Anomaly detection results STL method

This document presents the results of the Notebook entitled 'Anomaly detection 3', where we perform anomaly detection on the dataset data_clean.xlsx using: Seasonal trend decomposition using LOESS (STL). We are going to study each meterId as a separate time series, decompose it to its seasonal and trend components and extract the residue. The points where the residue is too big (more than 3 standard deviations away from the mean) suggest an anomaly.

- The percentage of total anomalies found in the dataset is : 1,41%

- The detailed results :

mediaType	percentage of anomalies	number of meters concerned	Main locations concerned
Electricity	1,38%	204	India kalwa (31,40%) Austria (26,46%) Switzerland Zug (15,15%) …
Heating	1,38%	13	Austria(100%)
ColdWater	1,47%	35	Austria (56,07%) Switzerland Zug (18,32%) Germany (14,13%)
WS_Brunnenwasser	1,84%	10	Austria (100%)
Electricity1	2,46%	1	Austria (100%)
Electricity2	1,75%	1	Austria (100%)
Power	0,3%	10	Germany (96,43%) Austria (3,57%)
WS_Blindstrom	1,95%	1	Austria (100%)
ElectricityGenEmergency	1,1%	7	India BL (48%) India kalwa (44%) India GU (8%)
ElectricityGenPV	1,41%	5	India kalwa (87,5%) Switzerland SH (12,5%)
NGas	1,62%	11	United State DA (24,84%) United State HU (16,56%) United State HE, United State BE , United State PO , United State FH , United State SQ (~8%) United State JS (4,45%)

GeneralElectricity	1,03%	1	Switzerland SH (100%)
DistrictHeating	1,18%	10	Germany (80,18%) Denmark (18,86%) Netherland (0,94%)
Cooling	2,58%	8	Germany (100%)
ElectricityGenCHP	0,00%*	-	-
HeatGenCHP	0,00%*	-	-
IntervalGas	2,31%	3	China (85,71%) India kalwa (14,28%)
ColdGenerated	3,18%	1	Switzerland SH (100%)
ElectricitySupplyPV	2,17%	2	India kalwa (100%)
Diesel	2,08%	2	India BL (80%) India kalwa (20%)

Observations

- The results given by this approach have some similarities with the results given by Z-score and iForest in the first approach, however this approach tends to detect more anomalies, which is logical because in this case we focus on each meterId separately which results in more anomalies than if we we study all the data of a specific mediaType at once