CATEGORISATION OF ELECTRICITY CUSTOMERS BASED UPON THEIR DEMAND PATTERNS

Magnus LINDÈN
SWECO – Sweden
Magnus.Lindén@sweco.se

Jakob HELBRINK
SWECO – Sweden
Jakob.Helbrink@sweco.se

Martin NILSSON
SWECO - Sweden
Martin.a.Nilsson@sweco.se

Daniel Pogosjan
SWECO – Sweden
Daniel.pogosjan@sweco.se

Jonathan Ridenour
SWECO – Sweden
Jonathan.Ridenour@sweco.se

Andrea Badano
SWECO - Sweden
andrea.badano@sweco.se

ABSTRACT

Historically, customer categorisation has usually implied categorization by fuse size, Industry-code, type of residence, estimated annual consumption and alternatively by actual consumption pattern. The latter has primarily been applied for larger consumers. Thanks to the hourly metering reform in Sweden, several calendar years of consumptions data with hourly resolution are now available, which presents new methods of categorization of customers. The logic behind categorizing customers is multi-layered, but often connected to retail, pricing, balancing, forecast, wholesale and portfolio management or enabling DR for balancing services.

By using a data-driven methodology taking into account the customer's hourly actual consumption, the insight of the individual as well as portfolio of customers will increase, in turn leading to fewer assumption having to be made in connection to pricing/rate-setting, development of template-profiles for subgroups of consumers as well as the overarching prognosis-work. Retailers will be able to develop more tailored deals and charge customers according to their hourly consumption at the same time as it will be easier to identify new smart grid-components such as energy storages, micro generation and customers that adapt their consumption to the electricity price (demand response).

The combination of a more complex cost-structure for network operators and the goal of a cost-reflective pricing presents a challenge. Some costs are more complex than others and are therefore harder to allocate to a certain customer, which is why they are profiled within a customer category. As the differentiation of costs between customer categories should be cost-reflective (different types of customers drives varying costs) it is important that the categorization is correct to start with.

The categorisation conducted in this study is carried out using actual hourly metering data for customers from five different concessional areas, which all in all sums to roughly 150 000 individual customers. The methodology of the workgroup has been to "cluster" customers into a given number of categories based on attributes derived from their consumption pattern. The attributes of the customers have been reduced to a number of KPI's (Key Performance Indicators) which are calculated from their hourly consumption profile. Following this, the customers are grouped based on how their KPI-values resemble each other. This have been done with the method K-means. As a single KPI can't reflect all nuances in the customers' consumption, a total of 9 KPI's have been analysed. This is expected to result in a categorisation where each category contains customers with similar consumption patterns.

Categorisation with several KPI's at the same time comes with pitfalls which were deemed as hard to handle during the project at the same time as the result becomes less intuitive to analyse. Therefore, the categorization has been conducted with one KPI at a time. Several analyses have been conducted, among them; the ability of the KPI's to reflect costs for the DSO (Distribution system operator) or electricity retailer, the similarity/unsimilarity between categorization with different KPI's, the effect of regional factors and geography on the categorization as well as how robust the categorization is over several calendar years. To link the results to present methods for categorizing customers comparisons have been made to fuse size and industry-code. This has resulted in a number of conclusions:

- *Categorization by simultaneously using more KPI's doesn't necessarily give a better result. Normalization and weighting of the KPI's are required (otherwise the KPI with larger values will dominate the results) at the same time as the result becomes harder to interpret.*
- *But with a sequential categorization with single KPI's it is possible to avoid above named pitfalls and gain enhanced knowledge of the customer's attributes.*
- *If the KPI is meant to be used for pricing, it is crucial that it contains a time-dimension.*
- *The effects of geography and regional differences are marginal with the datasets that have been analysed (which covers the areas for five DSO*

situated at different latitudes and with different degree of urbanisation).

- *Most of the KPI's display low correlation to each other.*
- *The results from the categorization differs significantly from year to year for the majority of the KPI's. This means one of two things; (1) the robustness is low or (2) customers have changed their consumption pattern.*
- *Customers within the same fuse size and SNI-code are spread out into several different groups regardless of which KPI's is used for categorizing them. This shows that "traditional" methods doesn't reflect the customer's consumption patterns very well.*

Furthermore, an overarching trend that becomes visible is that many customers are relatively alike but also that a significant share are more distinctive. Today's methodology for defining profiles customers, which is based on the assumption that a customer category is fairly homogenous, can therefore be questioned.

In the future, the customer's perceived "cost" for allowing their consumption being controlled by a third party, with the aim of lowering costs by adjusting to price signals, will decrease. Because it is a fairly reasonable assumption that, with more widespread hard-/software, more customers will see this as an obvious choice. If the rate-models at the same time are cost-reflective "everybody" will get the same incentive, which means that an average customer will consume electric energy "optimally" given their electricity use and prerequisites. In that case, knowledge regarding how the price signal affects the customer's consumption becomes key in predicting loads in the electricity market, designing networks and to create an optimal rate. All of this aims to enable a more efficient market and use of energy.

The results from this study can for instance be used for deeper analysis regarding particular customer-behaviours such as flexibility and new loads (electric cars, solar panels) as well as continued studies of how customer categorization can act as a complement present methods of developing template consumption profiles and rate-setting.

CATEGORISATION OF ELECTRICITY CUSTOMERS BASED UPON THEIR DEMAND PATTERNS

INTRODUCTION

DSOs in Sweden have traditionally classified domestic and smaller non-domestic customers for tariffing purposes according to their contracted peak capacity, while suppliers have classified their customers according to "type", for instance apartment, detached housing, or commercial. Previous studies indicates that this is not optimal, as demand patterns may vary substantially across groups of customers that have the same contracted capacity, or that belong to the same "category".

The rise of non-programmable electricity generation, often connected to electricity distribution networks, suggests that electricity prices will be more volatile in the future, and that networks will require more complex management at all voltage levels. Current classifications lead to electricity network and supply tariffs that are seldom aligned, and that do not properly signal the value of energy, capacity and network services. Attempts to better signalling often result in very complex tariff structures that do not suit all customers. Properly signalling the value of energy, capacity and network services is essential to achieve efficient markets. Proper signalling, however, will require a deeper understanding of electricity consumers' consumption patterns and alternative means of classifying customers.

Swedish DSOs have in recent years installed meters that record consumption in each hour. The resulting massive data streams provide the opportunity to derive detailed analytical information about consumption patterns. We make use of actual hourly data for approximately 150,000 customers covering a number of years (1 to 3 years depending on the DSO) and different regions that we have obtained from a number of Swedish DSOs. Using a clustering algorithm, we classify domestic and small non-domestic electricity customers into several categories based upon different aspects of their consumption patterns.

AIM

The aim of the study is to develop and investigate a methodology for categorising electricity customers based upon historic consumption patterns. The results from the categorisation are also analysed.

METHOD

On an overarching level, consumption-data provided by the DSO is inputted into the big data tool Lavastorm, where validation of the metering series as well as extraction of KPIs takes place. The categorisation is conducted by an external script, and afterwards the results are again inputted to Lavastorm (or MS Excel) for analysis.

To be able to see if there are clear patterns and correlations between different types of customers based upon on a certain KPI, the method *K-means* was used. Customers are grouped based upon their value in one or more KPIs (two in the figure below) to one of *K* different categories. A K-value of 10 has been used in this report. Each cluster is described by their *centriod*, which is defined by average

value of all observations in the given cluster. The algorithm optimizes the placement of the centroids to achieve a least distance possible for all customers to a centroid.

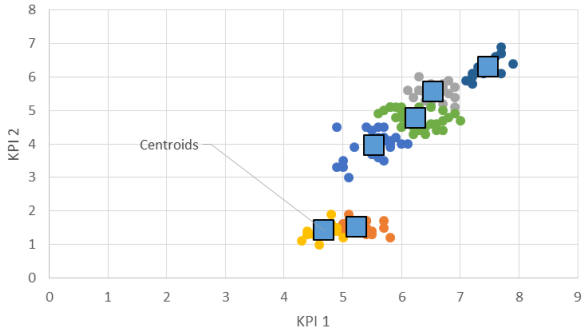

Figure 1. K-means-clustering based upon two KPIs generating six clusters. The squares mark the center of each cluster.

The complexity of the problem increases with one dimension for each added KPI, yielding a result which is harder to analyse. Because of this, the choice was made to categorise customers using just one KPI at a time, ensuring that the causation can be understood. While normalisation is a possible method for solving this problem, is it hard to conduct this in a fair and non-discriminating manner at the same time as maintaining transparency.

Data

Five different DSOs have contributed with hourly consumption data spanning over 1-3 calendar years and various number of customers. All in all, the analysis is based upon roughly 2 billion rows of consumption data.

Table 1. Summary of the consumption data provided by the five DSOs.

DSO	Years	No. Customers in analysis
Mälarenergi Elnät	2012,2014	53 000
Ulricehamns Energi	2013-2015	7 500
Umeå Energi Elnät	2014-2015	19 000 (2014) 49 000 (2015)
Göteborg Energi	2015	10 000
Luleå Energi	2015	32 000

KEY PERFORMANCE INDICATORS (KPIS)

To ensure an efficient energy-use it is essential to present the customer with the correct price signal. A crucial component in realizing the customers demand flexibility is therefore cost reflective pricing. To further understand how tariffs ought to be designed to offer the most appropriate incentive, knowledge of existing customer segments as well as their characteristics and drivers for demand are needed. These characteristics can, more or less, be approximated using KPIs.

Nine KPIs based upon on the hourly consumption profile of customers have been investigated in this study and are listed in the table below.

Table 2. Investigated KPIs. If a cell I shaded it implies that the KPI can reflect costs for the DSO and retailer.

KPI	Comment
Percentage of consumption during peak hours	Peak hours are defined as weekdays between 06-22 o'clock, november-march
Correlation, full year	Versus the aggregated load in the DSO network
Correlation, 10 highest hours	Versus the aggregated load in the DSO network
Correlation, 50 highest hours	Versus the aggregated load in the DSO network
Correlation, electricity price	NordPool Elspot for corresponding price areas
Capacity-factor	$\frac{Annual\ peak\ power}{Annual\ energy}$
Loadfactor	$\frac{\sum_{days}(\frac{avg.power}{peak\ power})}{days}$
Level-of-use	$\frac{\sum_{days}(\frac{peak\ power}{minimum\ power})}{days}$
Temperature sensitivity	Regression of consumption versus temperature

All correlations are calculated using the following equation.

$$Correlation\ (X,Y) = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2\sum(y-\bar{y})^2}}$$

COST REFLECTIVE KPIS

All KPIs are more or less suited for different uses, but all give a better understanding of the characteristics of the customer. One area of particular interest is their ability to describe actual costs for the electricity retailer and DSO. If so, the KPI can be used as a way to charging customers for the service or to separate the into different price categories. If two or more KPIs are used for categorising customers, costs can also provide an objective way of weighting the significance of the KPIs for comparison.

Costs for the DSO is approximated using the tariff design displayed in the picture below. The fixed costs are constant over the period, while the charge for capacity is directly proportional to the aggregated load in the grid. The energy fee is proportional to the square of the aggregated load in the grid, symbolizing that the load-losses increase with the square of the power.

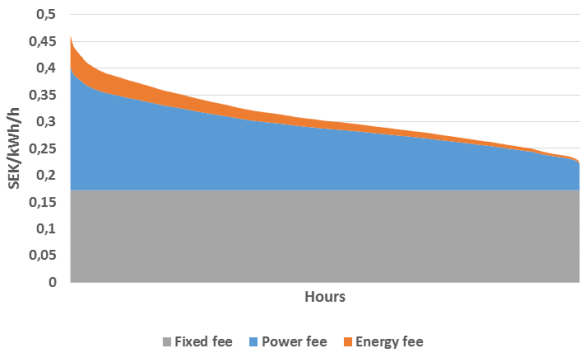

Figure 2. Design of the applied DSO tariff. Each hour is priced individually based upon the aggregated load in the network, with the most congested hour also being the most expensive.

Costs for the electricity he electricity retailer are approximated using the Elspot price, as seen in the picture below.

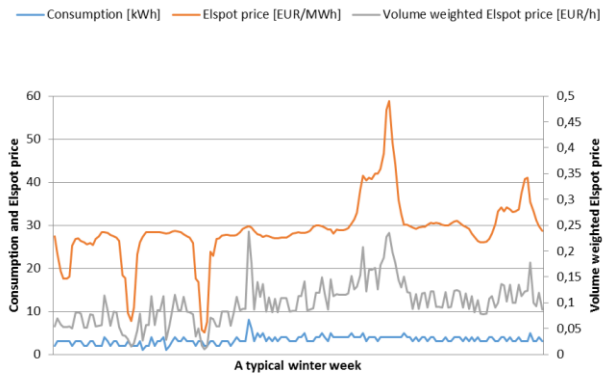

Figure 3. Volume weighted spot price a random customer during a winterweek.

Correlation between KPI and costs exists if a clear covariation between the two parameters is visible, e.g. like the picture below. As can be seen, Customers with a higher KPI-value generally also get a higher unit cost with the DSO tariff.

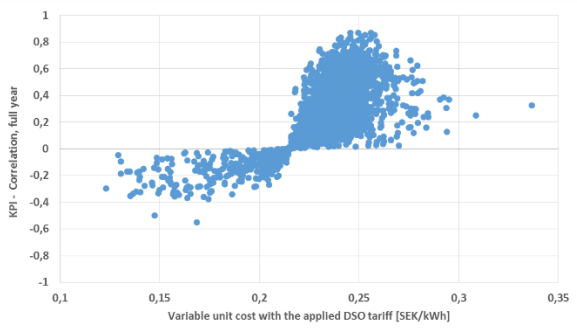

Figure 4. Example of a KPI which displays a clear correlation to costs for the DSO. Each dot represents a customer with their KPI-value on the y-axis and costs for the DSO on the x-axis.

After analysis it is concluded that a temporal dimension is needed for the KPI to be able to reflect costs, i.e. the shaded KPIs in Table 2.

ACCURACY OF TRADITIONAL METHODS OF CATEGORISATION

Segmentation of customers based upon their fuse size is a common method of categorising electricity customers into different price-groups in Sweden. But the fuse size does not give a perfect picture of the costs a customer entails for the DSO or retailer. As can be seen in the picture below, customers of the same fuse size are spread out in many different categories (each row represents a fuse category).

Luleå & Correlation, electricity price

Location of the Centroid	-0,03	0,14	0,21	0,28	0,35	0,44	0,53	0,62	0,71	0,84	
	Category/cluster										No. Customers per fuse category
Fuse	4	0	7	2	5	9	1	6	3	8	
250 A	0%	0%	0%	0%	4%	2%	11%	15%	32%	36%	47
200 A	0%	0%	0%	3%	0%	10%	2%	7%	22%	56%	59
160 A	0%	0%	1%	1%	4%	4%	9%	12%	33%	38%	129
125 A	0%	2%	1%	1%	3%	4%	14%	14%	31%	30%	142
100 A	0%	1%	1%	1%	1%	6%	19%	22%	22%	27%	139
80 A	1%	1%	0%	1%	3%	4%	4%	20%	42%	23%	137
63 A	1%	1%	1%	5%	37%	4%	9%	12%	20%	9%	478
50 A	1%	1%	2%	3%	4%	7%	18%	19%	31%	14%	213
35 A	1%	1%	1%	3%	9%	7%	18%	23%	29%	8%	423
25 A	1%	1%	1%	2%	4%	8%	14%	26%	38%	4%	1376
20 A	0%	1%	1%	2%	4%	8%	12%	29%	40%	3%	2792
16 A	3%	12%	22%	19%	14%	11%	9%	6%	3%	0%	25836
N/A	2%	2%	0%	1%	3%	1%	5%	11%	30%	45%	185
No. Customers per cluster	812	3248	5861	5112	4073	3289	2988	3059	2890	626	31956

Figure 5. Distribution of customers according to fuse size (rows) into 10 different clusters (columns) for the KPI Correlation, electricity price. The value of the centroid is seen at the top and each cell describes the percentage of customers in that cluster with the corresponding fuse size.

Large customers seem to have a high correlation to the electricity price (demands more electricity as the Elspot price gets higher) and implying larger costs for the retailer. As a stark contrast, some 16 A customers (3%) appears to consume less volumes as the price goes up. It can be argued that these 3% of 16 A customers constitute a very different type of customer than many of the other 16 A customers.

The claim of fuse size being an inadequate method of categorizing customers can be further validated by comparing the unit cost of customers within the same fuse segment. It is particularly visible for the fuse category 35 Amps where the spread is between different cost intervals is large.

Luleå - Unit cost vs. fuse size

	Unit cost [SEK/kWh]										
Fuse	0.14<x<0.19	0.19<x<0.24	0.24<x<0.29	0.29<x<0.34	0.34<x<0.39	0.39<x<0.44	0.44<x<0.49	0.49<x<0.54	0.54<x<0.59	>0.59	No. Customers per fuse category
250 A	72%	23%	4%	0%	0%	0%	0%	0%	0%	0%	47
200 A	73%	15%	8%	0%	0%	2%	2%	0%	0%	0%	59
160 A	63%	30%	2%	2%	2%	0%	0%	1%	0%	1%	129
125 A	43%	44%	8%	1%	1%	1%	1%	0%	1%	1%	142
100 A	32%	50%	14%	3%	0%	1%	0%	0%	0%	0%	139
80 A	28%	56%	7%	4%	2%	1%	0%	0%	0%	1%	137
63 A	5%	44%	20%	11%	5%	4%	2%	2%	1%	5%	478
50 A	3%	54%	20%	7%	4%	3%	2%	2%	0%	5%	213
35 A	2%	38%	26%	13%	9%	4%	1%	1%	2%	3%	423
25 A	0%	19%	39%	19%	7%	5%	2%	1%	1%	4%	1376
20 A	0%	7%	42%	25%	11%	5%	3%	2%	1%	3%	2792
16 A	0%	0%	2%	4%	6%	6%	6%	5%	4%	66%	25836
N/A	94%	5%	0%	1%	0%	0%	0%	0%	0%	1%	185
No. Customers per unit cost-interval	524	1279	2431	2203	1948	1739	1571	1263	1185	17278	31956

Figure 6. Distribution of unit cost for customers within a fuse category.

ROBUSTNESS OF THE CATEGORISATION

For all networks where two or more calendar years of data are available a robustness analysis has been conducted. These generally display that the similarity of the categorisations between two different years are low (<60%) for all KPIs. This implies that the composition customers within each cluster changes from one year to another to a large extent. E.g. if the robustness is 25 % it means that (between two years) a given customer end up i a category where 25% of the customers remain the same.

The exception to this is when a clustering is heavily affected by outliers, resulting in many small clusters and one large with the vast majority of the customers. This effect is visible for the KPI Capacity-factor in the picture below.

Mälarenergi 2012 & 2014 (51 300 customers)

	Correlation 2012 vs 2014
Percentage of consumption during peak hours	44%
Correlation, full year	44%
Correlation, 10 highest hours	30%
Correlation, 50 highest hours	40%
Correlation, electricity price	20%
Capacity-factor	99%
Loadfactor	43%
Level-of-use	58%
Temperature sensitivity	40%

Figure 7. Robustness for the KPIs. 100% robustness means that all customers end up in a category consisting of the same customers in year 1 as in year 2. A value below that indicates that customers switch categories from one year to another.

USING THE RESULT – AN EXAMPLE

By using a sequential approach, it is possible to combine the information provided from several KPIs without having to deal with the challenges of an overly complex analysis. This section provides an example of how this can be done to identify a given type of customer.

Suppose a DSO (Ulricehamn in this example) is aiming to reduce the costs for subscribed power towards the feeding grid. Customers with a larger (positive) correlation to the network load are more prone to consume larger amounts during congested hours and therefore more prioritized to target for demand flexibility measures. The larger the customer (large fuse), the greater the contribution. It is therefore of interest to identify large customers with a high correlation to the network load. These customers (62 in total) are circled in the picture below.

Ulricehamn 2015 & Correlation, full year

Location of the centroid	-0,24	-0,06	0,03	0,09	0,18	0,30	0,40	0,49	0,57	0,67	
	Category/cluster										No. Customers per fuse
Fuse	9	3	8	1	6	5	2	4	0	7	category
200 A	0%	50%	0%	0%	0%	0%	0%	0%	0%	50%	2
160 A	6%	0%	29%	12%	0%	6%	0%	24%	18%	6%	17
125 A	0%	0%	22%	6%	0%	17%	11%	17%	17%	11%	18
100 A	0%	0%	33%	4%	0%	13%	0%	13%	17%	21%	24
80 A	2%	0%	13%	0%	2%	8%	12%	17%	17%	29%	52
63 A	6%	0%	8%	3%	6%	21%	13%	25%	16%	8%	63
50 A	8%	1%	7%	5%	5%	16%	16%	11%	12%	18%	74
35 A	10%	1%	23%	6%	4%	8%	9%	13%	15%	13%	160
25 A	9%	0%	16%	4%	6%	17%	8%	10%	22%	7%	411
20 A	2%	0%	14%	3%	3%	23%	6%	14%	28%	7%	1725
16 A	12%	3%	3%	26%	17%	10%	8%	7%	8%	6%	4801
0 A	0%	2%	14%	0%	3%	22%	8%	26%	6%	17%	65
No. Customers per cluster	686	136	546	1349	928	988	567	695	1015	502	7412

Figure 8. The circled area denotes large customers with a consumption with high correlation to the aggregated network load.

But out of these 62 customers, not all will be able to shift load. But customers with a high Level-of-use (indicating a large volatility in the load) might be more capable than the others. By identifying large customers with a high Level-of-use and combining these with the previous selection 45 customers remain, which should be prioritized by the DSO as changes for these customers might be possible and will also yield large results.

Ulricehamn 2015 & Level-of-use

Location of the centroid	0	1	2	3	6	10	19	30	47	97	
	Category/cluster										No. Customers per fuse
Fuse	2	0	7	5	8	4	1	6	9	3	category
200 A	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	2
160 A	6%	0%	0%	6%	12%	35%	6%	35%	0%	0%	17
125 A	17%	17%	0%	0%	6%	17%	0%	28%	0%	17%	18
100 A	0%	8%	0%	0%	8%	38%	4%	29%	0%	13%	24
80 A	2%	13%	2%	2%	6%	27%	2%	35%	2%	10%	52
63 A	0%	14%	5%	0%	2%	35%	13%	24%	0%	8%	63
50 A	0%	14%	7%	0%	4%	22%	20%	26%	0%	8%	74
35 A	0%	16%	5%	0%	1%	31%	28%	14%	0%	6%	160
25 A	0%	35%	9%	0%	0%	21%	30%	4%	0%	1%	411
20 A	0%	46%	22%	0%	0%	14%	17%	1%	0%	0%	1725
16 A	0%	12%	13%	0%	0%	3%	73%	0%	0%	0%	4801
0 A	2%	15%	0%	5%	5%	35%	2%	23%	0%	14%	65
No. customers per cluster	6	1593	1036	5	18	589	3975	145	1	44	7412

Figure 9. The circled area denotes large customers with a high Level-of-use.