pricing program design.[4] By targeting a particular program at those most likely to benefit from it, utilities can maximize cost-effectiveness and achieve more "bang for the buck."

This dynamic can also help with the overall political economy of distributed energy resource program advocacy. Time-variant pricing program implementation, for example, is often stalled by questions around the likely impacts of such programs on low-income customers, a question that to date has been difficult to answer conclusively given the paucity of load shape data at the local level.

Second, while it raises complicated policy questions around the appropriateness of system-cost socialization, analytics based on individual usage patterns allow for a more accurate assessment of consumers' marginal system costs impact, a key input in economically

Abbreviations: ACS, American Community Survey; AIC, Ameren Illinois; AMI, Advanced Metering Infrastructure; CUB, Citizens Utility Board; ICC, Illinois Commerce Commission; LIHEAP, Low Income Home Energy Assistance Program; PLC, peak load contributions

* Corresponding author.
E-mail address: jzethmayer@citizensutilityboard.org (J. Zethmayr).

[1] See Faruqui et al., 2007

[2] It should be noted that more needs to be done to maximize the consumer benefits of AMI, particularly when it comes to advancing distributed energy resources.

[3] See, e.g., McLoughlin et al., 2015: "A wealth of new data exists for utilities, giving detailed electricity consumption at increased granularity for a large number of customers within the residential sector. The availability of this source of data can potentially be used by utilities to create customized electricity load Profile Classes (PC) and can assist in areas such as: improved load planning and forecasting; Time of Use (ToU) tariff design; electricity settlement; and Demand Side Management (DSM) strategies".

[4] Yu et al., 2018.

https://doi.org/10.1016/j.tej.2019.106643

efficient tariff design.[5],[6],[7]

In this paper, we apply k-means clustering, a machine-learning algorithm, to AMI data of more than 2.5 million customers of Commonwealth Edison (ComEd) and Ameren Illinois (AIC), the two largest electric utilities in Illinois. From our review of the literature, it appears this is the largest dataset yet used in an electric consumer segmentation analysis. This process generates subsets of similar customers based on their average summer load shape.

We then use a logistic regression model to determine the likelihood of a customer in each of the resulting six clusters to reside in locations associated with various demographic indicators, as defined in the most recent American Community Survey (ACS) from the U.S. Census Bureau. We construct demographic profiles for customers who consume electricity according to each cluster's average load shape. We run a separate regression to isolate the load shapes most prevalent in low-income communities. We identify six distinct load profiles, ranging from nearly flat to high-volume, peaky usage, and find that customers with flatter load shapes were substantially more likely to live in urban areas and low-income communities.

2. Theory

This paper relies on anonymous electric consumption data for residential customers of ComEd and Ameren Illinois, captured through smart meters. In 2017, the Illinois Commerce Commission (ICC) approved a plan for Illinois utilities to make customers' smart meter data available to researchers and other third parties, provided the data was kept anonymous.[8] The datasets provided through this plan include daily observations consisting of hourly interval volume readings from individual smart meters, identified by customer class, random ID number, and geographic location.

According to the rules approved by the ICC, an individual customer's data can only be included in the data release if it passes an anonymity screen, to ensure the identity of a customer cannot be determined or reverse engineered. In practice, the screen adopted in the rule requires that a customer's location cannot be provided if there are 15 or fewer customers in the given geographic area, or if they represent 15% or more of that area's load.

Due to the difference in population density in the utilities' service territories (ComEd serves northern Illinois including Chicago, while AIC serves central and southern Illinois), this anonymity screen results in very different levels of geographic specificity in the two datasets. For ComEd customers, we were able to use a dataset consisting of 1.5 million customers, all geo-located at the 9-digit ZIP code level (postal code). This allowed for precise matching of customers to demographic attributes, at the Census Block Group level. Few AIC customers passed the anonymity screen at this level, leaving the majority of observations in that dataset identified at the municipal level, requiring demographic information to be aggregated amongst block groups.

To construct our demographic attributes dataset, we downloaded Census Block Group level data from the 2017 ACS.[9] This dataset includes local information on a variety of categories:

- Age of head of household
- Construction year of residence
- Educational attainment
- Heating fuel
- Household makeup (number of residents, family vs. non-family)
- Home values
- Density (number of households in block group)
- Number of rooms in residence

In order to match postal codes to block groups, this study also relied on a commercially available dataset acquired from Melissa, a geographic data services company.[10] This dataset provided a direct concordance between postal codes and census block groups.[11]

2.1. Cluster analysis

In recent years, researchers have applied new data mining and statistical techniques to characterize consumer profiles according to their usage patterns.[12],[13] From our review of the literature on customer segmentation, we selected the k-means clustering method for this analysis.[14] Half-hourly smart meter datasets from ComEd and AIC were used to cluster residential customers in different groups based on their usage patterns.

K-means is an unsupervised learning algorithm, which assigns individual observations into similar subsets by minimizing the variance between those observations. The algorithm continuously generates random cluster assignments for all observations, to produce an optimal set of assignments. Rather than defining groups beforehand, clustering allows us to identify the organically formed groups within a dataset.

The inputs to the k-means algorithm are the number of clusters k and the dataset itself. The algorithm randomly selects k centroids from the dataset, then iterates between two processes to generate the final result: assignment, where each observation is associated with the nearest centroid and assigned a cluster, and centroid update, where each cluster's centroid is adjusted according to the mean of its assigned observations. These iterations continue until the sum of the Euclidean distances is minimized or the maximum number of iterations is reached.

2.2. Demographic analysis

To compare the demographic characteristics between clusters, we used customers' geographic identifiers, to pair them with block group census data from the annual American Consumer Survey. Block groups are the smallest geographic division for which this census data is publicly reported.

In our review of the literature, we evaluated several methods of associating clusters with demographic characteristics.[15] Of these, we selected multinomial logistic regression to link the clusters with demographic and household characteristics such as age, construction year, educational attainment, and income. This method uses one cluster as a baseline of comparison, and determines the likelihood of a cluster member residing in an area with a particular demographic profile, as compared to the baseline. Each explanatory variable is a binary indicator of whether or not a plurality of residents within a particular block group matches that demographic category.[16] Our model uses 43 independent explanatory variables.

Due to the lower level of geographic specificity in the AIC dataset, this analysis produced few significant results for these customers. This

[5] Chen et al., 1997.

[6] Burger et al., 2019a, "Fair, Equitable, and Efficient Tariffs in the Presence of Distributed Energy Resources".

[7] Burger et al., 2019b, "The Efficiency of Distributional Effects of Alternative Residential Electricity Rate Designs.

[8] See The Big Energy Data Center for information on Illinois' big data story. https://www.citizensutilityboard.org/welcome-big-energy-data-center/.

[9] See American Fact Finder, U.S. Census Bureau.

[10] Melissa Inc., Geo*Code, https://www.melissa.com/direct/reference-data/#3.

[11] For full methodology, see https://www.citizensutilityboard.org/wp-content/uploads/2019/06/ClusterAnalysisFinal.pdf.

[12] McLoughlin et al., 2015.

[13] Figueiredo et al., 2005.

[14] Al-Wakeel and Wu, 2016.

[15] Rhodes et al., 2014 and Beckel et al., 2014.

[16] The full list of variables can be found at https://www.citizensutilityboard.org/wp-content/uploads/2019/06/ClusterAnalysisFinal.pdf.

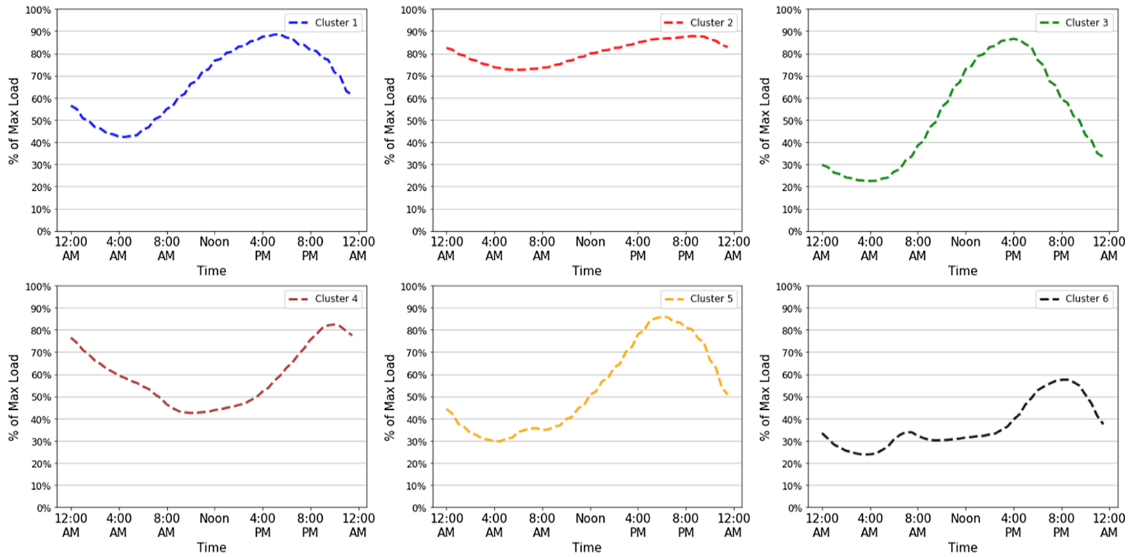

Fig. 1. Average summer weekday usage, in % of max loads.

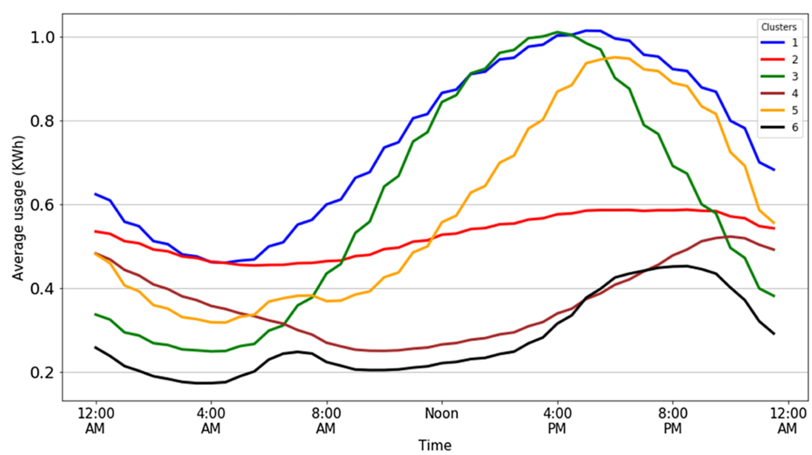

Fig. 2. Average volume usage, in kWh.

portion of the paper includes only results from the ComEd dataset.[17]

2.3. Low-income analysis

To further isolate the effect of lower incomes on electricity consumption, a separate regression was run on the ComEd dataset using only median income. This model used a binary variable designating block groups with median incomes low enough to qualify a family of four for the Low Income Home Energy Assistance Program (LIHEAP), or $37,650.

3. Results

3.1. Cluster load shapes

The clustering algorithm produced six average cluster load shapes, presented here in terms of percentage of maximum load and average volume (Figs. 1–3).[18]

Cluster 1 contains 706,174 customers, representing 25% of ComEd customers and 32% of AIC customers. This cluster has a wide afternoon through evening peak, ranging between 40–85% of maximum load. Their lowest usage is 0.5 kW at 4 a.m., while their highest usage reaches 1 kW. This load shape is the most similar of all clusters to ComEd's system-wide load shape.

Cluster 2 represents 421,782 customers, accounting for 18% of ComEd customers and 15% of AIC customers. This cluster exhibits a very flat load shape, staying between 70–90% of maximum load throughout the day. These customers' hourly volumes range between 0.45 kW at 6 a.m. to a prolonged, flat peak near 0.6 kW from 5 to 9 p.m.

Cluster 3 contains 493,306 customers, representing 18% of ComEd customers and 21% of AIC customers. These customers have the widest difference between base and peak load, ranging from 20 to 85% of maximum load, using 0.25 kW at 4 a.m. and 1 kW at their 4 p.m. peak.

Cluster 4 represents 282,773 customers, or 13% of ComEd customers and just 8% of AIC customers. This cluster exhibits a late evening peak at 10 p.m., using 85% of maximum load on, and a late morning trough at 40% of maximum. Their base load is 0.25 kW, and peak is 0.55 kW.

[17] For AIC demographic analysis, see https://www.citizensutilityboard.org/wp-content/uploads/2019/06/ClusterAnalysisFinal.pdf.

[18] Locations within five miles of the center of one of the 20 most populous Illinois cities outside the Chicago Metropolitan Area were considered exurban,

(*footnote continued*)
locations greater than five miles from those cities were considered rural.

Cluster Composition			Delivery Service			Area			
Cluster	Size	% of Total Data	Residential Single	Residential Multi	Ameren	Chicago	Chicagoland	Exurbs	Rural
1	706,174	27.50%	238,912	140,421	326,841	195,308	76,720	95,204	338,942
2	421,782	16.42%	136,566	133,898	151,318	200,721	35,644	44,011	141,406
3	493,306	19.21%	198,446	78,375	216,485	84,867	67,904	78,485	262,050
4	282,773	11.01%	86,076	111,036	85,661	142,489	26,684	30,995	82,605
5	451,540	17.58%	169,173	88,111	194,256	116,546	54,524	67,648	212,822
6	212,439	8.27%	57,079	102,600	52,760	79,699	28,737	30,790	73,213
Total	2,568,014	100.00%	886,252	654,441	1,027,321	819,630	290,213	347,133	1,111,038

Fig. 3. Cluster Composition.

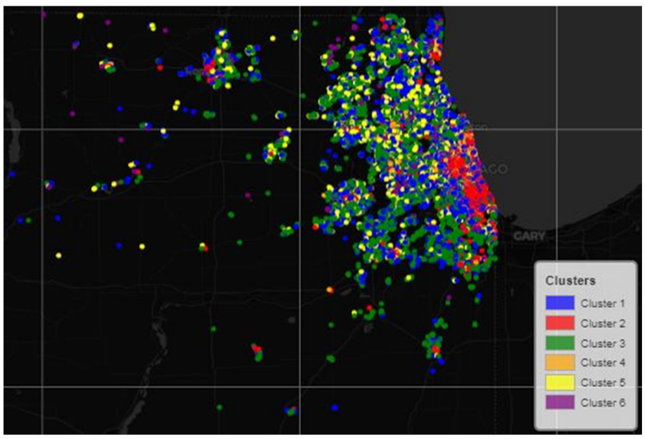

Fig. 4. ComEd service territory.

Cluster 5 contains 451,540 customers, accounting for 17% of ComEd customers and 19% of AIC customers. There is a slight mid-morning uptick in usage, followed by a steep, late-forming peak from late-afternoon through early evening. The volume for this cluster ranges between 0.35 kW at 4 a.m. to 0.95 kW at 6 p.m.

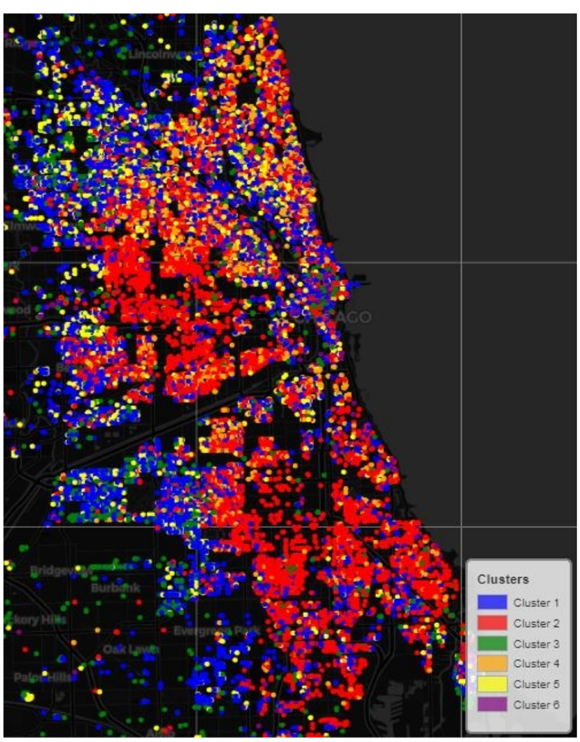

Fig. 5. Chicago.

Cluster 6 contains 212,439 customers, accounting for 10% of ComEd and 5% of AIC customers. This cluster has the lowest usage of all the clusters, ranging from 0.1 kW at 4 a.m. to 0.45 kW at 8 PM. These customers exhibit a bimodal load shape, with a slight morning peak at 7 a.m. and an evening peak at 8 p.m. The percentage range is between 25%–60% of maximum load.

While customers from each cluster appear throughout both the ComEd and Ameren Illinois service territories, there are strong correlations between geography and load shape. The maps below illustrate the prevalent load shapes in locations throughout the ComEd service territory. Each point in these maps represents a postal code, colored to represent the most frequently occurring cluster load shape in that location. They demonstrate the predominance of Clusters 2, 4, and 5 in Chicago and other city centers, and the frequency of Clusters 1 and 3 in suburban and rural areas (Figs. 4–7).

Interestingly, the tendency of Clusters 2 and 5 to appear in central city locations extends to smaller cities, as shown in maps of Rockford and Aurora. Outside of Chicago, significantly fewer locations meet the initial anonymity screen, leading to the relative scarcity of observations in the rest of the service territory.

3.2. Demographic analysis

The regression model produces results in terms of relative likeliness of a cluster member residing in a block group with a particular demographic profile, as compared to Cluster 1. Cluster 1 was used as the seed for the model because its load shape is the closest to ComEd's overall load shape and it is geographically distributed throughout the footprint.

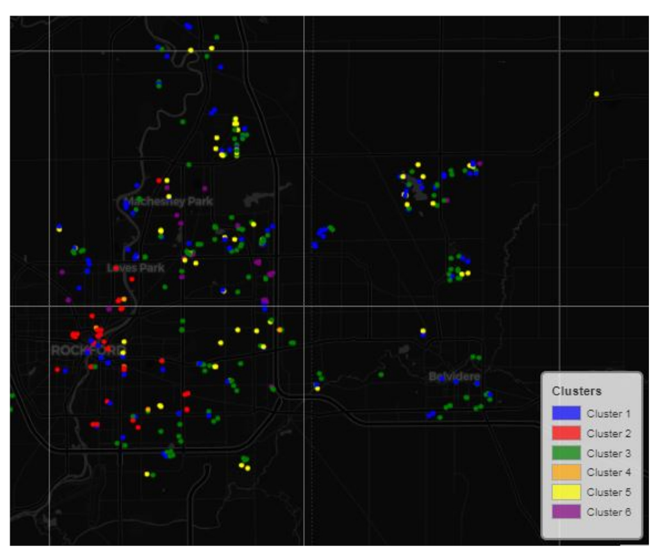

Fig. 6. Rockford.

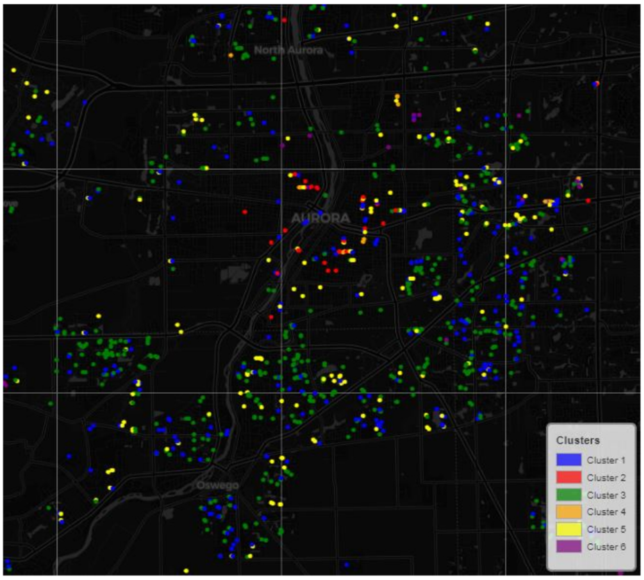

Fig. 7. Aurora.

Results greater than 1 indicate a cluster is more likely than Cluster 1 to be located in an area matching that variable; results less than 1 indicate that cluster is less likely to match the variable.

Cluster 1: **ComEd baseline**

This cluster has the fewest significant demographic indicators, likely because it is the load shape closest to the overall ComEd load. However, we are able to make two inferences about these customers based on the other clusters' results.

Compared to the other clusters, these customers are relatively likely to have high income. The other clusters are all between 0.55 to 0.78 times as likely to earn $150k or more. And while they are distributed throughout the footprint, they are relatively less likely to live in high density areas: Clusters 2, 3, 4, and 6 were between 1.1 and 1.5 times more likely to live in a block group with greater than 660 housing units. In other demographic categories, clusters had significant results both below and above 1, indicating Cluster 1 is relatively "average" in those traits.

Cluster 2: **City flat** (young, urban apartment-dwellers and low-income households)

Cluster 2 appears to contain two customer profiles. These customers are likely to be relatively young, educated, and live in a Chicago apartment. They are also the most likely of all clusters to earn less than $50,000 per year. These factors suggest this load shape pattern is representative of both young professionals and low-income households, which is supported by the map of Chicago; high concentrations of these customers are found in the west and south sides of the city, areas with high levels of poverty, and the affluent north side, which is popular with young professionals.

Cluster 2 customers are nearly 2.5 times more likely than average to live in Chicago, the highest of all clusters. They are 1.6 times more likely to be younger than 33, 1.2 times more likely to be between 33 and 56, and only 0.5 times as likely to be over 56. They are relatively likely to hold a post-graduate degree (1.3 for graduate degree, 1.2 for doctorate), and the least likely to not hold a high school degree (0.8). They are likely to live in older buildings in high density block groups, and 3.7 times more likely to earn less than $50,000. A flat overall load shape supports the likelihood of apartment dwelling without central air conditioning.

Cluster 3: Exurban retirees

These customers are likely to be older, live in exurban or rural areas, and dwell in buildings with five or more rooms built after 1951. This cluster is the most likely of all clusters to be over 56 and the most likely to be rurally located. Those in Cluster 3 are 2.2 times more likely to live in an exurb, 2 times more likely to live in a rural area, and half as likely to live in Chicago. While they are distributed throughout the income range, they are the least educated cluster, as they are the least likely to hold an advanced degree and the most likely to hold less than a college degree. Their load shape, with a high early peak, suggests a larger residence that is occupied in the afternoon.

Cluster 4: City duck curve (younger urban apartment-dwellers)

These customers exhibit very similar demographics to Cluster 2; they are likely to be relatively young (1.9 times likely to be under 33, 0.7 times likely to be over 56), and live in older, smaller residences in Chicago. Interestingly, these customers are both relatively likely to have no high school while also the most likely to hold a post-graduate degree. They most likely earn less than $75,000. With low average volume and a high, late peak, it is likely these customers are more likely than Cluster 2 customers to have all household residents employed.

Cluster 5: Exurban, middle class

Cluster 5 has relatively few significant demographic indicators. These customers are the most likely cluster to hold a bachelor's or associates degree, at 1.2 for an AA and 1.3 for a BA. They are 0.8 times as likely to be under 33.They are unlikely to earn more than $150,000, 0.7 times as likely as the baseline, and more likely to be over 56.

These customers are 0.6 times as likely to live in Chicago, and 1.2 times as likely to live in an exurb. The sharp, later peak and small morning peak suggest low daytime occupancy, supporting a profile of families with working parents.

Cluster 6: Low-volume space heaters

Cluster six customers are likely to reside in smaller residences (less than four rooms) in high-density block groups outside of Chicago; of all the clusters, in fact, they are the least likely to live in Chicago. They are the least likely to live with family members, and the most likely to hold an advanced degree. They are also less likely than average to earn less than $150,000. Finally, these customers are 2.5 times more likely to have electric space heating.[19]

3.3. Low-income analysis

Our analysis of low-income cluster assignments in ComEd showed that low-income households were significantly more likely to exhibit lower overall volumes and flatter load-shapes. The single most likely cluster for low-income households, Cluster 2, also has the flattest usage. Low-income customers were 3.5 times more likely than average to exhibit this usage profile.

The next most likely usage profile for these areas is Cluster 4, exhibiting low volume with a day-time trough. Low-income customers were 2.5 times more likely than average to exhibit this load shape. Customers in both of these clusters are highly likely to live in relatively dense areas of Chicago. All told, more than 50% of ComEd low-income customers have a flat, low-volume usage profile (Figs. 8–9).

To further illustrate the correlation between load shape and income, these maps show the location and local median income for members of each cluster, throughout the service territory as well as within Chicago. These maps indicate postal codes where each cluster is the most prevalent load shape, with each location colored according to the median income level (Figs. 10–11).

The service territory maps illustrate the significantly higher occurrence of clusters 1, 3, and 5 both in areas earning between $61,000 and $84,000, and $85,000 or more. By contrast, clusters 2 and 4 are highly concentrated in areas earning less than $61,000 and less than $43,000. This disparity is particularly salient in Chicago. The second set of maps show Cluster 2 customers highly concentrated in areas earning less than

[19] This is an interesting result and one we will examine further in a follow-up analysis focused on winter load shapes.

Demographics and household characteristics	Clusters									
	Cluster 2		Cluster 3		Cluster 4		Cluster 5		Cluster 6	
	Exp(β)	Std. Error	Exp(β)	Std. Error	Exp(β)	Std. Error	Exp(β)	Std. Error	Exp(β)	Std. Error
LIHEAP	2.472***	0.0077	0.666***	0.0102	1.774***	0.0088	0.702***	0.0102	0.973***	0.0111
Constant	0.637***	0.0027	0.749***	0.0026	0.489***	0.0029	0.699***	0.0026	0.415***	0.0031

Fig. 8. Low-income regression results.

Cluster Composition			Delivery Service		Area			
Cluster	Size	% of Total Data	Residential Single	Residential Multi	Chicago	Chicagoland	Exurbs	Rural
1	30884	20.68%	11341	19543	24037	3574	2440	833
2	48599	32.55%	17185	31414	42367	3141	2362	729
3	15405	10.32%	5477	9928	10020	2838	1910	637
4	26806	17.95%	9095	17711	23880	1652	920	354
5	15152	10.15%	5852	9300	11929	1652	1096	475
6	12469	8.35%	2634	9835	9658	1632	947	232
Total	**149315**	**100.00%**	**51584**	**97731**	**121891**	**14489**	**9675**	**3260**

Fig. 9. Low-Income Cluster Summary.

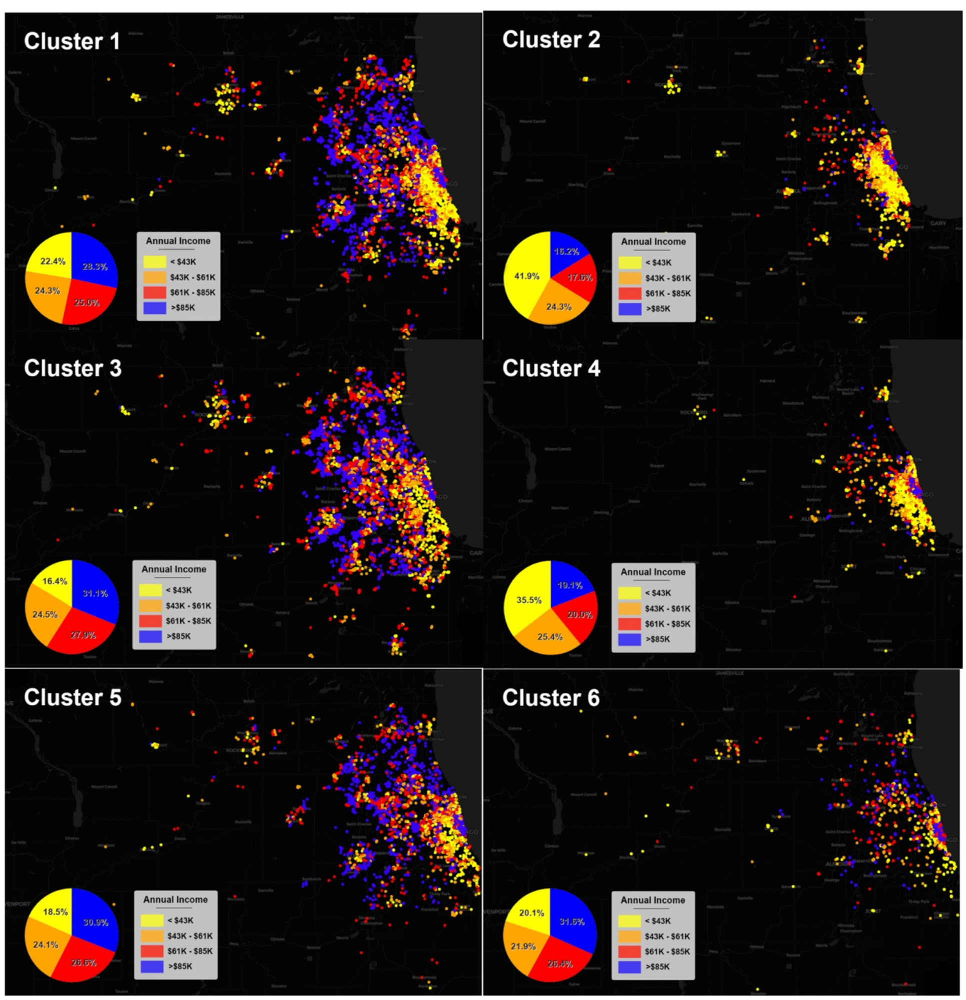

Fig. 10. Income Groups by Cluster.

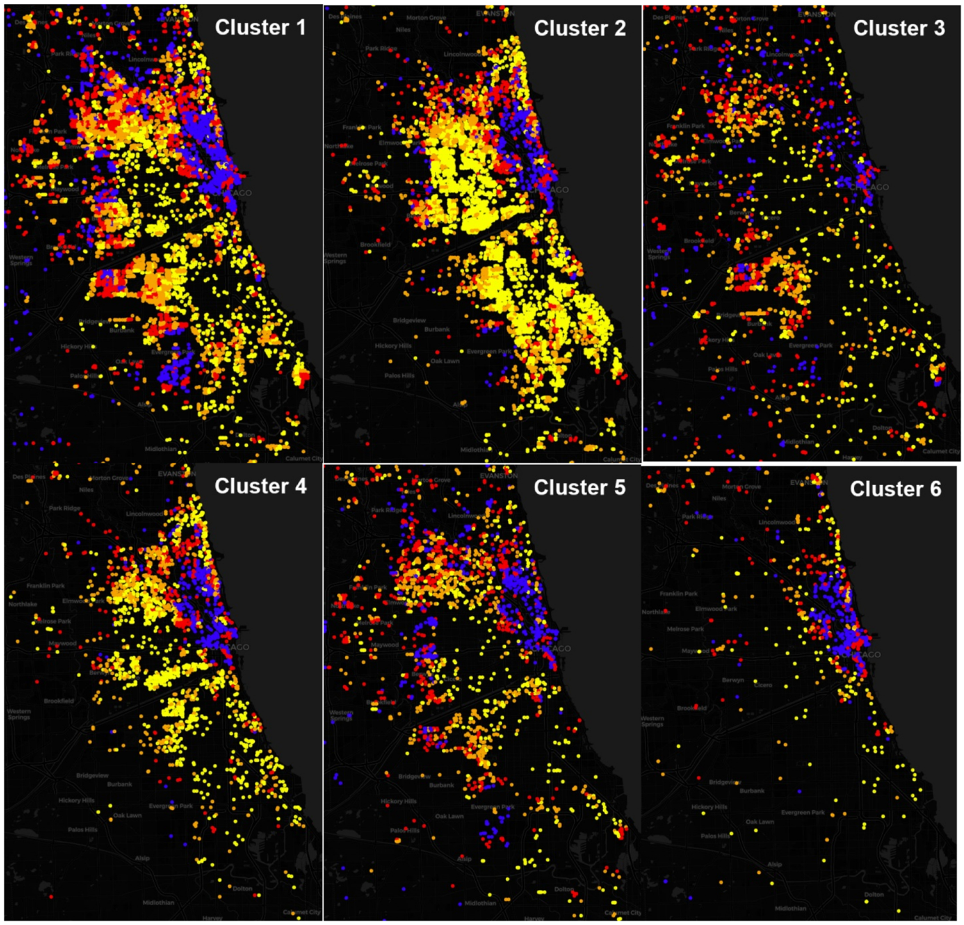

Fig. 11. Income Groups by Cluster, Chicago.

$43,000, on the west and south sides of the city: areas that have long suffered the effects of endemic poverty, racial discrimination, and disinvestment.

4. Conclusion

While previous studies seeking to segment electricity consumers had access to survey data and personal identifiers, this study sought to perform meaningful segmentation on these customers without the help of that information. By clustering anonymous data and applying a logistic regression model to a dataset using only geographically associated Census data, we were able to discover statistically significant results that marked clear demographic delineations between different types of electricity consumers.

This information can be used to improve the effectiveness of energy efficiency programs and dynamic rate designs by helping to target those initiatives at customers whose participation would have the biggest impact on the system, as well as those customers who would benefit the most. It also can facilitate more informed policy analysis, by helping to determine the potential impacts of rate design decisions on vulnerable communities, and to evaluate the equity of cost allocations.

High peak usage is a large driver of system costs, requiring more robust distribution and transmission grids, and higher capacity requirements in areas with capacity markets.[20] From this perspective, customers with high peaks relative to their overall volume likely pay less in their bills than the system costs they actually cause, while customers with flatter load shapes have higher volume relative to their peak usage, leading to them overpaying.

The high correlation between flat usage and lower incomes suggests this cross-subsidization has particularly harmful consequences, considering low-income households already pay a higher proportion of their income on utility bills. This finding should encourage utilities and utility commissions to adopt a wider offering of dynamic rate designs that may more accurately reflect customers' cost of service, reducing this cross-subsidization.[21]

Another conclusion we can draw from these results is the high value proposition of energy efficiency and distributed energy resources in reducing system costs. Programs encouraging energy efficiency adoption and distributed energy resources investment in urban areas are important and beneficial for low-income communities. However, expansion of demand response and price responsive demand programs in suburban and ex-urban areas may have a greater overall impact on system costs, reducing bills for all customers.

As a result of lower-density areas being excluded due to the anonymity screen, our ComEd dataset has a higher proportion of urban customers than the actual population of ComEd customers. One consequence of this, and our inability to do demographic analysis on Ameren customers, is under-sampling of rural low-income customers.

[20] This is a particular problem in ComEd territory, where PJM capacity market charges constitute 21% of the average residential customers' bill. See Monitoring Analytics, LLC. Q1 State of the Market Report for PJM.

[21] The implications for rate design and cost of service regulation is a thorny area that warrants further investigation. In future studies, we intend to examine more of the details and implications.

One avenue for further research would be to perform this analysis using ComEd's ZIP code dataset, which contains significantly more customers at a lower level of geographic granularity. This would likely increase our subset of rural low-income communities.

Finally, a crucial piece of usage data missing from our dataset is customers' peak load contributions (PLCs). With customer PLCs, we could perform more detailed cost of service estimates and rate design evaluations, with a higher degree of confidence. Utility commissions with data access policies such as Illinois' should consider requiring PLCs in the usage data, which can be done while still preserving individual customer anonymity.

Funding

A grant from Energy Innovation helped fund the research.

Declaration of Competing Interest

None.

Appendix A. Regression Results

Variables	Clusters									
	Cluster 2		Cluster 3		Cluster 4	Cluster 5			Cluster 6	
	Exp(β)	Std. Error	Exp(β)	Std. Error	Exp(β)	Std. Error	Exp(β)	Std. Error	Exp(β)	Std. Error
age_less_than_33	1.565***	0.083	1.259	0.315	1.946***	0.134	0.822**	0.095	0.952	0.118
age_33_56	1.162*	0.082	1.221	0.314	1.202	0.133	0.897	0.093	0.654***	0.117
age_more_than_56	0.539***	0.113	1.950**	0.319	0.656**	0.17	0.806*	0.121	0.546***	0.145
built_before_1951	1.767***	0.081	0.648	0.316	1.946***	0.132	0.803**	0.092	0.884	0.124
built_1951_1999	0.690***	0.078	2.062**	0.314	0.927	0.128	0.87	0.088	0.768**	0.118
built_after_1999	0.805*	0.094	2.245**	0.316	0.85	0.14	0.851*	0.096	0.501***	0.121
less_than_high_school	0.791***	0.048	1.048	0.053	1.256***	0.07	1.073	0.057	0.866	0.105
some_college_no_degree	0.904**	0.046	1	0.045	0.596***	0.066	0.838***	0.051	0.561***	0.096
associate_degree	0.899*	0.057	0.951	0.056	0.633***	0.094	1.158**	0.06	0.688***	0.118
bachelor_degree	0.753***	0.059	0.852***	0.05	1.002	0.079	1.260***	0.055	0.991	0.089
graduate_or_prof_school_degree	1.291***	0.065	0.636***	0.055	1.997***	0.082	0.984	0.057	1.439***	0.086
PhD	1.249***	0.071	0.807**	0.092	1.259***	0.078	0.929	0.079	1.157*	0.08
electric_heating	0.698***	0.096	1.097	0.083	0.721***	0.122	0.792**	0.101	2.515***	0.116
gas_heating	0.74***	0.099	1.226**	0.09	0.731**	0.124	1.046	0.108	0.756**	0.116
family_2_person	1.097	0.06	0.880**	0.057	1.146	0.085	0.851**	0.063	0.931	0.112
family_3_5	0.886**	0.053	0.985	0.052	0.959	0.076	1.015	0.058	0.562***	0.107
family_more_than_5	0.882***	0.047	0.913*	0.053	0.978	0.066	0.890**	0.056	0.453***	0.125
nonfamily_less_than_3	0.9	0.189	0.729*	0.19	0.664	0.25	0.743	0.205	1.709	0.408
nonfamily_3_5	1.024	0.191	0.639**	0.195	0.805	0.252	0.741	0.208	1.561	0.408
nonfamily_more_than_5	0.477	0.651	1.063	0.23	0.009	5.762	0.809	0.312	0.001	11.457
value_less_than_10700	0.24	10.099	71.483	5.228	0.73	14.595	0.569	12.395	0.847	19.288
value_10700_104999	0.929	0.125	0.744*	0.162	0.532***	0.168	0.794	0.187	0.947	0.197
value_105000_159999	0.523***	0.106	0.808	0.15	0.428***	0.124	0.744*	0.167	0.518***	0.166
value_more_than_250000	0.343***	0.112	0.851	0.154	0.327***	0.13	0.852	0.171	0.662**	0.168
h_units_less_than_361	1.082	0.066	1.136	0.294	1.024	0.109	0.832**	0.078	0.561***	0.139
h_units_361_475	0.869**	0.0631	1.238	0.237	0.942	0.104	0.906	0.073	0.572***	0.115
h_units_476_660	0.939	0.061	1.385	0.236	1.118	0.101	0.847**	0.07	0.800**	0.102
h_units_more_than_660	1.110*	0.06	1.540*	0.236	1.423***	0.099	0.931	0.069	1.323***	0.094
income_less_than_50000	3.726***	0.066	1.528**	0.194	2.553***	0.096	0.941	0.073	0.854	0.097
income_50000_74999	1.304***	0.062	1.459**	0.191	1.605***	0.092	1.045	0.063	0.825**	0.085
income_75000_99999	0.681***	0.068	1.503**	0.191	0.837*	0.098	0.932	0.062	0.751***	0.085
income_100000_149999	0.614***	0.077	1.386*	0.193	0.806**	0.103	0.983	0.07	0.823**	0.091
income_more_than_149999	0.549***	0.144	0.646**	0.212	0.555***	0.18	0.660***	0.104	0.781*	0.139
rooms_less_than_4	0.903	0.104	0.916	0.157	0.957	0.106	1.019	0.145	1.349***	0.103
rooms_4_5	0.728***	0.081	1.088	0.087	0.521***	0.094	1.237**	0.097	0.941	0.094
rooms_more_than_5	0.609***	0.096	1.482***	0.094	0.161***	1.141	1.166	0.105	0.516***	0.129
Occupied	0.804	0.434	0.07	2206	0.283	0.799	2.244	0.515	2.815*	0.576
Vacant	1.421	0.433	0.067	2.206	0.326	0.798	1.758	0.513	2.328	0.573
Chicago	2.482***	0.067	0.497***	0.237	1.777***	0.105	0.555***	0.072	0.406***	0.099
Chicagoland	0.977	0.069	1.35	0.236	0.857	0.111	0.784***	0.071	0.749***	0.096
Exurb	0.670***	92	2.165***	0.237	1.016	0.131	1.234***	0.076	1.091	0.103
Rural	0.604***	0.087	2.064***	0.236	0.991	0.127	1.108	0.073	1.026	0.099
Constant	0.981	0.212	2.999	0.939	1.535	0.365	0.595**	0.251	0.340***	0.319

Significance Levels: ***0.01 **0.05 *0.1.

Appendix B. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.tej.2019.106643.

References

Al-Wakeel, A., Wu, J., 2016. K-means based cluster analysis of residential smart meter measurements. Energy Procedia Vol. 88, 754–760. https://doi.org/10.1016/j.egypro.2016.06.066.

Beckel, C., Sadamori, L., Staake, T., Santini, S., 2014. Revealing household characteristics from smart meter data. Energy 78, 397–410. https://doi.org/10.1016/j.energy.2014.

10.025.
Big Energy Data Center, 2019. Citizens Utility Board. (Accessed 22 May 2019. https://www.citizensutilityboard.org/welcome-big-energy-data-center/.
Burger, S., Knittel, C., Perez-Arriaga, I., Schneider, I., vom Scheidt, F., 2019a. The Efficiency of Distributional Effects of Alternative Residential Electricity Rate Designs. NBER Working Paper No. 25570. http://www.nber.org/papers/w25570.
Burger, S., Schneider, I., Botterud, A., Perez-Arriaga, I., 2019b. Fair, Equitable, and Efficient Tariffs in the Presence of Distributed Energy Resources. Consumers, Prosumers, Prosumagers: How Service Innovations Will Disrupt the Utility Business Model. https://doi.org/10.1016/B978-0-12-816835-6.00008-5.
Chen, C., Hwang, J., Huang, C., 1997. Application of load survey systems to proper tariff design. IEEE Trans. Power Syst. 12 (4). https://doi.org/10.1109/59.627886.
Faruqui, A., Hledik, R., Newell, S., Pfeifenberger, J., 2007. The power of five percent. Electr. J. 20 (8) (Accessed 22 May 2019. https://www.smartgrid.gov/files/The_Power_Five_Percent_How_Dynamic_Pricing_Can_Save_35_Billi_200709.pdf.
Figueiredo, V., Rodrigues, F., Vale, Z., Gouveia, J., 2005. An electric energy consumer characterization framework based om data mining techniques. IEEE Trans. Power Syst. 20 (2). https://doi.org/10.1109/TPWRS.2005.846234.
McLoughlin, F., Duffy, A., Conlon, M., 2015. A clustering approach to domestic electricity load profile characterization using smart meter data. Appl. Energy 141, 190–199. (Accessed 22 May 2019. https://arrow.dit.ie/cgi/viewcontent.cgi?article=1063&context=dubenart.
Melissa Inc, 2019. Geo*Code, Melissa Direct. https://www.melissa.com/direct/reference-data/#3.
Monitoring Analytics, 2018. LLC. Q1 State of the Market Report for PJM. . http://www.monitoringanalytics.com/reports/PJM_State_of_the_Market/2018/2018q1-som-pjm.pdf.
Rhodes, J., Cole, W., Upshaw, C., Edgar, T., Webber, M., 2014. Clustering analysis of residential electricity demand profiles. Appl. Energy 135, 461–471. https://doi.org/10.1016/j.apenergy.2014.08.111.
US Census Bureau, 2017. ACS Estimates, American FactFinder. https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml.
Yu, Y., Liu, G., Zhu, W., Wang, F., Shu, B., Zhang, K., Astier, N., Rajagopal, R., 2018. Good consumer or bad consumer: economic information revealed from demand profiles. IEEE Trans. Smart Grid 9 (3), 2347–2358. https://doi.org/10.1109/TSG.2017.2662684.

Jeff Zethmayr conducts energy policy and economic research, with a focus on rate design for regulated utilities. Mr. Zethmayr's previous work includes expert testimony and analysis for utility rate cases, and research on the bill impacts of demand-based delivery rates. He holds a Masters of Public Administration from Columbia University's School of International and Public Affairs and a Bachelor of Arts from Macalester College.

Ramandeep Singh Makhija is a Data Scientist at the Citizens Utility Board, where he investigates utility customer behavior through advanced statistical analysis. He holds a Master's degree in Industrial Engineering from the University at Buffalo with specialization in Operations Research, and a Bachelor's degree in Mechanical Engineering from the University of Pune.