

SMARTER IMPACT ASSESSMENTS

How smart meter data can
inform regulators and lead to
smarter impact assessments

2020

SMARTER IMPACT ASSESSMENTS

HOW SMART METER DATA CAN INFORM REGULATORS AND LEAD TO SMARTER IMPACT ASSESSMENTS

REGULATORS WANT TO KNOW HOW POLICIES WILL AFFECT DIFFERENT GROUPS

Over the last few years, regulators and governments have been increasingly concerned about the distributional impact of their policies. Last year we published a [series of articles](#) about the what “fairness” might mean for different sectors.

Given this focus, Ofgem, the British energy regulator, needs to take into account how future policies will affect different types of customers. In May Ofgem consulted¹ on a framework for these types of evaluations. It proposes to assess the impact of policies on a series of customer groups. These include users it is statutorily obliged to protect, such as those of pensionable age, as well as 13 customer “archetypes” based on sociodemographic and behavioural characteristics.

One factor that distinguishes electricity customers from one another is the time of day when they use it. Although most households currently pay the same for their power no matter when they consume it, time-of-use tariffs (where the amount paid changes over the year or the day) may become more widely available in future.² Policies that affect these prices could therefore have a very different impact on customers depending on when they consume electricity.

Ofgem’s framework envisages adding to its archetypes consumption profile data showing when power is used. But, as we show in this article, standard consumer archetypes do not map neatly to distinct patterns of usage. In cases where the

EXEC SUMMARY

With fairness high up the agenda of regulators, Ofgem has been exploring how it can best capture the distributional effects of its policies. Smart meter consumption data can help make these assessments more useful. However, making sense of the vast amount of data requires both machine-learning techniques and careful judgment.

¹ https://www.ofgem.gov.uk/system/files/docs/2020/05/assessing_the_distributional_impacts_of_economic_regulation_1.pdf

² <http://www.frontier-economics.com/uk/en/news-and-articles/articles/article-i6106-time-to-pick-up-pace-of-dynamic-electricity-pricing/>

distributional impact of a policy is influenced by when electricity is consumed, these archetypes may work best alongside a classification based on the profiles of consumption themselves. As we explain, clustering techniques provide a way of transforming a huge dataset of customer profiles into an intuitive classification.

CLUSTERING HELPS MAKE SENSE OF THE WEALTH OF LOAD PROFILE DATA

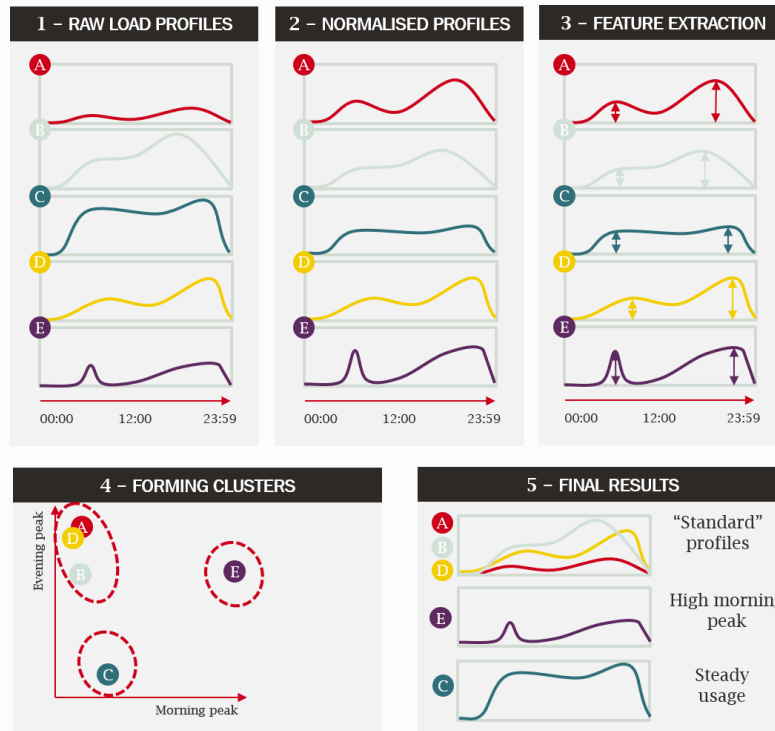
Our analysis uses data gathered as part of Northern [PowerGrid's Customer-Led Network Revolution \(CLNR\) project](#). Among the many datasets made available by this project is one showing anonymised half-hourly electricity consumption for around 9,000 households. It also includes archetypes taken from Experian's Mosaic system, one of the most commonly used customer classification methods. These archetypes differ from the ones Ofgem uses (there is no consumption dataset available that includes Ofgem's archetypes), but they follow the same principle, grouping together customers with similar sociodemographic characteristics.

For a single year, a customer will have 17,520 datapoints, describing each half-hour of the year. This adds up to over 150m datapoints just in this single dataset – far too many to make sense of.³ What is needed is a way to reduce the complexity of the data, classifying each customer in a way that distils the essence of their consumption profile while removing much of the detail.

We have used a clustering algorithm to do this. Clustering is an example of an “unsupervised machine-learning” technique, whereby an algorithm sorts a dataset into categories of similar observations, without being told ahead of time what those categories should be. The box below illustrates – in a very simplified form – what we have done.

³ If we were carrying out the analysis across all 26m electricity meters in Britain, the dataset might have 450bn observations.

HOW HAVE WE CLUSTERED PROFILES?



- We started with the raw data. Here, we have illustrated a day’s worth of data for five customers (in reality we had a year of data for almost 9,000 customers). The height of the line shows the amount of electricity being consumed at each point in the day.
- Whether a time-of-use tariff makes someone better or worse off will depend on the relative pattern of consumption (e.g. between the afternoon and the evening). The absolute level of consumption is less relevant. We have therefore normalised the profiles so they all have the same total energy consumption.
- At this point, each customer is still described by a datapoint for each half-hour, which is far too much detail. A process of feature extraction is needed to get to the “essence” of each customer. In the illustration, we have shown a simple method (our analysis is more complex) where we have defined each customer by the height of the morning and evening peak.
- Now that each user is defined by a small number of values, we can find which groups of customers resemble each other. In this example we have done so by eye – there is clearly a cluster of three similar customers (A, B, and D), while the other two (C and E) form their own distinct clusters. For our analysis we used a technique that can automatically form these clusters.

Looking at the final clusters in the illustration, we can see that they describe three intuitive types of customers. Those with “standard” consumption profiles (with a smaller morning peak and larger evening peak), those with a particularly high morning peak and those with very steady consumption through the day. This shows how this type of analysis can boil a large dataset down to a manageable number of clusters from which to draw conclusions.

From the description above, you will see that it is not simply a matter of feeding a large amount of raw data into a machine-learning algorithm that can “do the rest” without human intervention. The algorithm can base its decisions only on the data that is given to it, so the job of the researcher is to use their knowledge of the wider context to guide the algorithm to a better result.

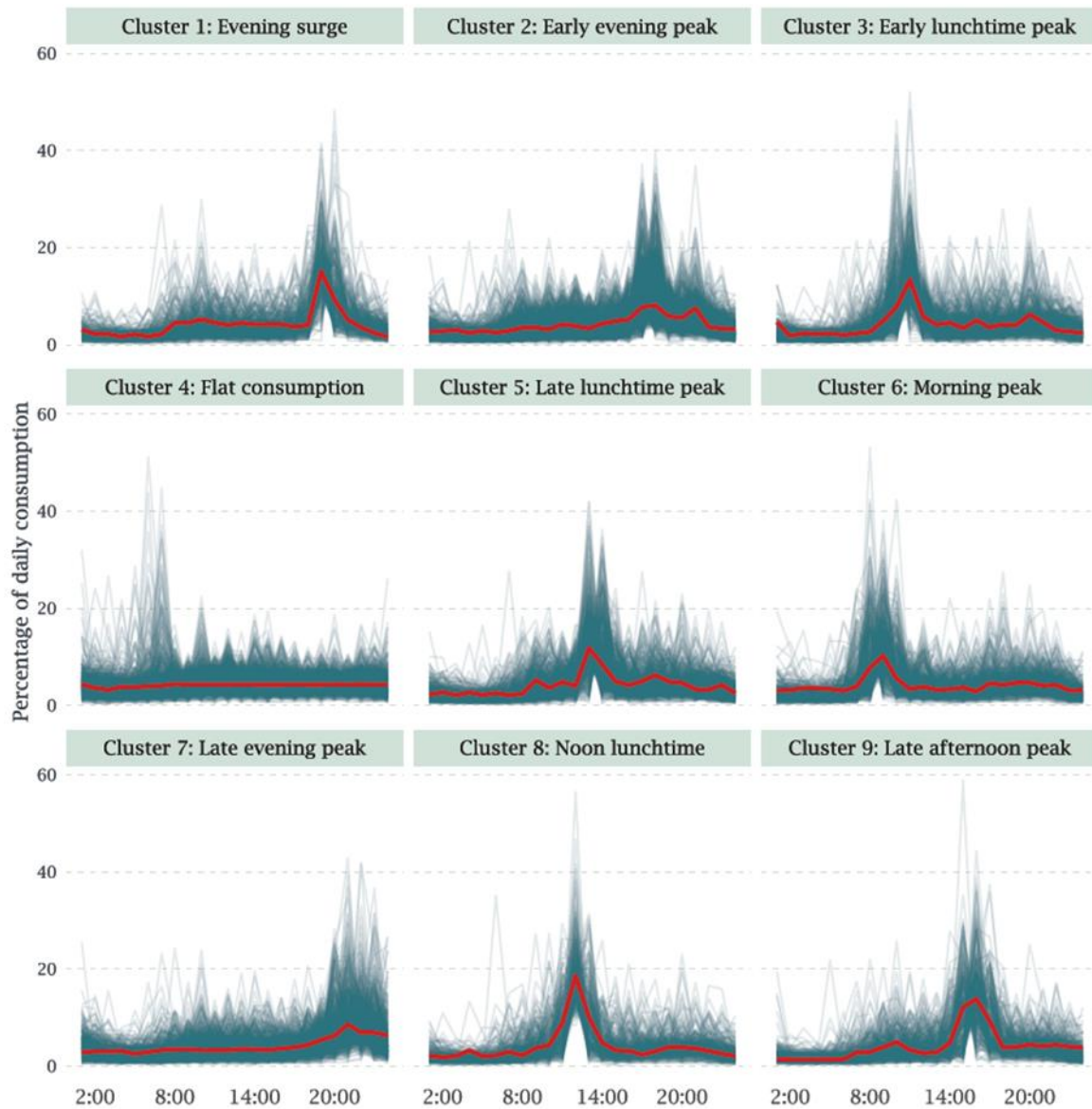
For example, as we knew that what matters for a distributional assessment is a customer’s relative use of electricity over the day rather than their absolute consumption, we applied a specific type of normalisation. If we left the process to the computer and omitted this step, the final choice of clusters would probably have been heavily determined by overall energy usage, which is irrelevant.

We have carried out this type of analysis on the full⁴ CLNR dataset and produced nine clusters, each with their own pattern of electricity usage. For example, some clusters show a significant spike in consumption at midday (perhaps people using an electric cooker for lunch) while others peak in the evening. **Error! Reference source not found.** illustrates each cluster. The red line shows the typical load profile of customers in this cluster. The blue lines show the load profiles of all the other customers that are in the same cluster.

We have created a tool that lets you explore how your own consumption compares to these clusters. [Click here to open the tool.](#)

⁴ For the purpose of this article we have reported results based on a single summer weekday.

FIGURE 1 LOAD PROFILE CLUSTERS FOR A SUMMER WEEKDAY

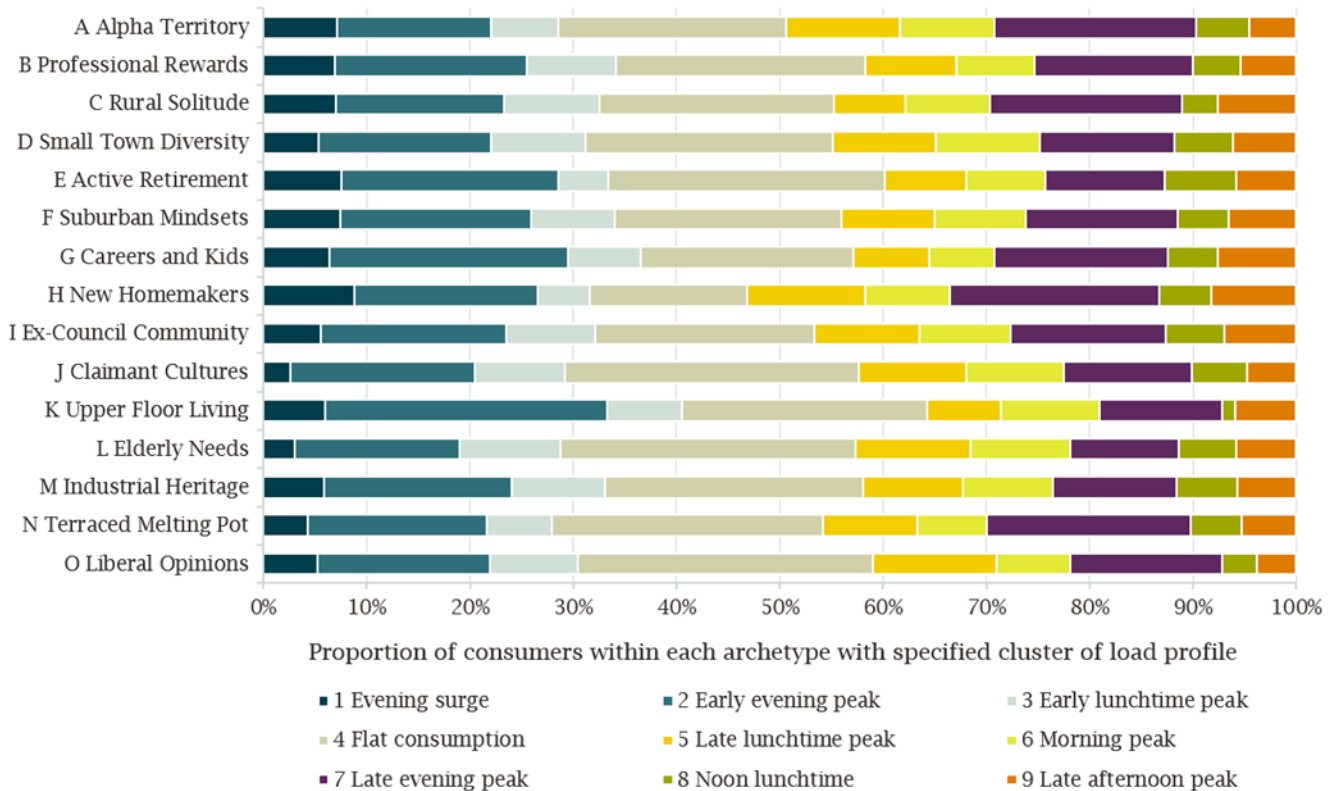


Source: Frontier analysis of CLNR data

STANDARD ARCHETYPES DO NOT CORRELATE STRONGLY WITH CONSUMPTION PROFILES

Now we have these clusters, we can check how well correlated they are with the archetypes. **Error! Reference source not found.** shows the proportion of load profiles of each cluster within each of the Mosaic archetypes.

FIGURE 2 LOAD PROFILE BREAKDOWN WITHIN EACH ARCHETYPE



Source: Frontier analysis of CLNR data

There is only a very weak correlation, and in general the customer archetypes are not a good predictor of consumption patterns. This is consistent with previous research, which has tended to show that while high-level demographic measures such as wealth may be correlated with total energy usage, they are not good predictors of consumption profiles. As you will have found in the interactive tool, electricity consumption is a reflection of highly individual patterns of behaviour, which will not always be a function of sociodemographic characteristics.

WHAT DOES THIS MEAN FOR DISTRIBUTIONAL ANALYSIS?

We have shown how general purpose archetypes are not well correlated with patterns of electricity usage. Ofgem’s archetypes are more directly linked to patterns of energy consumption than the Mosaic groups⁵ and may therefore be somewhat more correlated with consumption profiles than the Mosaic archetypes seen here. However, given the sheer variety of customers covered by each archetype, this correlation is still likely to be very weak.

⁵ For example, the archetypes split out customers with electric heating (who may be more likely to consume electricity overnight for storage heating).

Having a consistent set of such archetypes is an invaluable tool for distributional analyses as they allow different policies to be compared on a like-for-like basis. However, for policies where the main distributional impacts depend on consumption patterns (e.g. time-of-use tariffs), additional tools will be required.

Techniques like clustering offer a way to build a set of “consumption profile archetypes” that could complement the standard sociodemographic archetypes. This type of analysis is potentially useful for identifying the distribution of winners and losers in terms of their consumption patterns – whether any adverse impacts of a policy are relatively mild, or if there are pockets of customers for whom the effect will be much greater. However, a number of issues need to be borne in mind.

WE NEED TO UNDERSTAND WHO IS IN EACH CLUSTER

A distributional analysis ideally needs to identify who these groups of customers are. The clustering analysis might tell us, for example, that a policy is disadvantageous for “people who use a lot of electricity on weekday lunchtimes”. But to assess this finding in its full context, we need to understand whether these customers are already disadvantaged or vulnerable in some way.

ACCESS TO DATA IS NEEDED

Carrying out this type of analysis requires access to a large dataset of representative smart meter data showing customers’ hour-by-hour consumption.

Given the sensitivity of such data, it is (rightly) subject to strict controls on who can access it and for what purposes. However, such restrictions may hinder regulators and governments from fully understanding the effects of their regulations.⁶

Anonymised samples like the CLNR dataset can avoid some of the privacy concerns. But such samples may not be large enough to capture relatively unusual patterns of consumption which could still affect many tens or hundreds of thousands of people and might be important for the assessment of some policies. They may also fail to capture the richness of sociodemographic data needed to understand the types of customer within each cluster.

THERE MAY NOT BE ONE TYPE OF CLUSTERING THAT WORKS FOR ALL POLICIES

As we have shown, a machine-learning approach such as clustering still requires the researcher to make decisions based on their knowledge of the overall context. One key decision involves the “feature extraction” step that streamlines a load profile with many thousands of observations into a manageable series of values. Much data is lost in this step, so the art of feature extraction lies in preserving the aspects of the consumption profile that are most relevant for the analysis being carried out.

⁶ These issues are being explored by the Smart Meter Data Public Interest Advisory Group - <https://www.smartenergydatapiag.org.uk/>

Different policies may have different types of distributional effect. For some it could be the size of the peak that is more important; for others, the overall shape. So rather than one set of “fits all” clusters, it may be most appropriate to build the clusters with the policy in mind.

AUTHORS

ALEX WHITTAKER
Manager

SUMAIYA RAHMAN
Analyst

WANT TO KNOW MORE?

WWW.FRONTIER-ECONOMICS.COM

HELLO@FRONTIER-ECONOMICS.COM

+44 (0) 207 031 7000