Analyzing Unoccupied Setback Effectiveness and Prevalence via K-Means Clustering

Jay Herron
https://github.com/NeedleInAJayStack/RTEM-Hackathon_2022

May 30, 2022

1 Intro

Reducing energy use in buildings is expensive. Despite advances in fault-detection and diagnostic programs, achieving energy savings still requires manually evaluating issues to determine importance and making physical or virtual changes to rectify them. Because of this, better prioritization systems are needed.

Historically, energy reduction analytics can be separated into two categories:

1. **Top-down**: Industry standard metrics like energy use intensity are used to compare buildings to their peers and grade their efficiency. While these are very useful for determining large energy users, they do little to suggest actions for energy reduction within a given building.

2. **Bottom-up**: Equipment-level analytics identify discrete operational issues, and include specific actions to take to resolve the issue. However, due to data availability issues and complex system interactions, energy savings estimates, when even provided, are unreliable.

There is an opportunity for a middle-of-the-road approach that uses high level metrics to analyze common low-level operational issues, and provides prioritization along with clear implementation guidelines.

This document proposes a set of key performance indicators (KPIs) and an analysis framework to identify and prioritize energy reduction through unoccupied setbacks. Nearly every building can achieve significant energy reductions by basic changes to their unoccupied operation, typically by modifying zone air temperature setpoints, minimum discharge air flow setpoints, or lighting based on occupancy. It is possible to implement these changes on nearly all controls equipment and the results are extremely cost-effective.

2 Approach

This framework requires whole-building energy use at hourly frequencies or faster. Whole-building energy usage is then cleaned by removing meter rollover artifacts, aligning time-stamps to a consistent frequency, and interpolating missing values. Next, the values are dis-aggregated if usage is represented by a monotonically increasing counter, and finally the history is passed through an outlier filter. After this process, we are left with clean, consistent energy usage data for the whole building.

Next, the cleaned data is split into weekly periods and each is passed through a k-means clustering algorithm[1] that identifies 2 nodes: a high-usage cluster and a low-usage cluster. This process provides a mapping between each historical observation and whether it belongs in the high or low cluster, as well as average values for the high and low clusters. We interpret the high-usage period to be the occupied time in a building, and the low-usage period to be the unoccupied time.

From these results, we can compute two meaningful weekly KPIs:

1. **Unoccupied Turndown Factor**: The high-usage cluster value divided by the low-usage cluster value. Scaled from 0 to 1, lower values indicate more effective unoccupied setbacks. Lower is better from an energy reduction standpoint.

2. **Occupied Duration Factor**: The amount of time the building was in a high-usage state, divided by the total time in the KPI period. Scaled from 0 to 1, lower values indicate more prevalent unoccupied setbacks, in terms of time. Lower is better from an energy reduction standpoint.

This simple approach offers several distinct advantages. First, the data requirements are extremely low; only a single historical datastream for total energy use is required. Of the 230 buildings in the RTEM dataset, 99 have *Electric Consumption* points, making it the 5th most prevalent point type behind *Virtual*, *Zone Temperature*, *Outside Air Temperature*, and *Status*. Second, it is widely applicable. The analysis process is not impacted by the type of building, the pattern of occupancy throughout the week, or the type of equipment within the building. Since no assumptions are made on the time of occupancy, it functions as well on irregularly occupied buildings (like community centers or theaters) as on regularly occupied buildings (like commercial office buildings or retail). Finally, it delivers actionable results. Even the most primitive control systems offer features for modifying building operation according to occupancy.

3 Usage

These KPIs alone do not scale according to energy usage, so they are best used as part of a top-down Energy Use Intensity-based analysis. For example, if comparing buildings within a portfolio, the EUI could be used to determine the worst-performing facilities, and then the KPIs above could be used to determine whether these facilities would benefit from improving the unoccupied operation.

The different KPIs suggest different actions:

- A large Unoccupied Turndown Factor suggests that unoccupied operation should be investigated. When unoccupied, zone air temperature setpoint deadbands should widen significantly, minimum discharge airflow setpoints should be set to 0, and lights should be turned off. This KPI may also suggest that there are areas of the building that never enter an unoccupied state.

[1] `https://en.wikipedia.org/wiki/K-means_clustering`

- A large Occupied Duration Factor suggests that occupancy scheduling and detection systems should be investigated. Of course, unoccupied energy saving measures must conform to each building's unique occupancy patterns. However, if occupancy is based on a schedule, that schedule should be checked to ensure that it matches actual occupancy patterns. If occupancy is detected using a sensor, the sensors should be validated for correct operation.

While a KPI value on a particular building during a particular week is meaningful, KPIs are typically most useful when they are compared between entities or over time. A building with a large KPI value compared to its peers, or with a KPI value that increases over time may provide context on what constitutes a "large" KPI value.

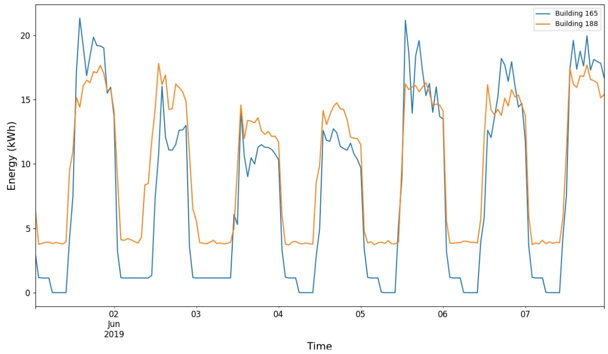
4 Results

Building ID	Type	Square Footage	Start Date	End Date	Point Count	Low Cluster Avg	High Cluster Avg	Unoccupied Turndown Factor Avg	Occupied Duration Factor Avg
100	Multifamily	153,750	2019-06-17	2020-10-26	8	137.66	174.04	0.861	0.567
105	Food/Beverage	56,000	2017-11-20	2019-09-30	34	2,772.69	6,974.23	0.416	0.328
108			2019-10-28	2020-05-04	6	11.75	15.63	0.755	0.415
112	Commercial Office	210,000	2017-09-11	2019-03-11	34	25.64	81.99	0.489	0.475
121	Commercial Office	236,067	2017-11-27	2018-05-28	35	244.62	591.74	0.416	0.326
122	Food/Beverage	56,000	2017-11-27	2019-09-30	38	1,901.78	6,605.24	0.324	0.388
123	Commercial Retail	330,000	2017-09-11	2018-09-10	45	61.12	80.22	0.769	0.371
125	Commercial Retail	8,500	2019-03-11	2020-01-06	3	1.30	3.38	0.409	0.377
127	Multifamily	789,079	2018-06-25	2018-12-31	319	21.55	28.69	0.867	0.615
132	Commercial Retail	198,057	2018-08-20	2019-08-19	16	0.20	0.83	0.272	0.497
135	Commercial Retail	200,000	2018-12-31	2019-12-23	5	12.85	15.78	0.854	0.544
137	Commercial Office	221,704	2016-12-26	2019-12-09	217	33.76	76.47	0.463	0.385
139	Food/Beverage	9,000	2019-04-29	2020-02-24	5	5.02	12.61	0.401	0.855
146	Commercial Office	281,755	2019-02-11	2019-12-30	363	369.00	746.05	0.500	0.343
147	Commercial Retail	198,057	2018-07-02	2019-07-01	18	0.37	1.88	0.270	0.362
151	Commercial Retail	198,057	2018-08-20	2018-08-20	13	0.09	0.87	0.105	0.733
155	Commercial Retail	144,000	2018-08-27	2019-08-26	24	0.78	2.80	0.268	0.533
157	Food/Beverage	56,000	2017-12-04	2019-09-30	35	2,133.13	6,906.57	0.332	0.368
158	Food/Beverage	9,000	2019-04-29	2020-02-24	5	5.46	6.74	0.812	0.532
164	Food/Beverage	56,000	2017-12-04	2019-09-30	38	1,676.93	4,564.45	0.497	0.459
165	Commercial Retail	144,000	2018-09-17	2019-09-16	23	0.45	3.05	0.147	0.526
166	Commercial Retail	198,057	2018-07-02	2019-07-01	19	0.30	2.19	0.192	0.475
167	Commercial Retail	144,000	2018-10-08	2019-10-07	21	0.26	0.98	0.274	0.262
170	Multifamily	789,079	2018-08-27	2018-08-27	6	2.01	2.29	0.877	0.986
172	Commercial Retail	198,057	2018-07-16	2019-07-15	14	0.26	1.80	0.215	0.492
173	Commercial Retail	198,057	2018-08-20	2019-02-18	12	0.58	1.83	0.427	0.423
175	Commercial Retail	198,057	2018-07-02	2019-07-01	19	1.17	1.96	0.658	0.495
176	Commercial Retail	150,000	2018-12-31	2019-12-23	5	33.73	40.35	0.829	0.422
177	Commercial Retail	144,000	2018-09-17	2019-09-16	38	0.58	4.27	0.136	0.488
179	Food/Beverage	56,000	2017-11-06	2019-09-30	36	2,261.94	5,714.80	0.419	0.379
187	Commercial Retail	198,057	2018-07-16	2019-07-15	24	0.42	1.93	0.263	0.557
188	Commercial Retail	144,000	2018-09-03	2019-09-02	22	1.16	2.88	0.424	0.550
191	Commercial Retail	198,057	2018-07-02	2019-07-01	19	0.38	1.92	0.248	0.373
194	Commercial Office	1,700,000	2016-12-26	2018-12-10	16	841.38	1,253.40	0.674	0.486
205	Food/Beverage	8,500	2019-04-29	2020-02-24	7	5.49	6.72	0.814	0.512
206	Multifamily	789,079	2018-06-25	2018-12-03	7	21.19	26.24	0.852	0.539
210	Commercial Retail	198,057	2018-07-30	2019-07-29	28	0.14	0.70	0.246	0.409
217	Commercial Office	229,154	2019-12-09	2021-12-06	14	31.67	85.35	0.491	0.515
220	Commercial Retail	198,057	2018-07-16	2019-07-15	24	0.20	2.13	0.093	0.458
225	Commercial Retail	198,057	2018-07-02	2019-07-01	24	0.31	0.98	0.311	0.411
226	Commercial Retail	144,000	2018-09-17	2019-09-16	31	0.42	1.58	0.270	0.344
236	Commercial Retail	144,000	2018-09-03	2019-09-09	28	0.42	1.59	0.339	0.522
248	Multifamily	127,000	2019-04-08	2019-10-07	34	58.97	78.77	0.747	0.438
249	Commercial Retail	120,000	2018-12-31	2020-06-08	6	12.32	15.01	0.848	0.414
250	Multifamily	120,721	2019-02-25	2019-08-26	3	105.21	123.03	0.855	0.345
252	Commercial Retail	198,057	2018-08-20	2019-08-19	20	0.40	2.37	0.259	0.505
253	Multifamily	789,079	2018-06-25	2019-01-14	5	14.18	15.69	0.911	0.685
260	Commercial Retail	144,000	2018-09-03	2019-09-02	34	0.07	0.43	0.491	0.564
264	Commercial Retail	144,000	2018-11-19	2019-11-18	17	0.40	1.53	0.353	0.363
265			2019-09-23	2020-05-04	5	3.43	6.18	0.617	0.551
274	Commercial Retail	144,000	2018-09-17	2019-09-16	38	0.58	4.27	0.136	0.488
275	Commercial Office	522,000	2019-10-28	2020-04-27	9	366.20	448.35	0.821	0.337
277	Commercial Retail	198,057	2018-07-16	2019-07-15	22	0.26	1.47	0.252	0.299
282	Food/Beverage	56,000	2017-12-18	2019-09-30	37	1,143.91	5,270.35	0.384	0.439
285	Commercial Retail	198,057	2018-08-20	2019-08-19	17	0.28	1.01	0.354	0.494
294	Commercial Retail	144,000	2018-09-10	2019-09-09	38	0.42	1.87	0.215	0.407
297	Commercial Retail	198,057	2018-07-16	2019-07-15	24	0.46	1.48	0.327	0.451
307	Food/Beverage	56,000	2017-12-18	2019-09-30	47	2,476.87	5,782.70	0.481	0.380
316	Commercial Retail	144,000	2017-03-20	2019-03-18	59	2.01	8.78	0.237	0.523
322	Commercial Retail	198,057	2018-07-16	2019-07-15	19	0.29	2.84	0.113	0.458
327	Commercial Retail	198,057	2018-08-20	2019-08-19	12	0.11	0.29	0.375	0.583
332	Commercial Retail	120,000	2019-08-19	2020-02-24	7	54.45	65.96	0.845	0.390
350	Commercial Retail	198,057	2018-07-16	2019-07-15	19	0.37	1.79	0.399	0.324
353	Commercial Retail	198,057	2018-07-16	2019-07-01	19	0.28	1.99	0.201	0.449
372	Commercial Retail	198,057	2018-07-16	2019-07-15	14	0.20	0.90	0.266	0.509
373	Multifamily	127,577	2019-11-25	2020-06-01	10	20.36	23.56	0.864	0.559
380	Food/Beverage	56,000	2017-11-27	2019-09-30	49	3,213.26	7,497.76	0.552	0.467
386	Commercial Retail	198,057	2018-12-03	2019-12-02	29	0.03	0.03	0.988	0.470
390	Food/Beverage	8,500	2019-04-29	2020-02-24	5	6.73	15.71	0.429	0.798
393	Commercial Office	1,100,000	2019-07-29	2020-07-27	12	508.76	1,620.37	0.336	0.404
394	Commercial Office	725,000	2018-01-01	2018-12-31	386	3.52	11.92	0.445	0.496
420	Healthcare	236,000	2018-04-09	2018-12-31	6,068	144.76	227.08	0.723	0.527
423	Multifamily	119,710	2019-12-30	2020-11-30	463	2,922,233.79	6,979,961.45	0.613	0.665
426	Hospitality	80,000	2017-12-04	2020-11-30	959	7.99	12.36	0.692	0.439
440	Commercial Office	268,000	2020-09-28	2021-09-27	57	64.43	183.73	0.400	0.242
442	Multifamily	205,000	2020-12-28	2021-05-10	69	29.94	62.07	0.480	0.665

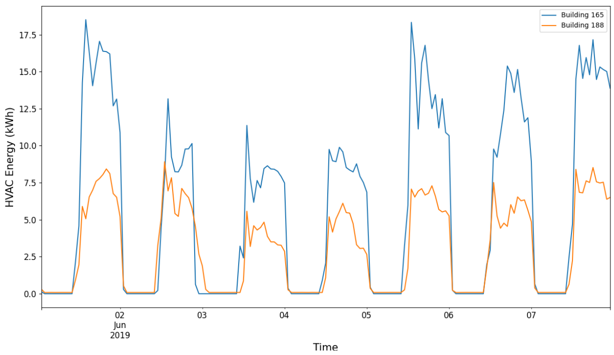
5 Analysis

5.1 Building 188

Building 188 and Building 165 are similarly sized commercial retail spaces with similar occupied electricity consumption. However, Building 188 has an average Unoccupied Turndown Factor of 0.424 while Building 165 has a value of 0.147. By observing the respective usage over a representative week in June 2019, we can indeed see larger energy consumption in Building 188 during unoccupied times.



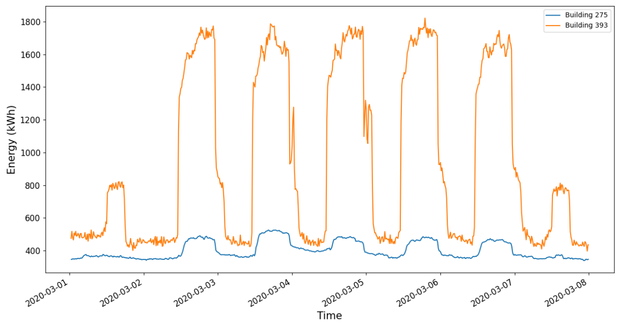
Since these facilities have HVAC submetering, we can start to disaggregate the total electricity consumption. By charting the total HVAC energy use over the same timeframe, we find that the nighttime usage in Building 188 is unrelated to the HVAC equipment, suggesting that it is likely related to light or plug loads



If Building 188 is able to achieve a turndown equivalent to that of Building 165, the KPIs suggest that its energy consumption will be reduced by about 11,500 kilowatt-hours (kWh) per year, which is roughly 16% of its total yearly energy usage.

5.2 Building 275

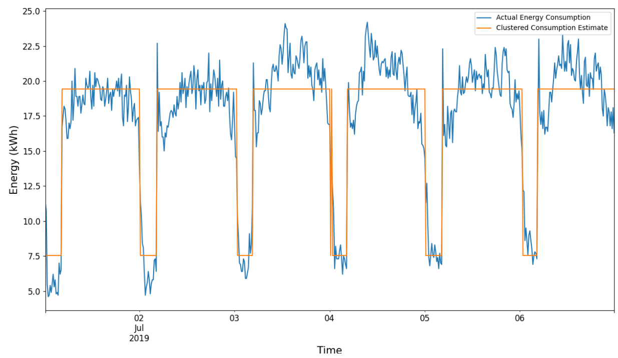
Building 275 and 393 are both large office buildings. Building 393 has more than twice the square footage and three times the occupied electrical consumption of 275. However, with an unoccupied turndown factor of 0.821, Building 275 has much less effective unoccupied setbacks than Building 393, whose factor is 0.336. By charting their usage together for a week in early March 2020, we can see the difference in the unoccupied consumption.



If Building 275 was able to achieve the same unoccupied turndown factor as 393, which is not unusual in the commercial office group, it would reduce its total energy use by an estimated 36%. Extrapolating from the 6 months of available energy usage data, this would result in savings of 5,000,000 kWh or nearly $750,000 of reduced consumption charges per year, based on a 15-cent per kWh cost.

5.3 Building 390

Building 390 is a food and beverage facility, with a large occupied duration factor of 0.798. An investigation of the eletrical consumption over a week in July 2019 shows that usage is high from roughly 4AM to midnight each evening.



There are certainly food and beverage facilities that keep these hours, as staff prep in the morning and clean up at night. However a quick investigation could determine if the consumption profile reflects actual occupancy schedules. If not, and the building could be run in an unoccupied mode even just 2 additional hours each day, that would save nearly 5% of this building's total energy consumption.

6 Future Growth

As the RTEM dataset expands, these weekly KPIs can be automatically computed on new data. In fact, computing KPI values for the complete RTEM dataset throughout all time currently takes only a few minutes on commodity hardware. Scripts to compute these KPIs are provided in the GitHub repository linked below the title.

However, the automation of this process suffers from incomplete data modeling of electrical metering within the RTEM dataset. For example, given a site with multiple electric consumption readings, it is not denoted whether a specific reading is a subset of another or fully separate. Top level consumption readings are not differentiated from low-level submeters. Due to this, the final point list used within the analysis was manually revised to determine which points (or collection of points) represented each building as a whole. Project Haystack models metering trees using the `submeterOf`[2] tag, and denotes top-level metering using a `meterScope`[3] tag. Brick Schema models building-level meters with a `Building_Meter`[4] class. Integration of these ideas into the RTEM dataset would improve future automation efforts.

7 Additional Applications

The clustering approach shown in this document is not specific to total-building electric energy use. It could potentially be applied in very similar ways to:

- Other building-level utilities: Total building gas, chilled water, or steam use
- Submetering: Zone or tenant-level unoccupied setback analysis and tracking
- Downstream occupancy effectiveness metrics: Zone air temperature setpoints, VAV discharge airflow

8 Conclusion

This document presents a strategy for analyzing the effectiveness and prevalence of unoccupied setbacks by using a clustering algorithm on a historical whole-building consumption record. It outlines the KPIs that can be calculated from this record, and demonstrates how they can be used on real-world data from the RTEM dataset. Finally, it discusses the feasibility of expanding this analysis as the RTEM dataset grows.

Energy conservation is the most cost-effective means of decarbonization. Of course, investments must be made in switching fossil fuel-based technologies to clean energy, and expanding renewable energy production. However, in the meantime, we must also make sure that our existing infrastructure is working as efficiently and effectively as possible. Poor unoccupied operation is the largest

[2] `https://project-haystack.org/doc/lib-phIoT/submeterOf`

[3] `https://project-haystack.org/doc/lib-phIoT/meterScope`

[4] `https://brickschema.org/ontology/1.2/classes/Building_Meter`

opportunity in most buildings that does not require a capital investment. We ought to make sure it is working well and continues to do so.