Return to Kronos

Aaron LaFevers

University of Denver

COMP 4449: Data Science Capstone

Claudio Delrieux

August 18, 2023

Return to Kronos

Purpose

In this analysis I examine two weeks' worth of employee proximity and building systems data, with the goals of identifying patterns of life among employees, among building systems, and to detect anomalies in these data, and lastly to discern relationships between the employee and systems data.

Scenario

GAStech, a company headquartered in Abila, Kronos, moved into a new office building in late 2015. The building is equipped with proximity sensors which detect "Prox" cards, carried by employees. The building is also equipped with a modern HVAC system with a suite of sensors. This includes four sensors dedicated to the detection of a new and little-understood gas, called Hazium. There is a mail robot that moves about the building freely and is itself equipped with a proximity sensor. It is now mid-2016, and GAStech has employed me to investigate a two-week sample of their data streams to uncover patterns, anomalies, and relationships.

Dataset Description

The data provided by GAStech includes prox data, both from static sensors and from the mobile sensor. There are approximately 30k and 12k entries in the static and mobile datasets, respectively. An employee directory was included with 125 employees, their departments and office numbers. Building systems data was provided in 5-minute logging intervals from 415 data sources, resulting in 4032 entries. A separate dataset provides 5-minute data during the same period for the four Hazium sensors. Lastly, GAStech provided a set of floor plans, a set of HVAC zones, and a set of prox zones. One map per floor, per type, for a total of nine JPEG map files. Data Preprocessing

Data Cleaning

The data was fairly clean, with no missing data. There were duplications in the mobile prox data caused by the high sampling rate of the mail robot. It polled for nearby prox cards on 1 second intervals,

logging at 1 minute intervals. Due to its rate of travel it therefore frequently pinged the same prox card multiple times from the same location. This resulted in a high number of duplicate entries in the mobile prox data which were removed, leaving 2.5k entries from this source.

Data Preparation

To facilitate a programmatic analysis of the prox data, and in particular to combine the coordinate-based mobile data with the zone-based static data, the provided JPEG maps had to be distilled into a for capable of being queried. To accomplish a script for Adobe Photoshop was written, in JavaScript, which output the bounds of each image layer in *left, top, right, bottom* format, as a CSV. Each map was then brought into Photoshop, cropped to the building outline, and scaled to 10x the dimensions of the building in feet (189ft -> 1890px). A series of rectangles were then drawn over each of the prox zones, HVAC zones, and offices, one rectangle per layer, with the name of that zone/office set as the layer name in Photoshop. The script was then utilized to capture a CSV representation of the coordinates for the boundaries of each zone, scaled back down to feet.

These CSVs allowed the office numbers provided in the employee directory to be mapped to HVAC and prox zones. The employee directory did not include a Prox-ID field, and therefore a prox ID was estimated for each employee using a *first initial* + *last name* + *001* format. The enriched employee dataset was joined to the prox data. There were 1.8k instances in the prox data that did not match the estimated prox IDs, which stemmed from hyphenated last names and other forms of straying from the implied naming convention. These prox IDs were all matched to employees in the directory.

The mobile prox data was then parsed with the prox zone mappings to convert the provided coordinates to prox zones. This allowed the mobile prox data to be concatenated to the static prox data, and sorted by time. This full record of prox data was then interpolated to determine, on a 1-minute basis, where every employee was at each instance during the observation window. There is some attenuation of the data in this step as it is common for multiple prox zones to be crossed within a

minute. An illustration of this process, from employee data enriching to prox data interpolation is shown in Figure 1. The full location matrix was then down-sampled to 5-minute intervals to match the fidelity of the systems data and ease computational load.

The building systems data was combined with the Hazium data, and fit into a multi-indexed data frame to facilitate easy filtering by floor, zone, and sensor type, hereafter referred to as 'node'. The systems data was transformed into three major datasets, with three main forms of each. The three main datasets were the raw systems data, standardized systems data, and binned systems data. The binning was done based on the standardized data, where signals between +/- 1 standard deviation were considered *flat*, 1-3 std were *low*, 3-5 std *medium*, and over 5 std *high*. There was a fourth dataset created that binarized this binned data, mapping all *medium* and *high* readings to 1. Each dataset was then provided in a *default*, *transposed*, and *stacked* form. The default data used timestamp as an index, and each sensor node as a column. The transposed version was required for clustering, treating each node as an instance. The stacked version was a tall form, with a timestamp index and a column specifying node, required for the visualizations. Figure 2 illustrates these main systems data resources.

Pattern Identification

Prox Data

To programmatically identify patterns in the employee prox data, an unsupervised clustering approach was taken. The prox data was transposed to treat each employee ID as an instance, with 4,032 features corresponding to each 5-minute interval in the two-week observation window. This transposed data was filtered to include only single-ID holding employees, and iteratively fed to a k-means clustering algorithm while varying the value of k from 1 to 20. The within-cluster variances were recorded for each iteration and an elbow plot was constructed (Figure 3). Analysis of the elbow plot suggested 4 main groupings for the employees. The reason for excluding the multi-ID holding employees was because each Prox ID represents only a fraction of that employee's "pattern". These

employees were included in a synthetic fifth cluster. The cluster assignment for each employee was then added to the employee dataset, and a time series visualization of every employee's location for the entire two-week period was generated, grouping each employee by cluster. To simplify the visualization employee location was binned into *building*, *office*, *away*, or *undefined*, corresponding to instances where the employee was in the building, in their personal office, out of the building, or unknown, respectively. The unknown case arises prior to the first prox card reading, and after the final prox card reading, as location can only be accurately interpolated between two readings. The resulting visualization is shown in Figure 4. Patterns identified in the prox data include:

1. Cluster 0: day shift, working 7:30-17:00, with regular morning and afternoon meetings, and a deli lunch break (within the building). This group comprises most of the engineers, and about half of the IT and Facilities staff.
2. Cluster 1: day shift, working 8:00-17:00, with sporadic meetings and lunch breaks, some taken offsite, and less rigorous working hours than cluster 0 with some early departures and late arrivals. This group consists of the majority of Administration and Executive departments.
3. Cluster 2: night shift, working 16:30-23:45, with regular meetings and lunch breaks like cluster 0. This group is made up of half of the IT and Facilities departments and a few Engineers.
4. Cluster 3: day shift, arriving between 7:30 and 8:00, departing 17:00, short consistent stints away from the office in morning and afternoon, with an offsite lunch every day from 12:00 to 13:00. This group consists predominantly of the entire Security department.
5. There are 3 employees in the facilities group who leave every day around 15:00, despite arriving at work at 7:30. They spend only the first hour of the day in their office. They are R. Parade, D. Scozzese, and C. Staley

6. There are 2 facilities employees who work the 3rd shift, from midnight to 7-7:30. They both leave their office at 2:30 each morning for an hour, and again about 1 hour before departing for the day. They are E. Arpa, and V. Awelon.
7. The executives lead the most irregular hours, often arriving late or leaving early. There are only 3 employees who missed a full workday during the observation period, and 2 of them were executives.

Systems Data

The building systems data was clustered in a similar manner to the prox data. The binned and transposed version of the data was utilized, treating each node as an instance, with 4,032 features corresponding to each 5-minute interval in the observation window. The elbow plot for this data is shown in Figure 5. Analysis of this elbow plot suggests 8 main groupings for the systems data signals. The cluster assignments for each node were recorded and the stacked version of the data was utilized to create a similar time series visualization, grouping the signals by predicted cluster. The resulting visualization is shown in Figure 6. This visualization highlights deviations from the signal mean, in absolute terms, and therefore in the discussion that follows a "peak" in the signal refers to an event beyond 1 standard deviation above **or below** the mean value. This may also mean that a "peak" represents the off or idle state of a system that, on average, is on or running. By definition though these states are abnormal for the given node. Patterns identified in the systems data with this understanding include:

1. Clusters 0, 1, 2, 3, 5, and 7 include nodes which all have peaks during non-business hours, these systems run at or near their mean level while the building is occupied.
2. Cluster 6 is a notable exception, with peaks corresponding closely to business hours. This cluster consists mostly of lighting power and equipment power signals.

3. Cluster 7 consists almost entirely of floor 3 systems, such as thermostat settings, which peak in non-business hours except for the period from Friday afternoon to Monday morning. It is likely that the HVAC system is running all weekend on the 3rd floor, which should be addressed to save on operating costs.
4. Cluster 4 exhibits almost no peaks in any signals, day or night, during normal operation. There is an anomalous event between June 7th and 8th which affected these nodes as well. This group consists of thermostat setpoints and temperature readings from all 3 floors.
5. The AC and heating setpoints in all zones usually follow a very regular schedule, stepping down (AC) and up (heat) to be within a tighter temperature range to cover business hours. This stage starts at 5:00 and lasts until 22:00, at which point both setpoints are relaxed giving a much wider acceptable temperature window when the building is mostly unoccupied. This behavior is displayed in Figure 7.
6. Clusters 0, 2, 4, 5, and 7 all reacted strongly during a system-wide anomalous event on the 7th and 8th. Only clusters 0 and 5 however reacted strongly to a separate, Hazium related event on the 10th/11th. Only cluster 3 reacted to the latter event and not the former. Further analysis of the systems anomalies were conducted.

Anomaly Detection

The process of anomaly detection, outside of the macro events captured by clustering, was undertaken by manual evaluation. The visualizations discussed above, as well as further, more focused visualizations in different formats, were created using the Plotly python library, which provides interactive plots with tooltip information “out of the box.” This enabled the visualizations themselves to be used as powerful tools for digging into anomalous events. The seven most interesting anomalies are described below, in order of lowest to highest priority.

Multiple IDs for G. Florez

Geneviere Florez is the only employee, at the time the data was collected, to have received more than two IDs. During the first week of observation, she lost her ID every day. She is now on her 5th ID. It is notable that none of her lost IDs have been detected at the building after activation of the next ID. G. Florez is one of six employees who have lost an ID. A close up of the activity corresponding to these six employees is shown in Figure 8. While most of their behavior is innocuous, issuing multiple IDs represents a security risk, and management should impress upon the employees the importance of securing their IDs.

Weekend CO_2 spikes with no Prox activity

There are a couple spikes in CO_2 concentration seen in the building in off-hours, but these events typically coincide with some Prox activity indicating someone entered the building. There was a CO_2 event on June 5th between 5:00 and 6:00 however that did not coincide with any Prox activity. It is possible that someone entered the building without using their prox card. This event only affected the 1st floor, but on a few sensors was a higher magnitude concentration than what was seen during the two major CO_2 events on the 7th and 8th. The CO2 concentration for all zones, colored by floor, is shown in Figure 9.

Weekend entries

Weekend entries were detected from L. Carrara, O. Strum, L Lagos, and M. Bramar. Some of these entries were adjacent to high Hazium concentration events. This activity in relation to the Hazium events is depicted in Figure 10. It is noteworthy that Bramar and Strum were both at the office on Saturday the 12th immediately prior to a large Hazium event. They previously had anomalous behavior adjacent to Hazium releases, with only a couple hours worked amid the previous Hazium event on the 9th, and Strum missing work altogether during the Hazium event on the 3rd. The weekend activity on

its own is anomalous, considering only four employees entered on a weekend during the observation period, but the adjacency to the Hazium events makes Strum and Bramar's activity suspicious.

Midnight entries

There were two occurrences of employee Prox cards entering the building at midnight. This was seen for both G. Florez and S. Sanjorge, on the mornings of the 7th and 2nd, respectively. In both cases the cards were detected entering the first floor zone 1 (main entrance and common area), and not detected leaving or entering any other zone until around 8am in G. Florez's case.

For S. Sanjorge, his card didn't leave the first floor zone 1 until about 2pm that afternoon. This is also strange because on the 1st and 3rd S. Sanjorge only worked for a couple hours in the morning. Was this afternoon activity genuinely S. Sanjorge, or someone else using his card? This prox activity is shown in Figure 11.

Patrick Young's extra Prox card

All employees who have been issued more than one Prox card have ceased any activity on the previous card(s). There are two exceptions to this. The first is V. Morlun, who used a new card for one day on the 2nd, and then resumed using his original card after presumably finding it. The second is P. Young, who lost his Prox card '001' on the 1st, and began using Prox card '002' on the 2nd. Young's '001' card, however, began exhibiting simultaneous activity on the 2nd, which continued until the 13th. Upon closer inspection of the Prox data, it was discovered that Young's original ID has been kept in Loretta Bennett's office on the second floor. It appears she, or someone else, has been using Young's ID to, among other things, visit the 3rd floor server room four times during the observation window. Neither Young nor Bennett typically visit the server room, based on the data available. The activity for both Young and Bennett is plotted in Figure 12, with Young's activity being represented by a combination of his first ID, prior to losing it, and his second ID. The simultaneous activity from his first ID is shown in a separate trace, with the Y-axis representing the floor and zone of detection, in *Floor_Zone* format.

Young's office is located immediately across the hallway from the 3rd floor server room. So, for someone to use his ID to enter the server room on four separate occasions, and fail to return the card to Young, suggests they were keeping his ID intentionally. It is likely Young himself was unaware of this activity with his original ID. It is further likely that the person using Young's ID was Bennett herself.

Major systems anomaly on the 7th and 8th

There were major systems anomalies on all floors and nearly all zones on June 7th and 8th. It was a cyclical event starting around 7:00 the morning of the 7th, peaking around 17:00 that evening before dipping, starting over again at the low point at 7:00 on the morning of the 8th. Most systems had an anomalous spike at this time, but perhaps the most prominent and worrying is the CO_2 concentration. There were Hazium spikes on the 7th and late on the 8th into the 9th, but they did not coincide quite as succinctly as most of the HVAC systems did. It is likely the events are related, and ultimately it is becoming clear that an employee, or a group of employees are manipulating the building systems and precipitating the Hazium spikes. Figure 13 shows the CO2 concentration averaged by floor.

Hazium events

Little is currently known about Hazium gas and its specific effects on the human body. It is generally understood to be hazardous though, and GAStech saw fit to install four sensors to detect its presence. During the observation period there were three major Hazium spikes detected, affecting 2-3 floors simultaneously. There were two lessor spikes, also involving multiple floors, and then a multitude of minor events. The standardized signal from the Hazium sensors, averaged by floor, is plotted in Figure 14. There is no regular pattern of Hazium concentration similar to the CO2 concentration, and it is not known what the safe level of Hazium is for human occupation. The large Hazium spikes however are anomalous, and potentially very dangerous.

Relationships in Prox and systems data

The four following relationships were identified between employee activity and systems activity. There are other relationships present in the data, however these are the most significant and they are presented here in order of lowest to highest priority.

Sten Sanjorge (CEO) - erratic thermostat + odd hours

It is observed that S. Sanjorge's office thermostat settings were modified on June 2nd. Despite the summer temperatures outside (29C/84F), Sanjorge's thermostat heating setpoint moved to 32C at 13:00 each day. The AC cooling point at this time was 35.2C. These are very hot temperatures for an office environment and would have been uncomfortable to work in. At 5:00 the setpoints drop, with the AC setpoint falling to just 10C. In the time between 5:00 and 13:00 the office temperature holds around 20-21C (68-69F), despite the low setpoint. This is chilly but a common target range. The room temperature and HVAC setpoints for Sanjorge's office are plotted in Figure 15. It is possible that the CEO runs hot and simply wanted the room colder while working in the first half of the day. It is strange however that the room is allowed to heat to over 32C (90F) in the afternoons. The CEO's Prox activity is plotted in relation to periods during which his office exceeded 26C, in Figure 16.

It does not appear, based on Figure 16, that the high temperature inside the CEO's office dramatically affected his work pattern. He seems to only work a half day on Fridays and looks to have left the office on the 3rd prior to the room heating up. The following week he worked Monday through Thursday in the office in the afternoon with temperatures around 32C/90F. This is certainly strange behavior. Sanjorge may be ill, or someone has modified his thermostat intentionally, perhaps to encourage the CEO to leave the office/building. It is noteworthy that on the day of the thermostat changes, the 2nd, Sanjorge's ID entered the building at midnight. The ID made no movements beyond the first-floor common area until normal business hours, but the entry is strange and coincident.

Full building erratic thermostats + CO_2 surges

It is observed that most thermostat set points were modified on/around the morning of June 6th. The thermostat settings for every zone, plotted separately by floor, are shown in Figure 17. There are isolated anomalies throughout, but a major upset to the 3rd floor zone 1 system (CEO's office) starting on June 2nd, and a full-building upset beginning late on the 6th.

This behavior is compared to the CO_2 concentration and VAV damper position, standardized and averaged by floor, in Figure 18. Looking closely at the signals for thermostat setpoint (AC or heat), damper position, and CO_2 concentration, a causal chain can be established. When the thermostat setpoint lowers and the HVAC system starts to cool the space, the damper opens. When the damper opens, the air within the building circulates out, and this lowers the CO_2 concentration. The CO_2 concentration climbs during the day while the building is occupied, even with an open damper. Due to the thermostat setpoints being simultaneously changed on the 6th, the heat kicked on at 7:00 on the 7th. This setpoint change led to the damper closing, which while the building was occupied caused a major spike in CO_2 concentration. There are many other system spikes associated with this event, as seen in Figure 6, mostly related to power consumption and heating system function. It is the summer, and the heating system is not typically in use, so any activity in the heating system would be a dramatic departure from summertime mean levels.

Strum and Bramar + Hazium surges

It is observed that both M. Bramar and O. Strum had anomalous location behavior during times adjacent to large Hazium spikes. Upon closer inspection of their location data, additional anomalous behavior was discovered. Figure 19 shows all employee Prox activity related to the most prominent Hazium events. Based on observations of Figure 19, several people have been singled out as having suspicious or anomalous behavior in temporal proximity to the Hazium events. The records from the employee directory corresponding to these individuals are provided in Table 1. The locations of these

individuals as detected by the mail robot (mobile Prox sensor), are plotted over building floor plans in Figure 20.

Activity by employees outside of their personal offices or common conference rooms could be suspicious. In certain cases the employee's job duties may dictate the activity, for instance Strum, an executive, visiting the CEO's office, or the CEO visiting A. Campo's office (another executive). There were several instances though of employees visiting vacant offices, such as Bramar visiting room 3505 and Strum visiting room 2692. We examine further the activity of these and other suspicious individuals in the next section.

Suspicious activity and a plot against the GAStech CEO

This section builds on the patterns and anomalies discussed so far in this analysis. It has been shown that P. Young's ID was fraudulently used to enter the 3rd floor server room, that thermostat changes threw the building systems into chaos on the 7th and 8th, and that M. Bramar and O. Strum are potentially involved in the Hazium outbreaks. Tables 2 and 3 list all mobile Prox entries during the observation period for activity in the vacant rooms 3505 and 2692, respectively. No other personnel visited room 3505, and no specific event from the systems data can be tied directly to this activity on the morning of the 1st, however Bramar is under suspicion for other activities as well, so this is further enforcement of the theory that Bramar is a nefarious actor.

According to Table 3, Strum visited the vacant office 2692 around 14:30 on the 31st. There was no anomalous systems activity coinciding with this exact moment, however the following day at the same time C. Lais visited the same vacant office. Neither Strum nor Lais have offices on the 2nd floor, so these visits are immediately suspicious. At approximately 13:05 on the 1st, there was simultaneously a Hazium surge exclusively in the CEO's office, as well as thermostat setting changes in one zone on each of the three floors. This timing coincides closely with Lais' visit to the vacant room, and to the moment

that P. Young's original ID was "lost". It is also noteworthy that the very next morning room 2692 was visited by three security workers.

To investigate the suspicious use of Young's ID vis-a-vis the server room, all employees who accessed the server room at any point are listed in Table 4. No other employees from the Facilities department enter the server room during the observation window. There is also reason to believe that Young did not actually access the server room himself, as only his lost ID entered the server room. Therefore it is only L. Bennett and the Administrator C. Lais who have unexplained visits to the server room. Lais being implicated in the dubious access of the vacant office 2692 as well. The mobile Prox data from all of the employees who accessed the server room is mapped over floor plans in Figure 21, which corroborates Lais and Bennett's activity as described above.

Conclusion

We can conclude with a high degree of certainty that Strum and Bramar have acted abnormally, with odd visits to vacant offices and weekend activity in association with multiple Hazium outbreaks. Taken in total with the other anomalies, patterns, and relationships discovered in this analysis, it is possible that Strum and Bramar are central to a plot against the CEO, in which they are attempting to exploit the as-yet poorly understood Hazium gas. It is also highly likely that C. Lais, and L. Bennett are party to this plot, with their abnormal activity in relation to the 3rd floor server room and the temporal proximity of that activity with systems events. The illicit use of P. Young's missing ID by L. Bennett is perhaps the clearest sign of nefarious intent.

Recommendations

1. Update the thermostat settings on the 3rd floor to prevent the cooling system running all weekend.
2. Launch an investigation into O. Strum, M. Bramar, L. Bennett, and C. Lais.
3. Confiscate P. Young's ID from L. Bennett.

Tables

Table 1

Employee directory records for individuals with suspicious activity during Hazium events.

Prox ID	Last Name	First Name	Department	Office
earpa001	Arpa	Emile	Facilities	1070
vawelon001	Awelon	Varro	Facilities	1070
mbramar001	Bramar	Mat	Administration	3100
lcarrara001	Carrara	Lise	Administration	3420
dcoginian001	Coginian	Dante	Facilities	2260
llagos001	Lagos	Linda	Administration	3210
amorlun001	Morlun	Adan	Facilities	2660
morlunv001	Morlun	Valeria	Facilities	2300
jsanjorge001	Sanjorge Jr.	Sten	Executive	3000
ostrum001	Strum	Orhan	Executive	3200
cwhaley001	Whaley	Clemencia	Engineering	2560

Table 2

All mobile Prox activity detected in vacant office 3505.

Timestamp	Type	Prox ID	Floor	Zone	x	y
2016-06-01 09:45:00	mobile-prox	mbramar001	3	2	82	90

Table 3

All mobile Prox activity detected in vacant office 2692.

Timestamp	Type	Prox ID	Floor	Zone	x	y
2016-05-31 14:33:00	mobile-prox	ostrum001	2	7	118	62
2016-06-01 14:26:00	mobile-prox	clais001	2	7	118	62
2016-06-02 09:32:00	mobile-prox	kherrero001	2	7	118	62
2016-06-02 09:32:00	mobile-prox	mvollan001	2	7	118	62
2016-06-02 09:32:00	mobile-prox	slea001	2	7	118	62

Table 4

Employee directory records for all individuals who accessed the 3rd floor server room at any point.

Prox ID	Last Name	First Name	Department	Office
lbennett001	Bennett	Loretta	Facilities	2345
csolos001	Solos	Celestina	IT	2585
sflecha001	Flecha	Sven	IT	2160
ncalixto001	Calixto	Nils	IT	2155
pyoung001	Young	Patrick	Facilities	3230
clais001	Lais	Cornelia	Administration	3500

Figures

Figure 1

Prox data preparation process, illustrated

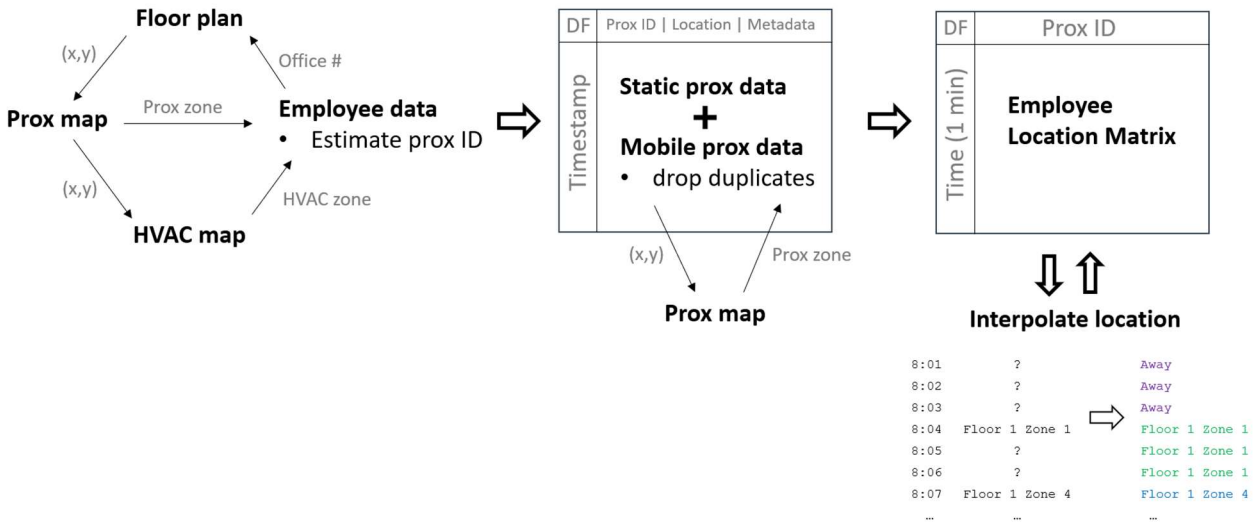

Figure 2

Main systems data formats

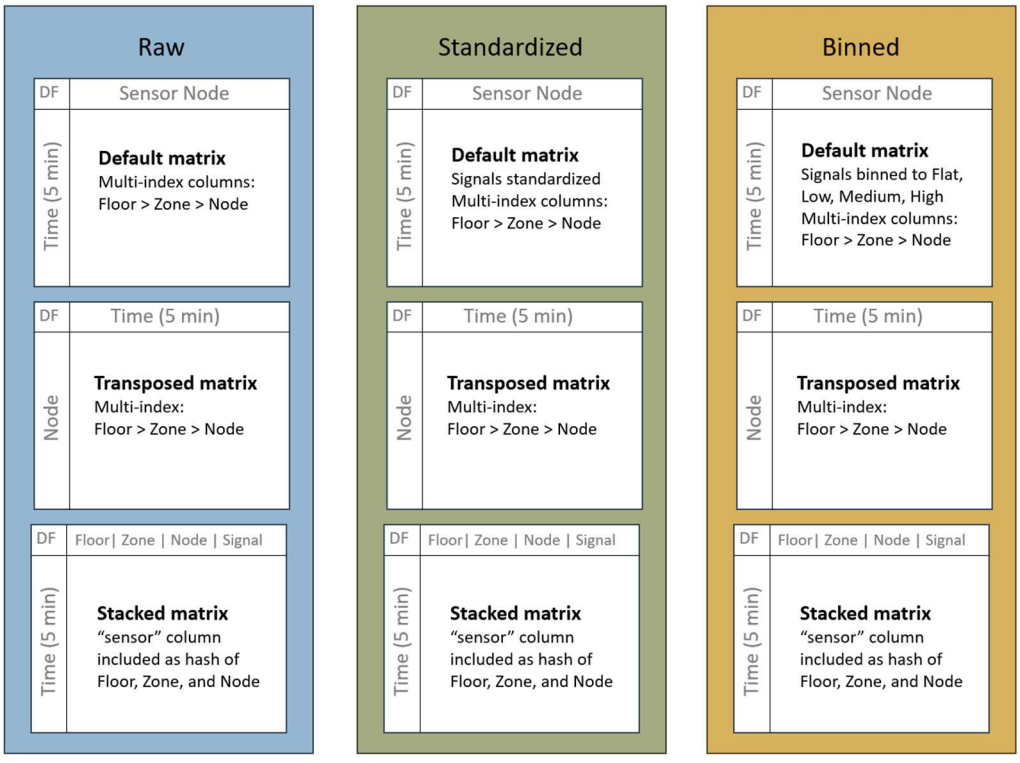

Figure 3

Elbow plot for Prox data K-Means clustering

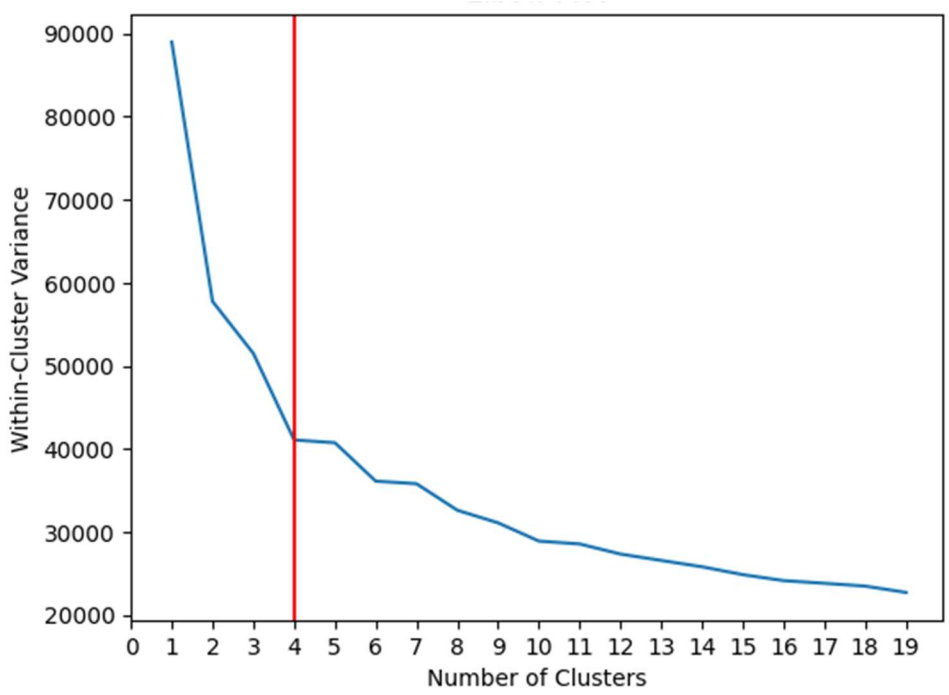

Figure 4

Time series visualization of all employee Prox activity, grouped by predicted cluster

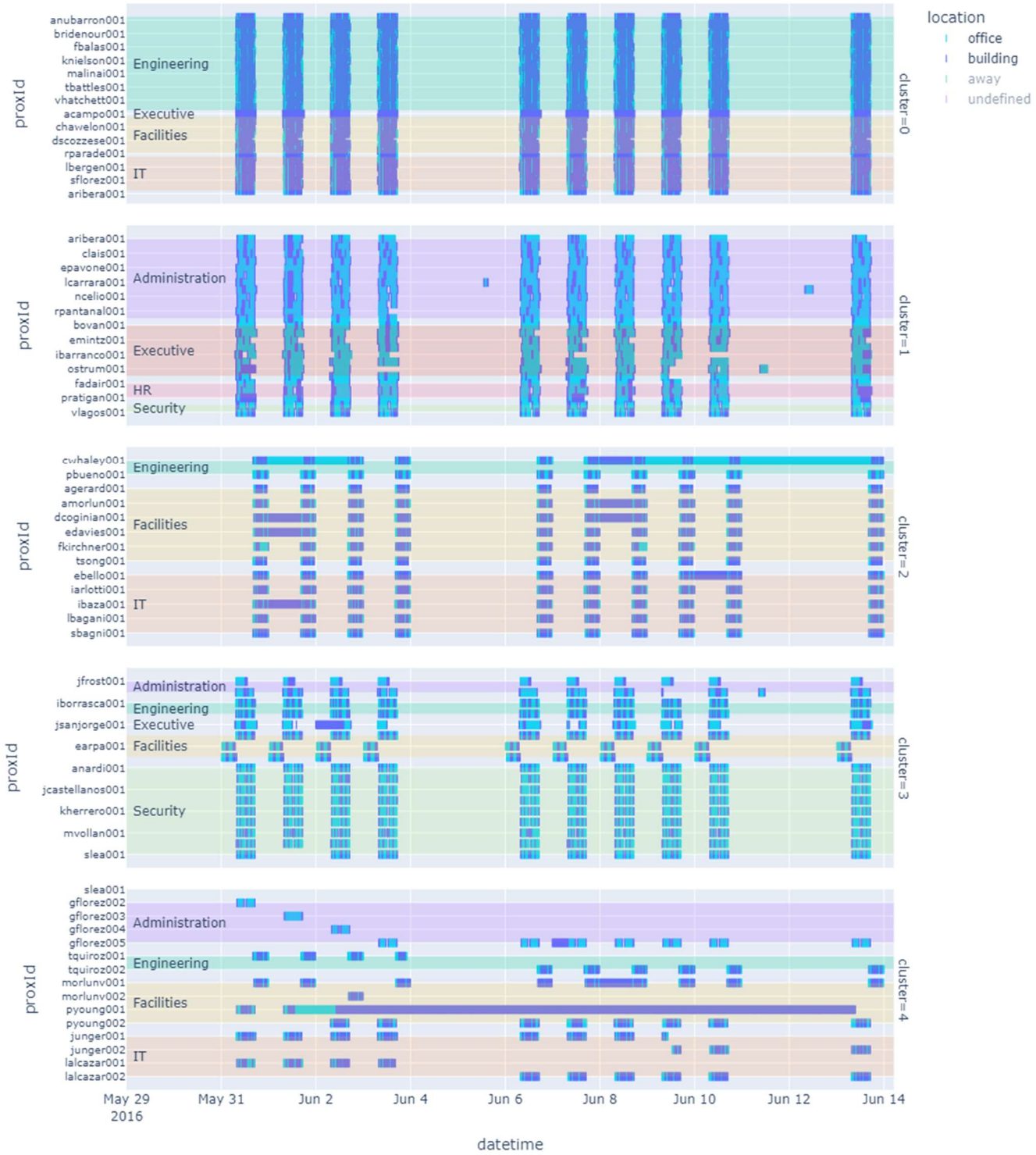

Figure 5

Elbow plot for binned systems data K-Means clustering

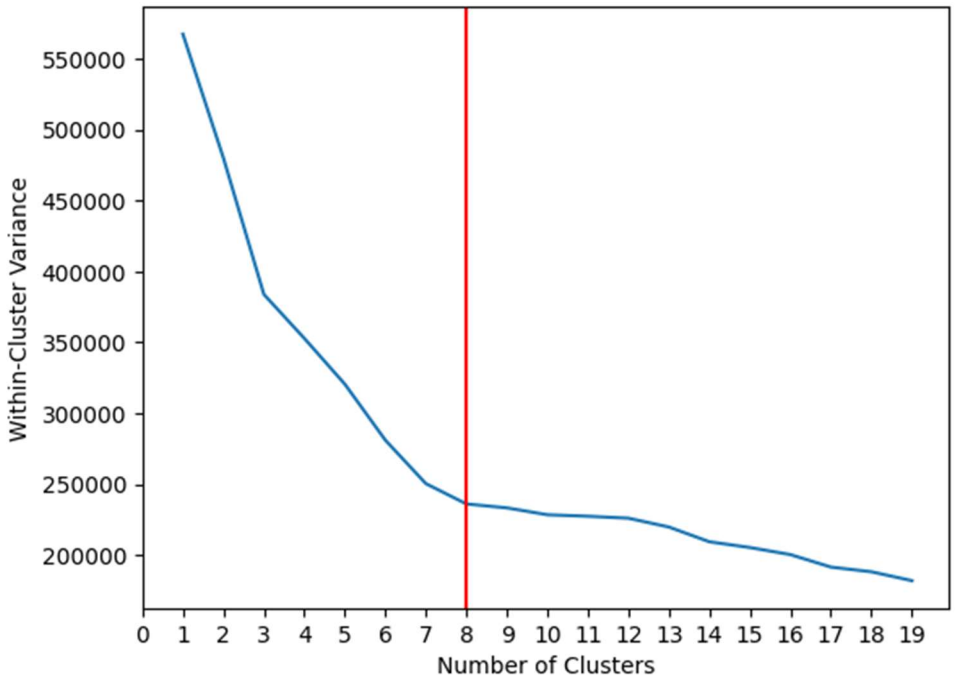

Figure 6

Time series visualization of all systems data signals, binned and grouped by predicted cluster

Figure 7

AC and Heating setpoints, floor 2 average

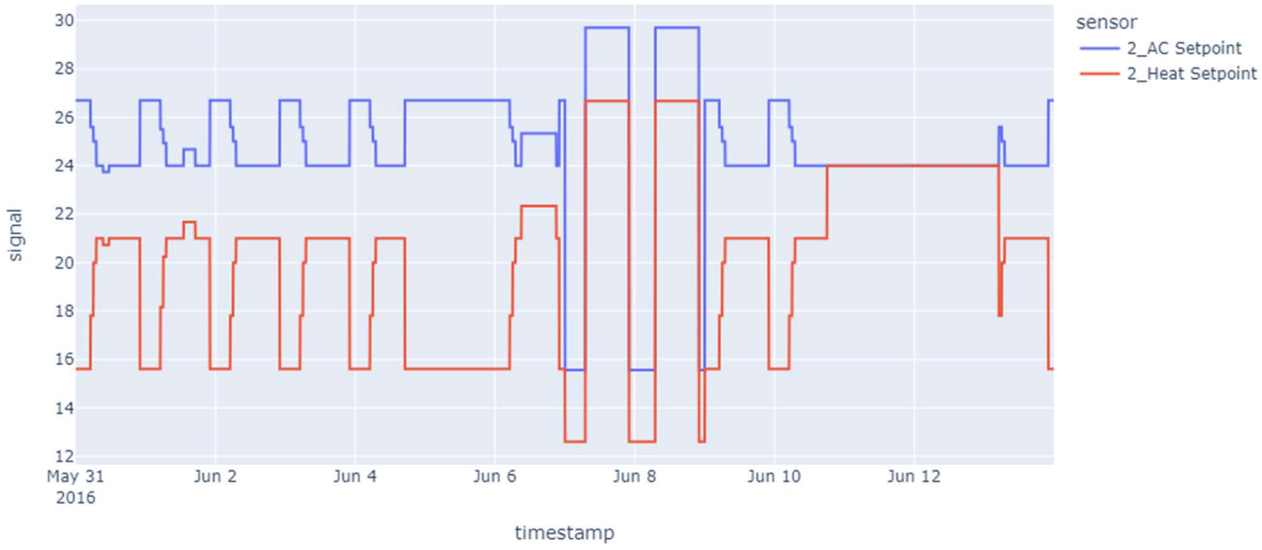

Figure 8

Multiple-ID-holding employees, Prox activity

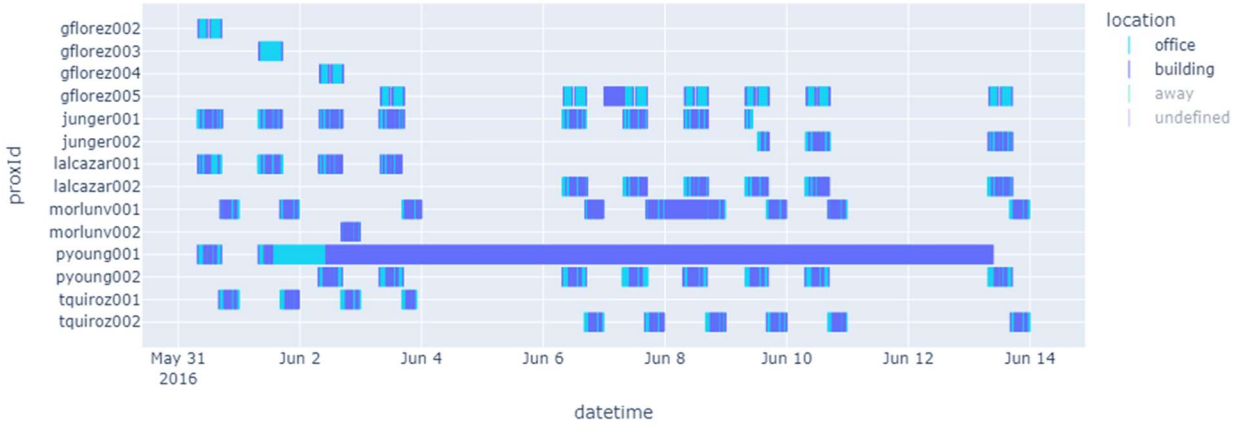

Figure 9

CO2 concentration spike on June 5th

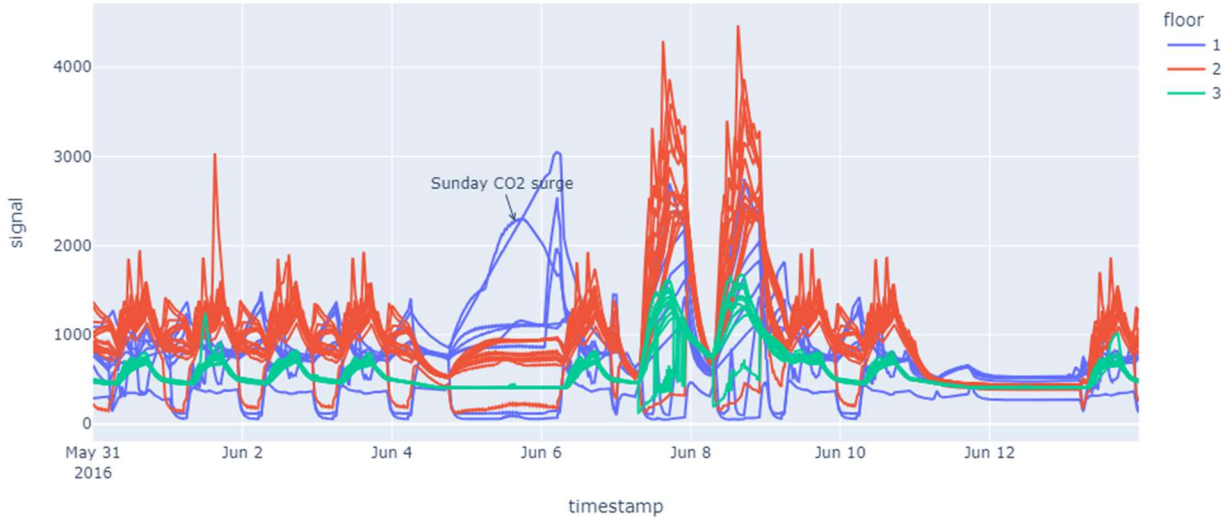

Figure 10

Weekend Prox activity

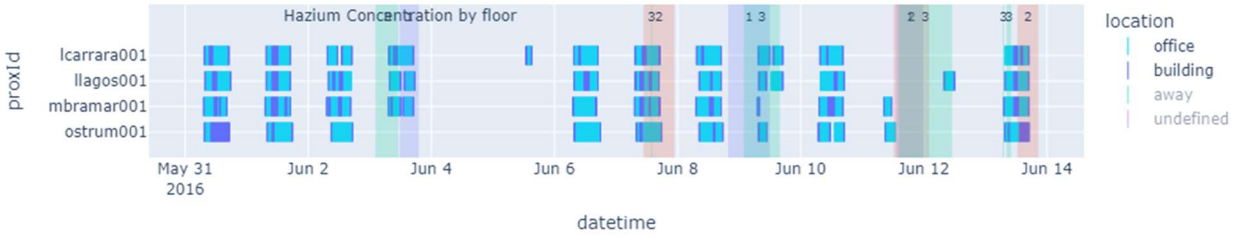

Figure 11

Midnight entries

Figure 12

P. Young and L. Bennett activity May 31 – June 3

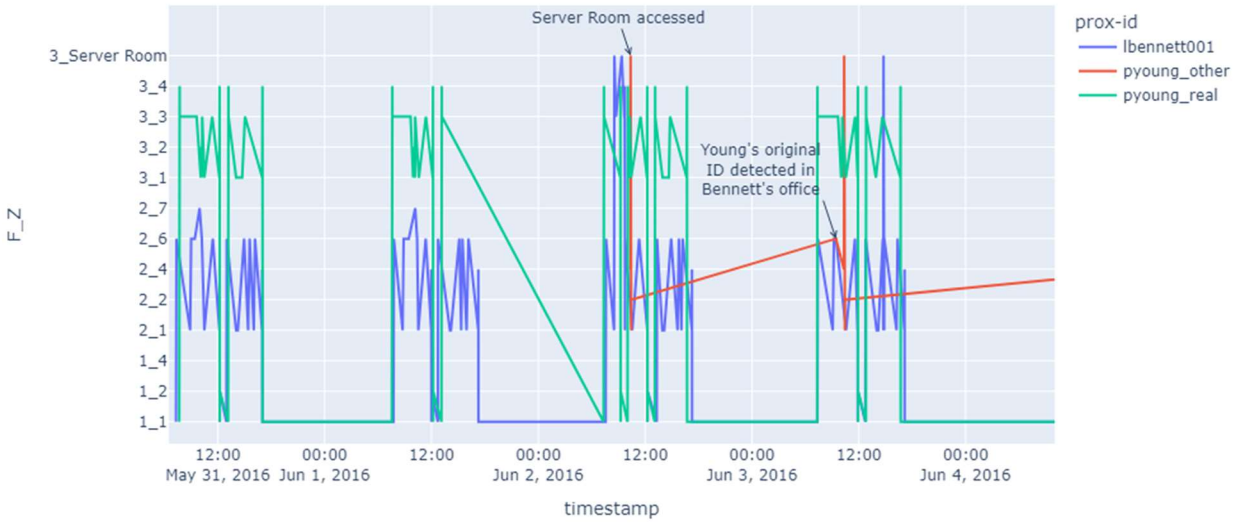

Figure 13

CO2 concentration (ppm), average by floor

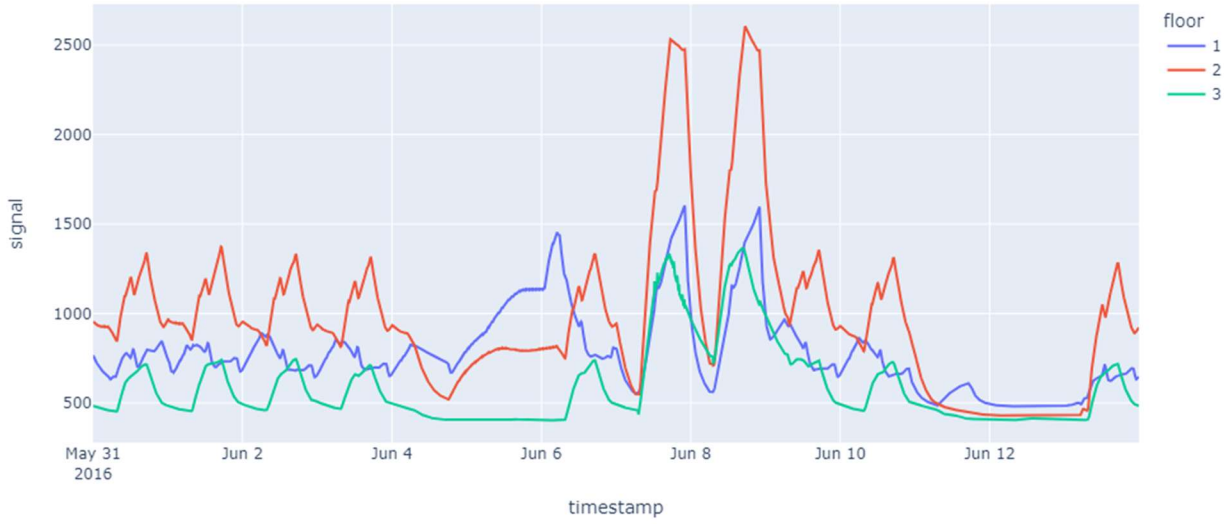

Figure 14

Hazium concentration, standardized, average by floor

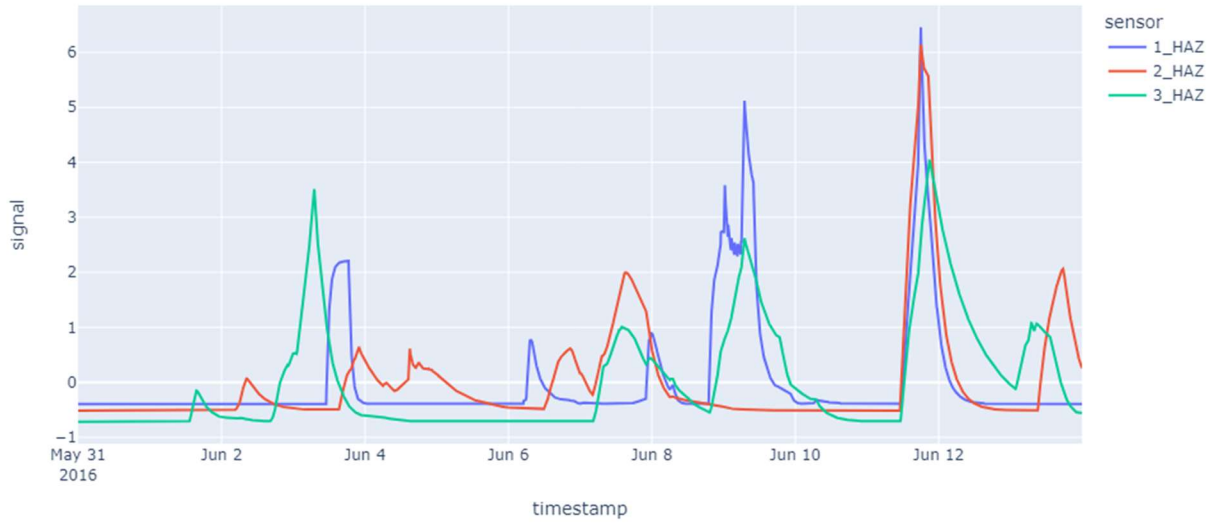

Figure 15

CEO's office – temperature and thermostat setpoints

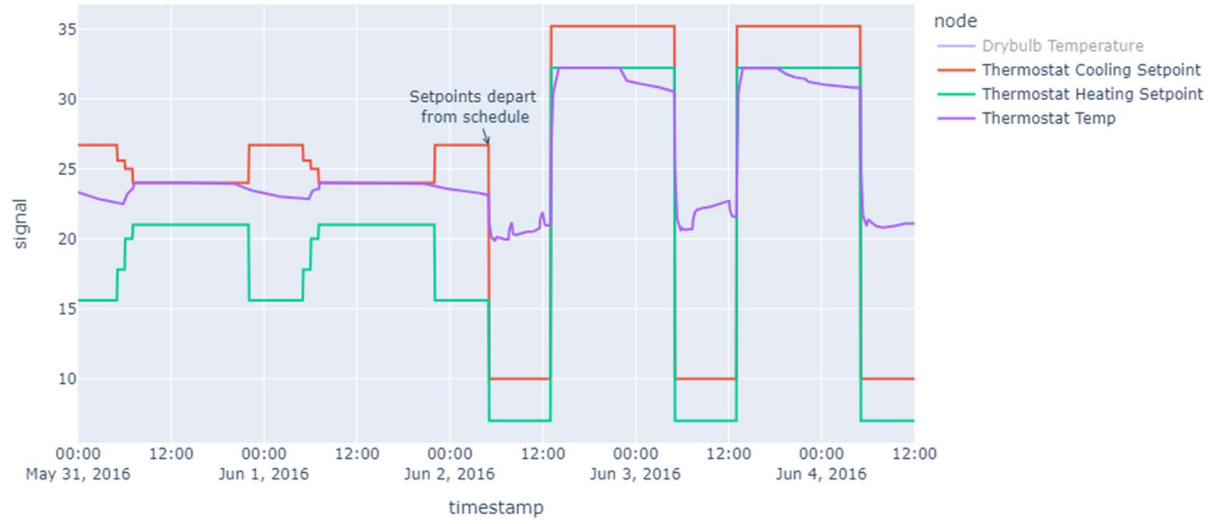

Figure 16

CEO activity related to office temperatures >26C (red shading)

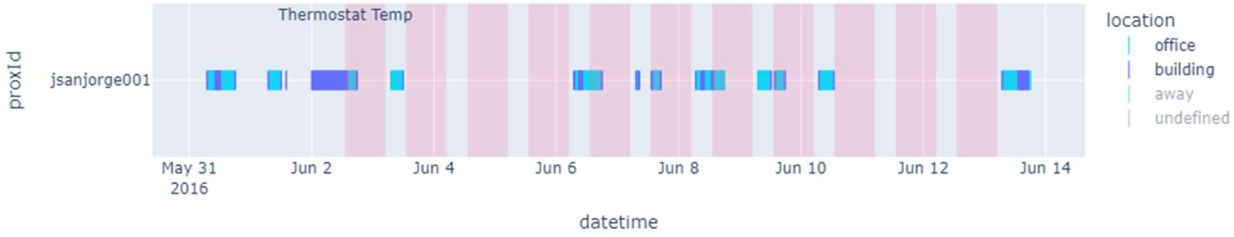

Figure 17

AC thermostat setpoints by floor (centigrade)

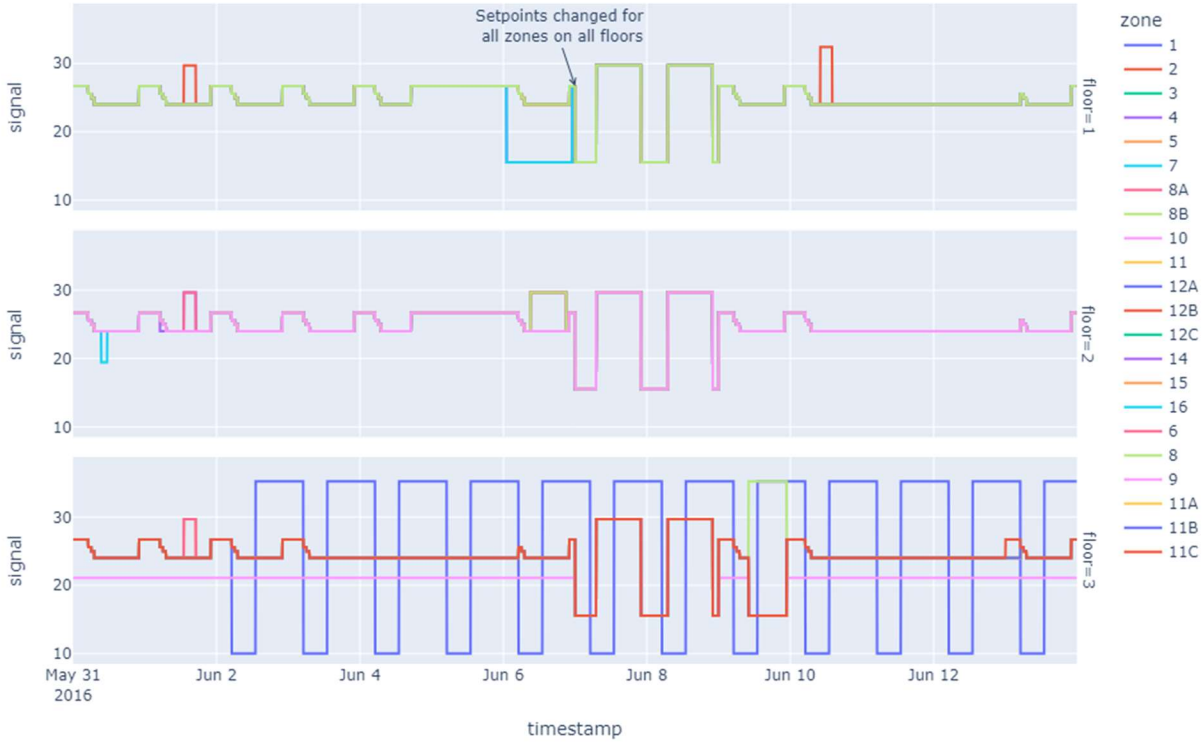

Figure 18

CO2 concentration versus VAV damper position, standardized, average by floor

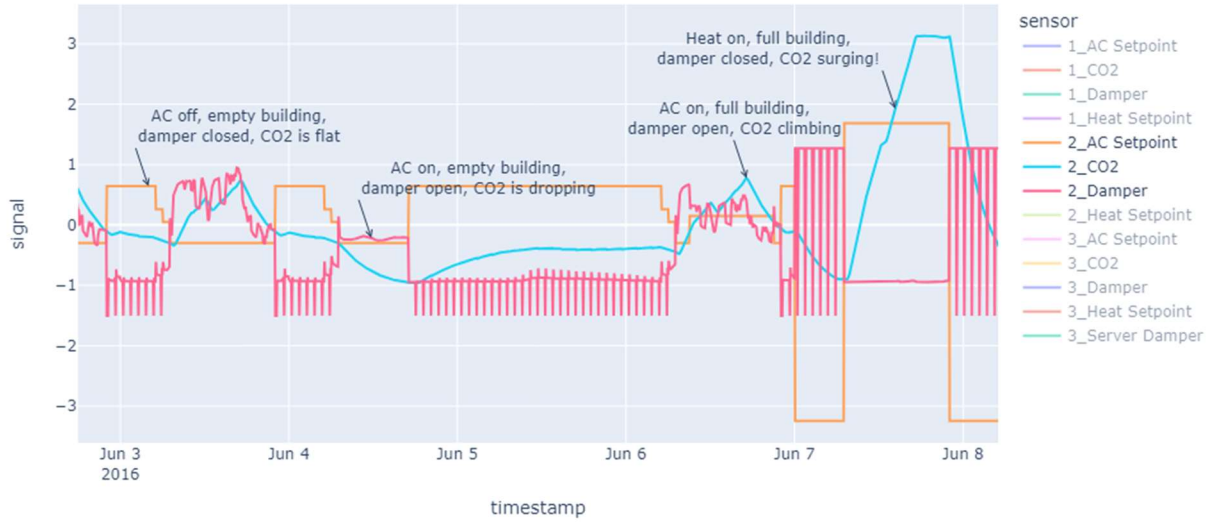

Figure 19

All employee activity versus elevated Hazium concentration events, grouped by employee cluster

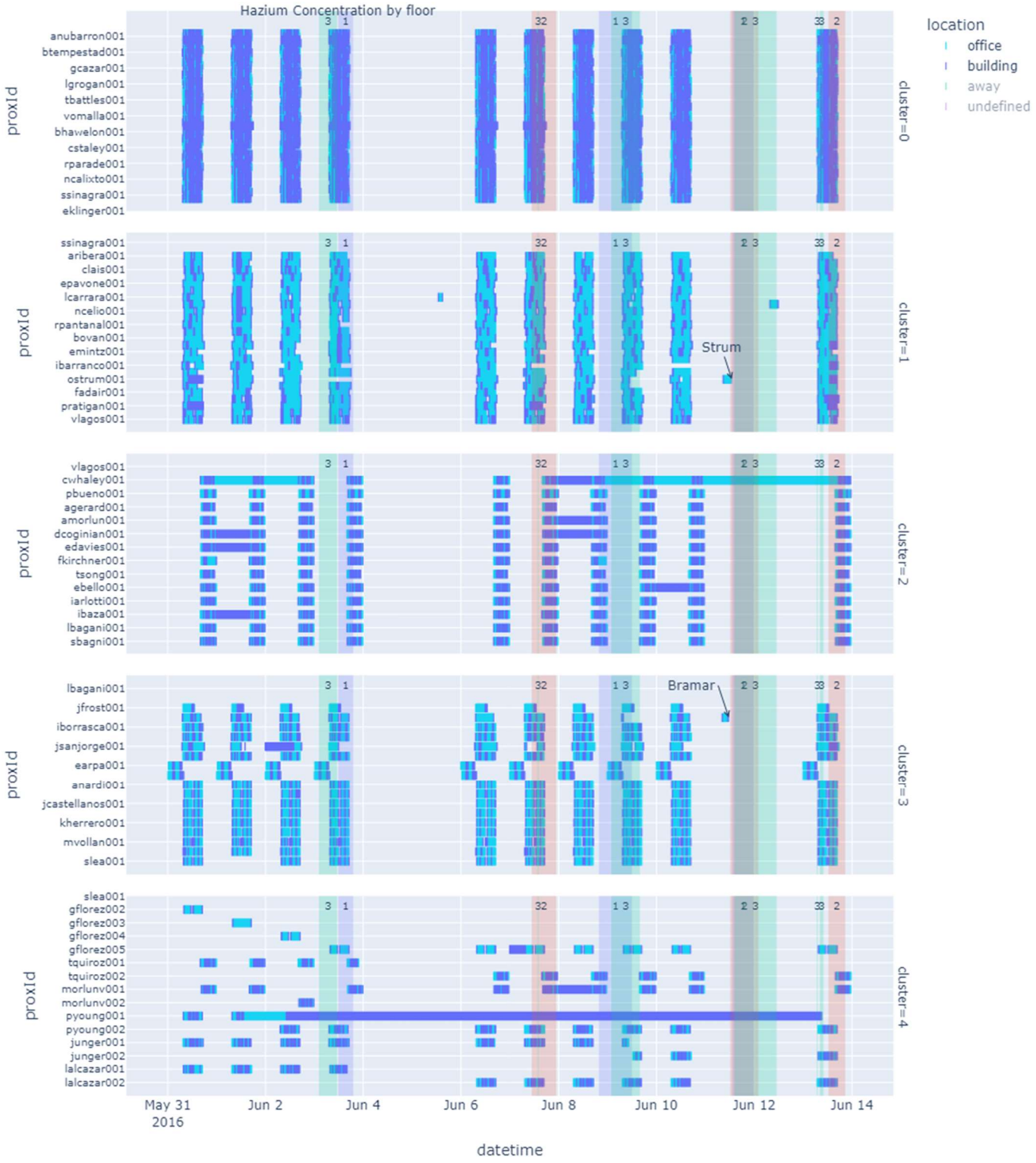

Figure 20

All mobile Prox activity for all "suspicious" individuals

Figure 21

All mobile Prox activity for all individuals who accessed the 3rd floor server room at any time