



Background

At the AUG sub-Committee meeting on 30th September 2021 we took an action to present some detail on the input data we use relating to thefts at Supply Meter Points in the EUC band 01NI. These data are part of our master theft dataset which is consolidated and validated from the following industry datasets:

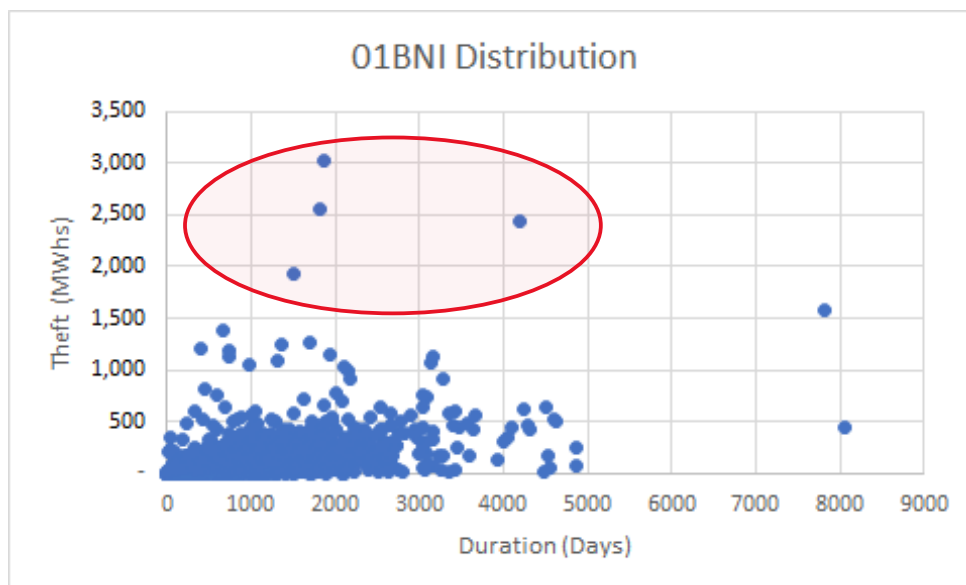
Theft Data	A report of the thefts from Smart and Traditional meters provided by a sub-set of EUK members	EUK
TRAS Theft Information	The data outcome file from TRAS, verified and enhanced by the CDSP with meter type data	Electralink/ CDSP (via CDSP)
TOG Theft Information	Details of theft provided to Xoserve within CMS	CDSP

This short paper presents the distribution of datapoints within the dataset visually, to show the grouping and datapoints which might be considered as outliers.

We have also undertaken some further analysis to understand whether the 01NI EUC band is particularly impacted by such outliers.

01NI Data Distribution

The 01NI theft dataset comprises 1788 datapoints. The plot below shows these datapoints distributed according to duration and size of reported theft, covering all theft that has taken place within the last ten years at Supply Meter Points currently in the 01NI EUC band.





The plot indicates a handful of reported thefts for which the size of theft was several times larger than most of the dataset (red circle). Whilst these might be considered outliers, their scale is plausible in our view.

Outliers in duration are of limited interest, unless they are also unusually large.

Outlier comparison to other EUC bands

To understand whether the 01NI EUC band is disproportionately impacted by potential outliers in the dataset, we compared its data to the same theft data used to allocate theft UIG to other EUC bands. Specifically, we removed the top 1% of thefts (by size) from the dataset for each EUC band. For 01NI, that is the largest 18 recorded thefts.

We recalculated the theft proportions allocated to traditionally metered sites and compared these numbers with the actual outcome using full datasets.

Our hypothesis was that a redistribution between EUC bands would show the varying impact of a small number of large thefts in one or more EUC bands.

The results in the table to the right show that if we were to remove the top 1% of thefts by size, there would be a redistribution of UIG allocation between EUC bands. Notably, however, the theft proportion for 01NI would increase, suggesting that the impact of the largest 1% of thefts is greater in EUC bands other than 01NI.

	All Thefts	Highest 1% of Sites Taken from Total
1ND	35%	34%
1NI	22%	23%
1PD	22%	24%
1PI	0%	0%
2ND	2%	2%
2NI	6%	7%
2PD	0%	0%
2PI	0%	0%
3-8	13%	9%

Based on this analysis, we see no justification in seeking to identify or eliminate outliers as part of our methodology.

If you have any questions about this analysis or outcomes, please contact us at auge@engage-consulting.co.uk.