**Abdul-Rahman Abdel-Fattah**

**SID: 210046015**

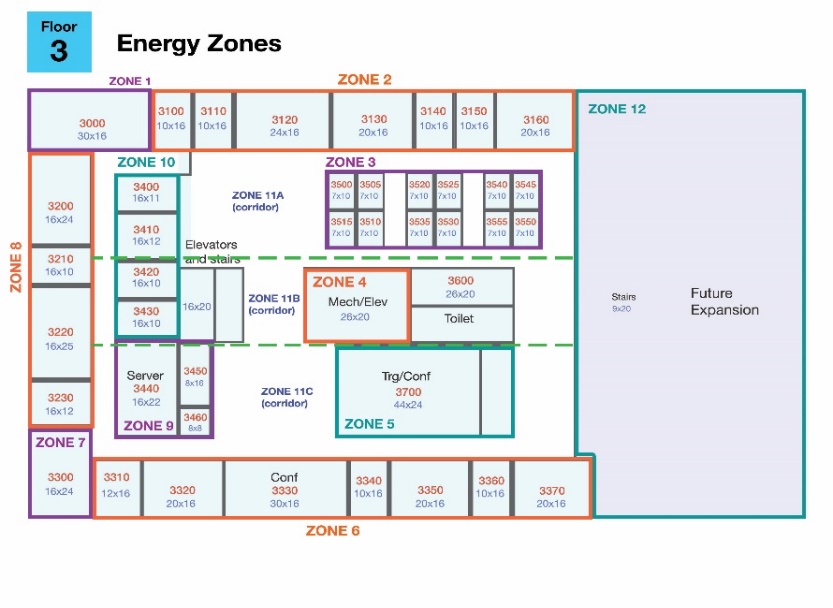
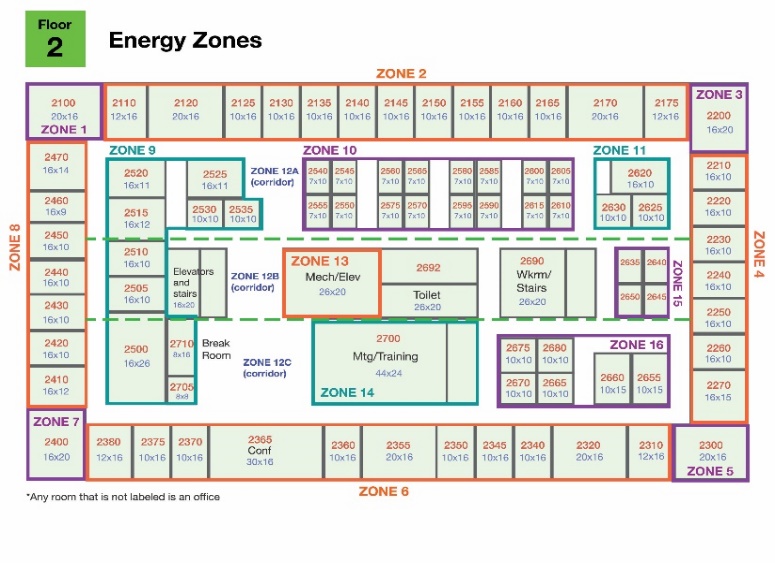
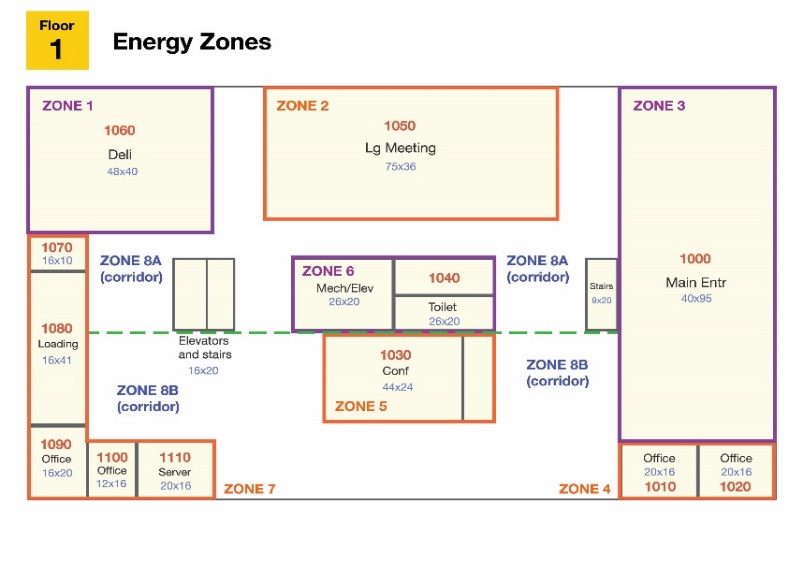
**Data Science Tools and Applications**

**EDA Challenge**

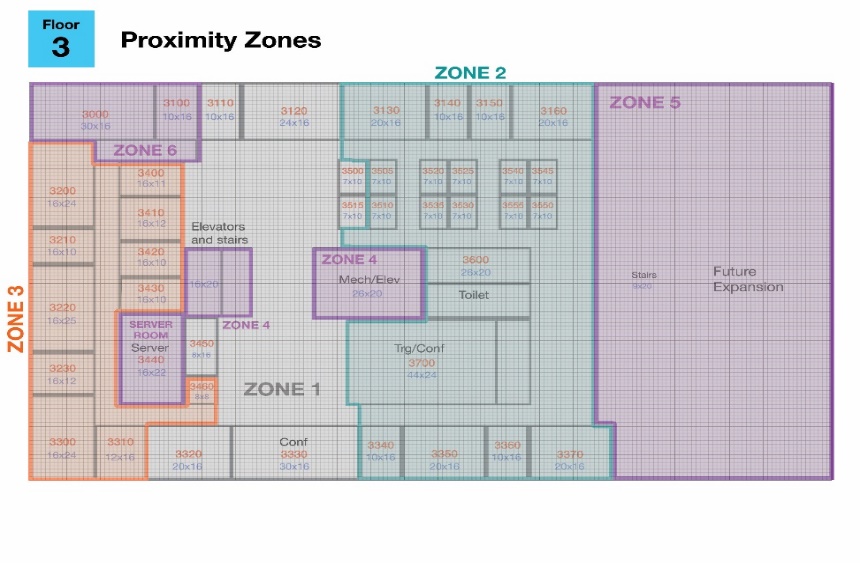
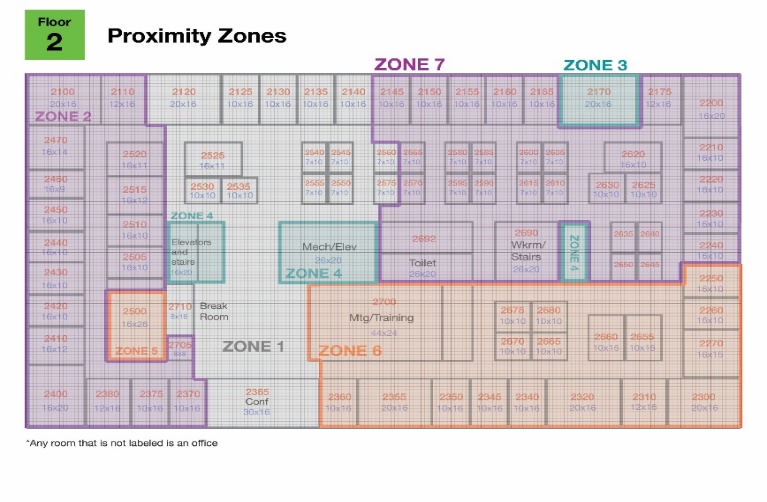
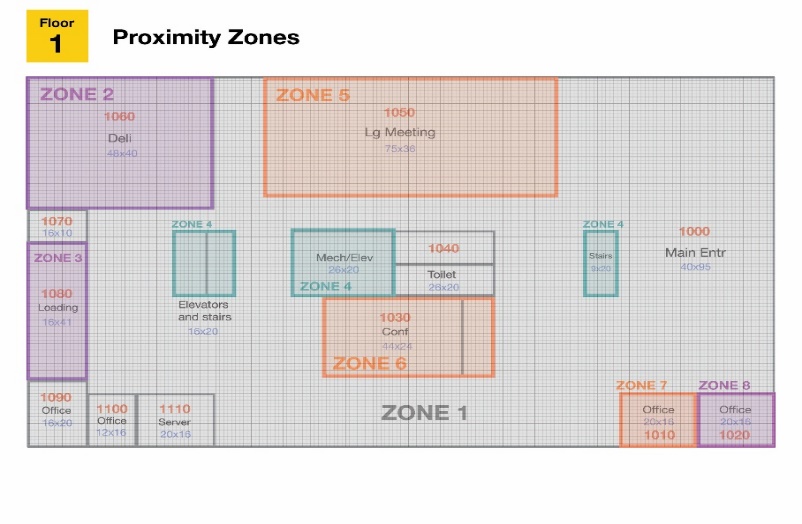
**Scenario:**After a kidnapping at GAStech’s Abila office in 2014, you have been hired as a visual analytics expert. In response to the kidnapping, GAStech’s security introduced a policy whereby all employees must wear proximity ( prox ) cards while in the building, so that they can be tracked. You have been hired specifically to analyze and understand this steady stream of data. The company needs your help in operational issues as well as security issues.Can you identify issues to safeguard the company’s employees ?

GAStech’sAbila office is divided into multiple HVAC (heating, ventilation, and air conditioning) zones. Each zone is instrumented with fixed sensors reporting building temperatures, heating and cooling system status values, and concentration levels of various chemicals such as carbon dioxide (CO 2 and hazium (a recently discovered and potentially dangerous chemical).

**Scenario : HVAC Zones**



Company staff are required to wear proximity ( prox ) cards and the building is instrumented with passive prox readers, covering individual building zones. Prox zones do not generally correspond to HVAC zones. When a prox card passes into a new zone, it is detected and recorded. Most, but not all, zones are open to staff members who forgot their prox cards. In addition, a robot travels the halls periodically and takes readings, identifying prox cards in the areas travelled by the robot.



As an expert in visual analytics, you have been hired to help GAStech understand its operations data. In this challenge, you are given two weeks of building and prox sensor data. Can you use visual analytics to identify typical patterns of and issues of concern?

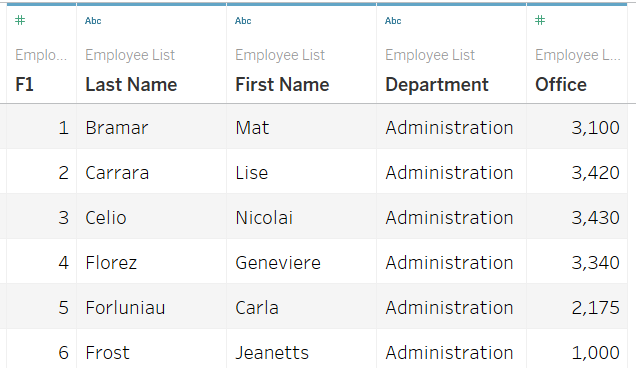
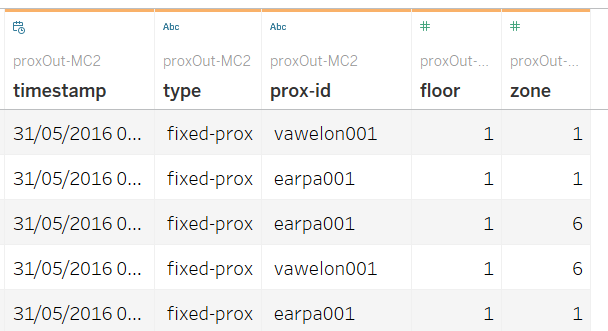
You will have the following data and supporting information at your disposal: (The **dataset** covers two weeks, May 31 to June 13, 2016.)

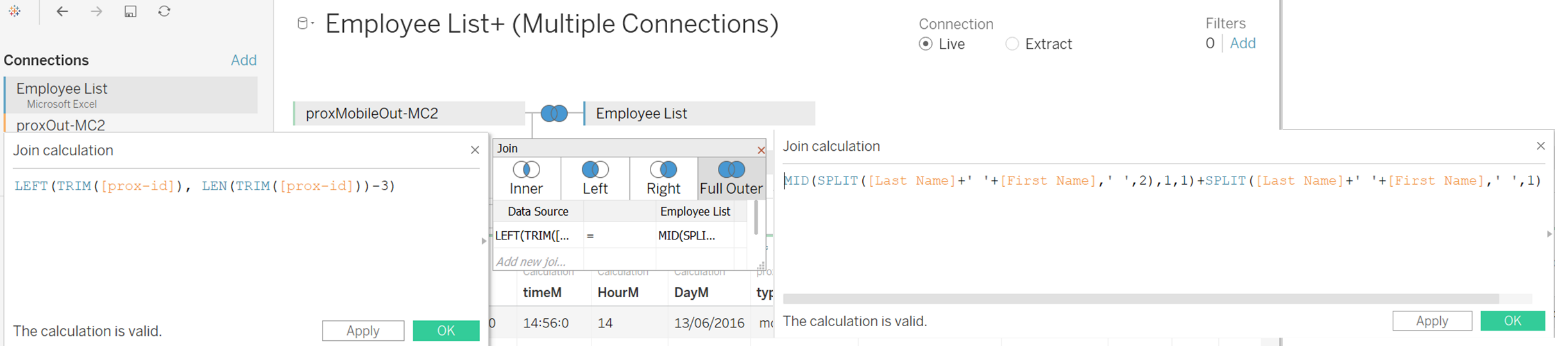
* A building layout for the GAStech offices, including the maps of the prox zones and the HVAC zones
* A current list of employees, roles, and office assignments
* A description of the data formats and fields provided
* Proximity sensor data for each of the prox zone regions
* Proximity sensor data from Rosie the mobile robot
* HVAC sensor readings and status information from each of the building’s HVAC zones
* Hazium readings from four sensors.

**PART 1: Joining sheets:**All the created worksheet workbooks using tableau were saved on the zipfolder attached.

1. **Joining Employee List & ProxOut data:**

The employee list dataset has the Last Name, and First Name of employee, while the ProxOut dataset has a pox-id, as shown below.



To combine the two datasets, the prox-id was generated from the employee dataset list. This was done through the formulas which join the first and last name and remove the last three numbers form the prox-id. In addition, to take into consideration some inconsistency in given prox-id, last names that have spaces have been split as well, where the first part was taken instead. As shown below.

1. Other dataset joining was done straightforward through using the timestamp.

PART 2: Use Visual Analytics to answer the following questions:

**Question 1 + 2**:

What are the typical patterns in the prox data? What does a typical day look like for company employees?

1. **Mobile ProxOut dataset:**

Using the Mobile ProxOUT dataset, the following visualization shown in Figure 1 was obtained by displaying the timestamps of each recording (to the right) and then grouped per hour (to the left). From this figure, we obtain the information that most employees start work at 9 am, and leave at 2 pm, however, there seem to be some 2 distinct count recordings outside of this pattern around 10 am, as shown clearer in Figure 2.

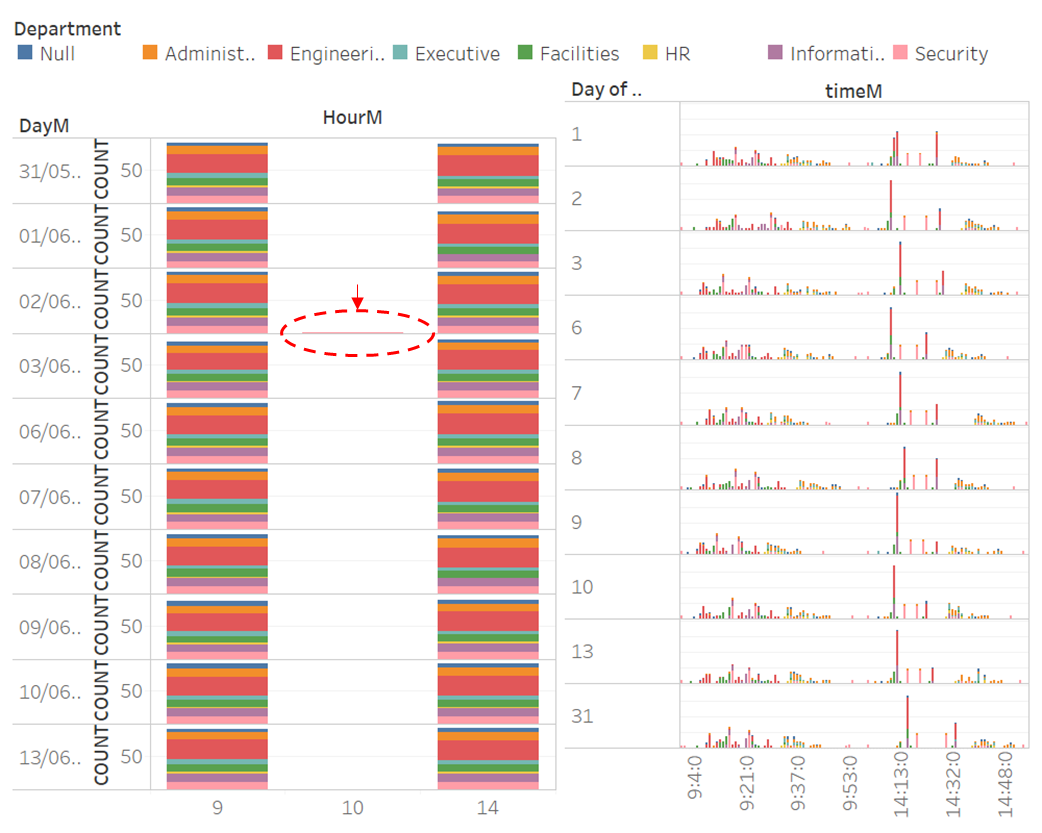


Figure :Mobile prox data – count of employees during the two weeks colored by departments. (a) per hour. (b) per second.

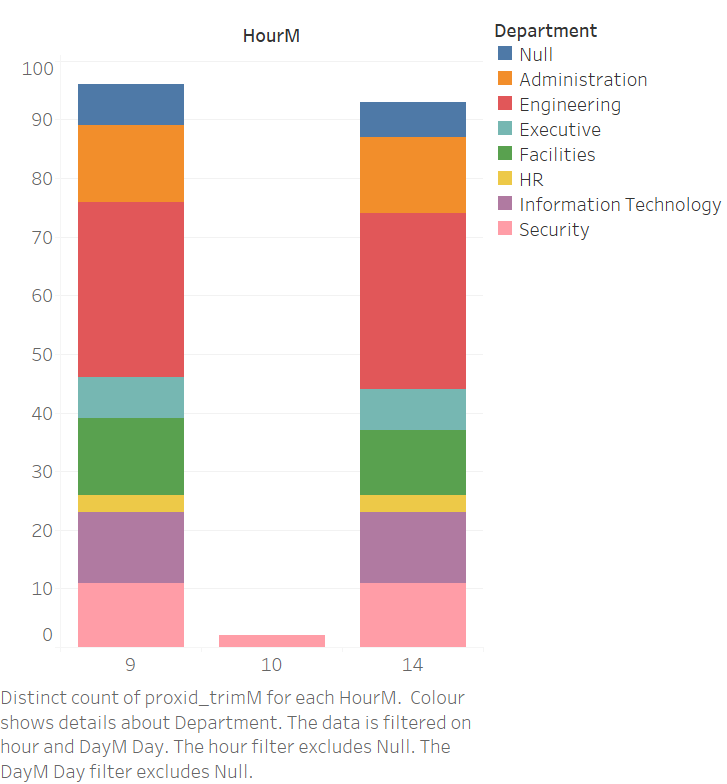


Figure : Mobile prox data – count of employees during the total two weeks colored by departments.

1. **Fixed ProxOut dataset:**

Figure 3 below explored the fixed-prox recorded data count of employees during the two weeks, with annotation in red of suspected unusual events that will be discussed in the next question. The color is providing a label of the department for each count each employee's timestamp. It is noticed from the figure that employees working hours are from 7 am to 5 pm. While on the Saturday 5th, Saturday 11th and Sunday 12th of June, the distinct count of people in the building is very low, which indicated that Saturdays might be an official weekly day off, whereas Sunday 12th June might have been a holiday for the company.

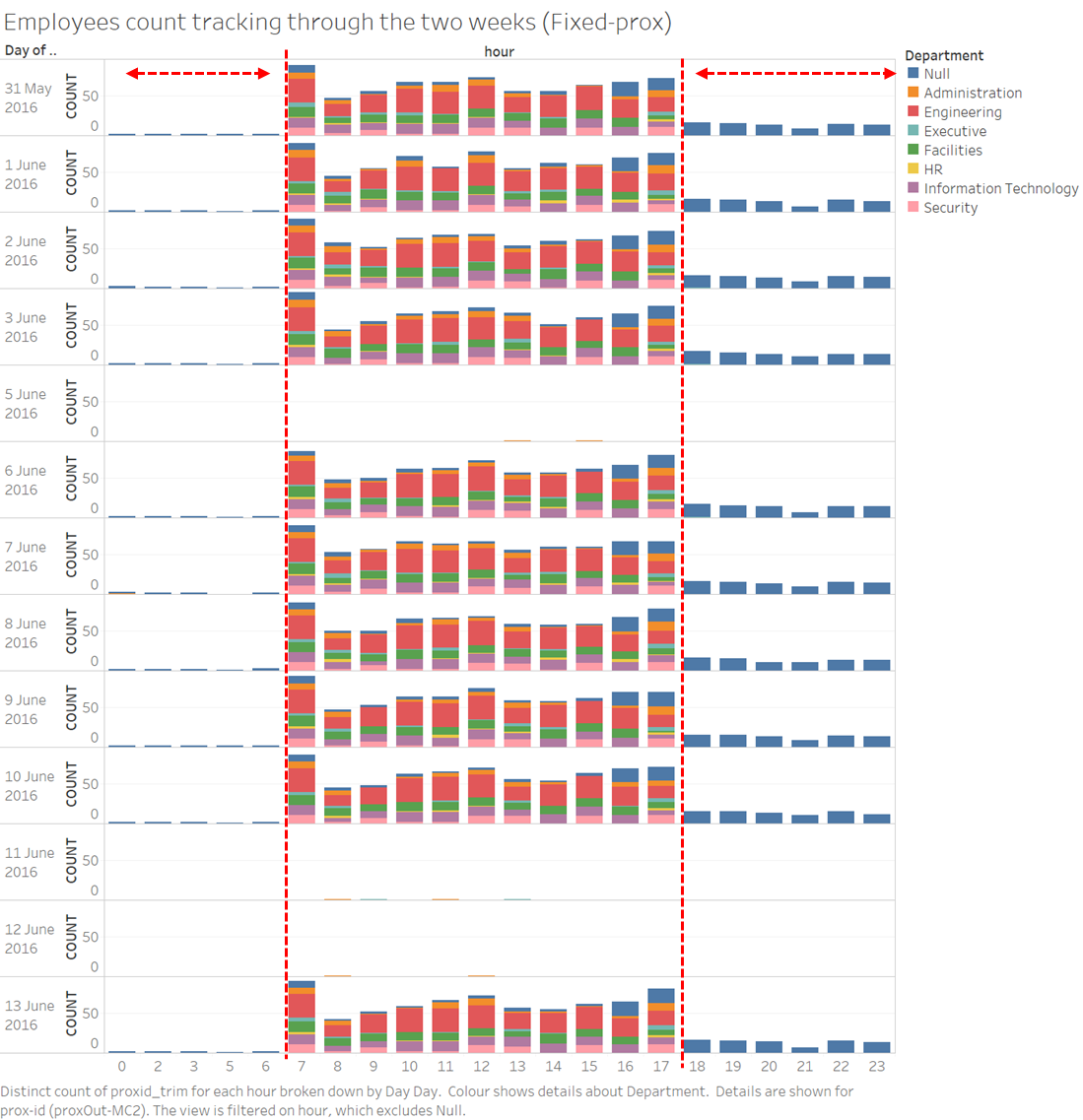


Figure : Fixed prox data – count of employees during the two weeks, with annotation in red of suspected unusual events discussed in question 3.

Figure 4 and Figure 5 below show the timestamp of each department employees grouped by zone and floor information. This visualization helps in identifying employees per department present in each zone of the three floors during each hour of the two weeks recordings, making it easy to visually track the movement of the employees between floors and zones. Taking the zoomed-in example in figure 4 below, we can see the following employee from the Administration department.

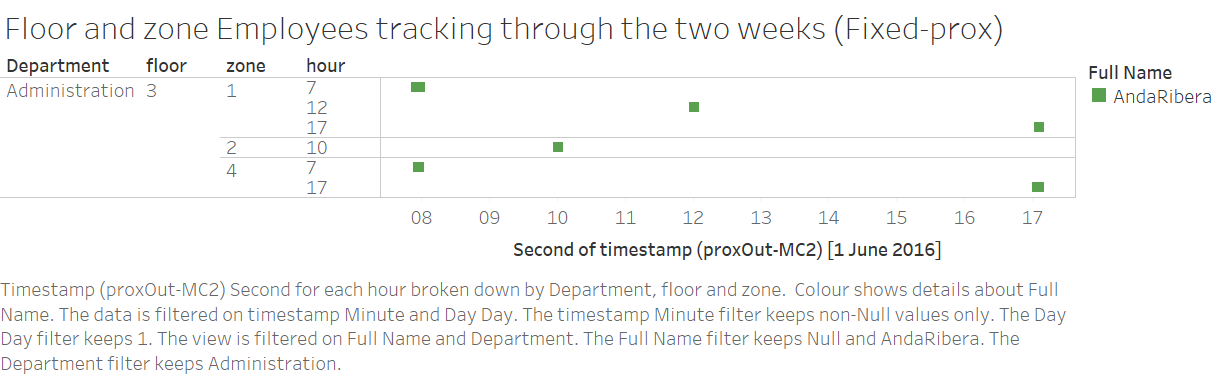


Figure : The movement of "Anda Ribera" Administration department employee around floor 2, on 31st of may 2016.

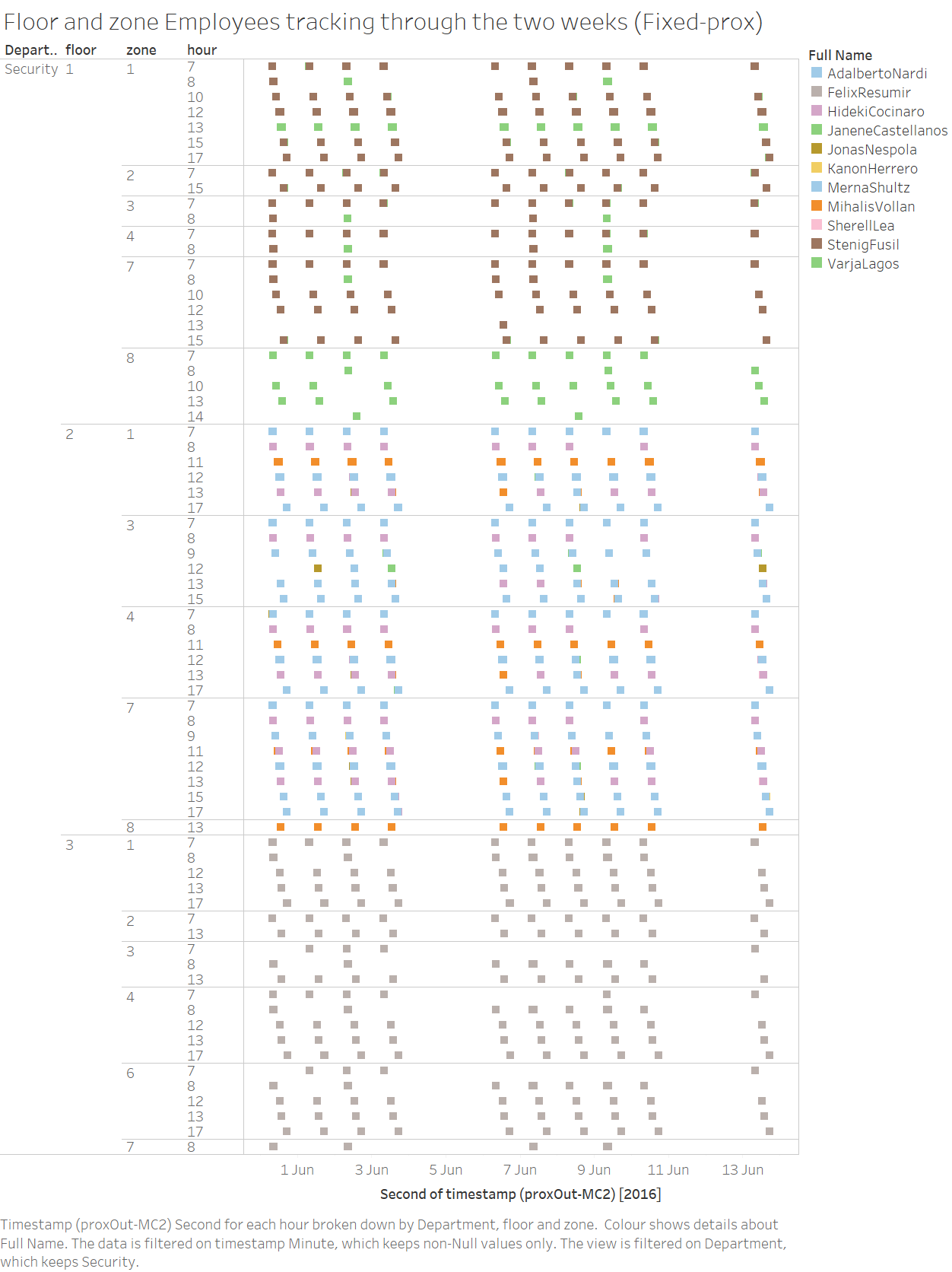


Figure : Security Department Employees tracking through the two weeks (Fixed-prox).

**Question 3:**

Describe up to ten of the most interesting patterns that appear in the building data. Describe what is notable about them and explain their possible significance.

1. **Return outlet Co2 concentration recordings:**

Figure 6 shows the average co2 concentration of each zone per the 2016 year, which gives a picture of the overall zones concertation in the building during the recordings two weeks.

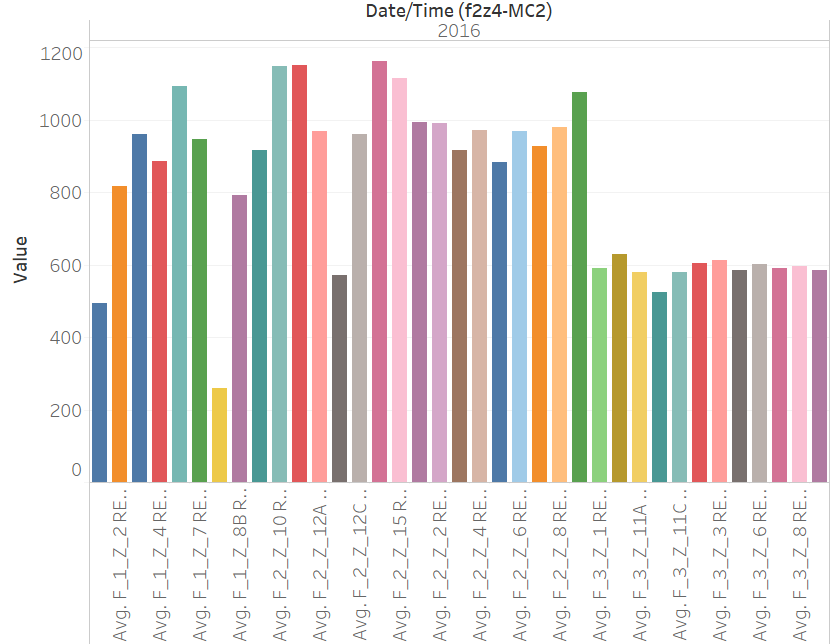


Figure 6: the average of all the zones co2 recordings.

The following three figures are obtained through plotting the line and area of co2 concentration recordings versus each day of the two weeks, where hours during the day are shown using colour label.

Floor 1 Co2 recordings patterns:

1. Zone 2, zone 3, zone 4, and zone 5, and zone 7 seem to share the same pattern happening on the 7th, 8th and 9th of June 2016, as annotated below. This could be an indication of the everyday normal pattern happening on the floor.
2. Zone 1, Zone 2, Zone 3, Zone 5, and the both of Zone 8 have the same patterns happening on several days ( 1st until the 5th June), however, some zones have higher levels of concentration for the same pattern on these days. Higher concentration zones could be indicating that unusual events are occurring in those specific zones, which needs further study and monitoring by managmnet to ensure the building’s safety.

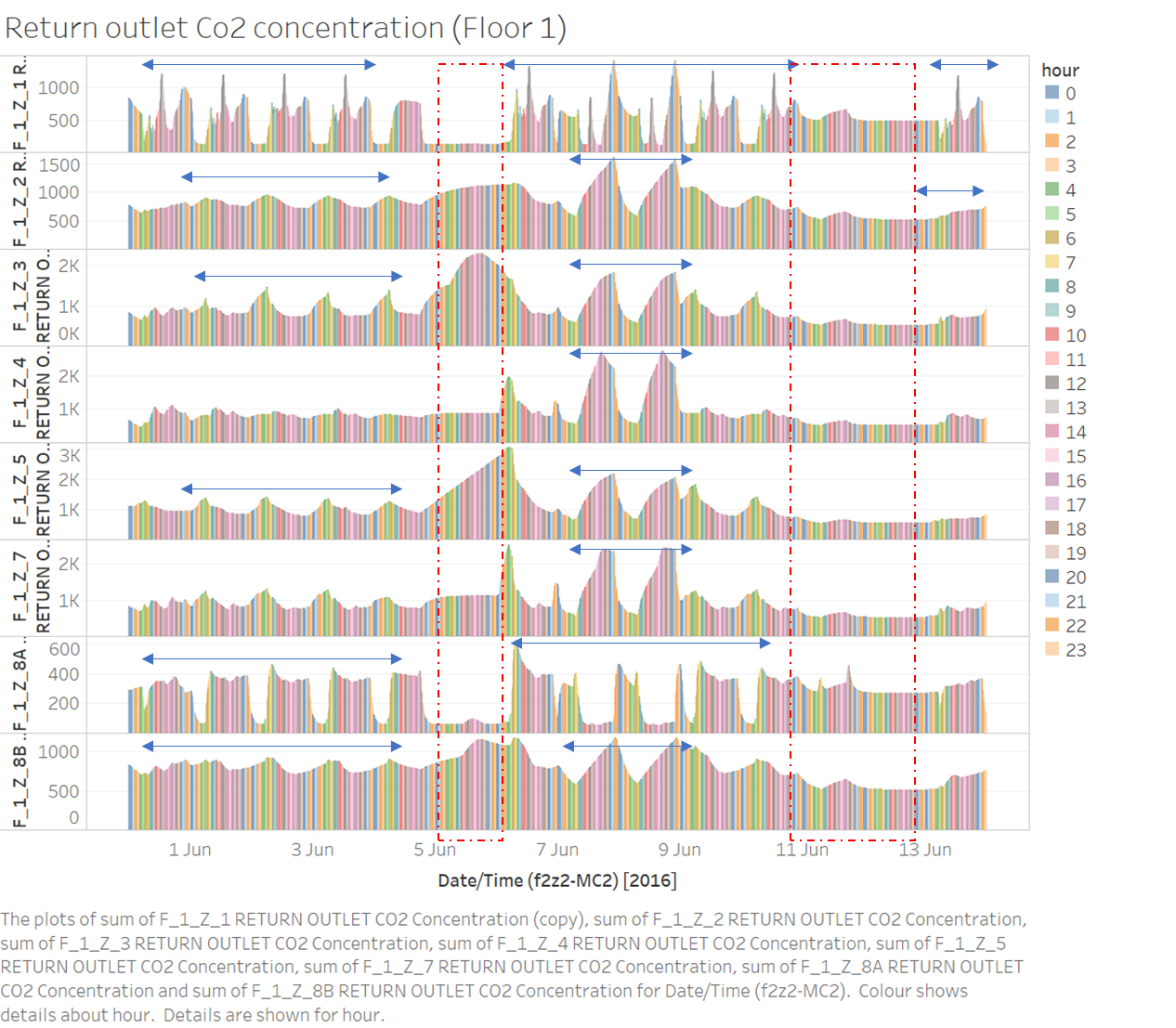


Figure : Return outlet Co2 concentration (Floor 1), annotated in blue are noticed pattern, and annotated in red are noticed anomalies for question 3.

Floor 2 Co2 recordings patterns:

1. Floor 2 recordings, as shown in Figure 8 below have the same pattern happening throughout the day (increasing then decreasing) for all of the zones, however, some days show higher concentration levels of the pattern than other, as expected previously from figure 9 which showed all zones averages throughout the year.
2. On the 7th and 8th of June, there seems to be a high concentration pattern happening among all of the zones.
3. On the 5th, 11th, and 12th of June concentrations are relatively low with no peaks, which is aligned with the previous assumption that these days are off days for company employees.

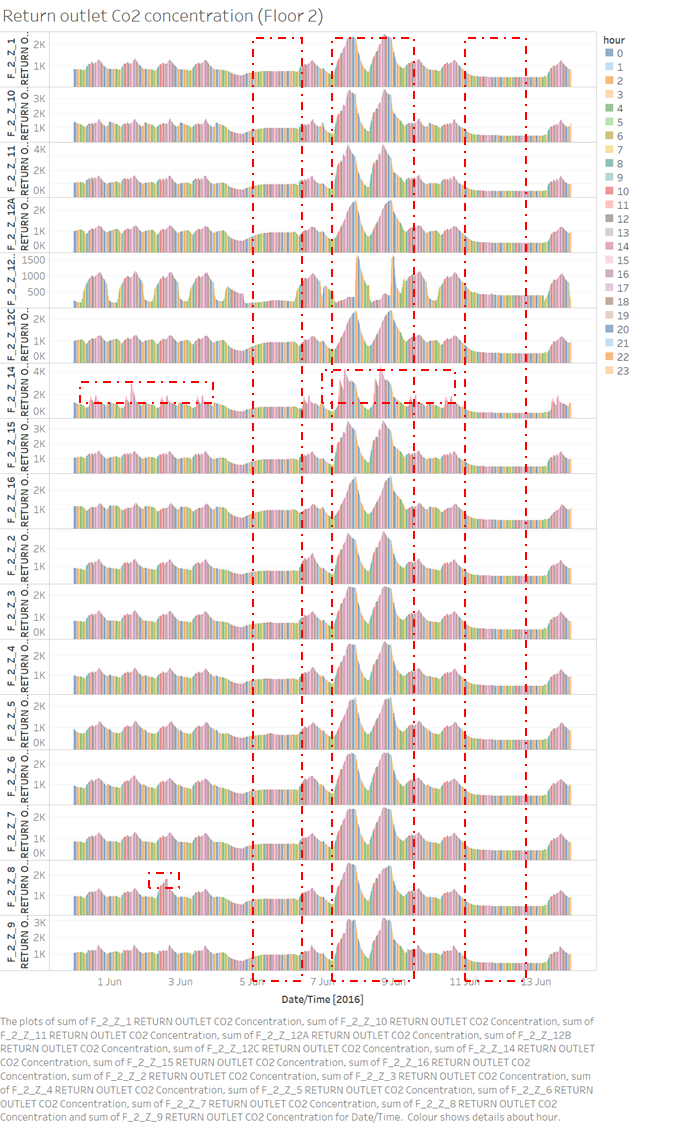


Figure 8: Return outlet Co2 concentration (Floor 2) and annotated in red are noticed anomalies for question 3.

Floor 3 Co2 recordings patterns:

1. Out of the three floors, Floor 3 has the most common patterns with all days and zones, and relatively similar concentration levels (1000-1500). There exist some anomalies to the patterns as annotated in red, which will be addressed in the next question.
2. On the 5th, 11th, and 12th of June concentrations are relatively low with no peaks, which is aligned with the previous assumption that these days are off days for company employees.

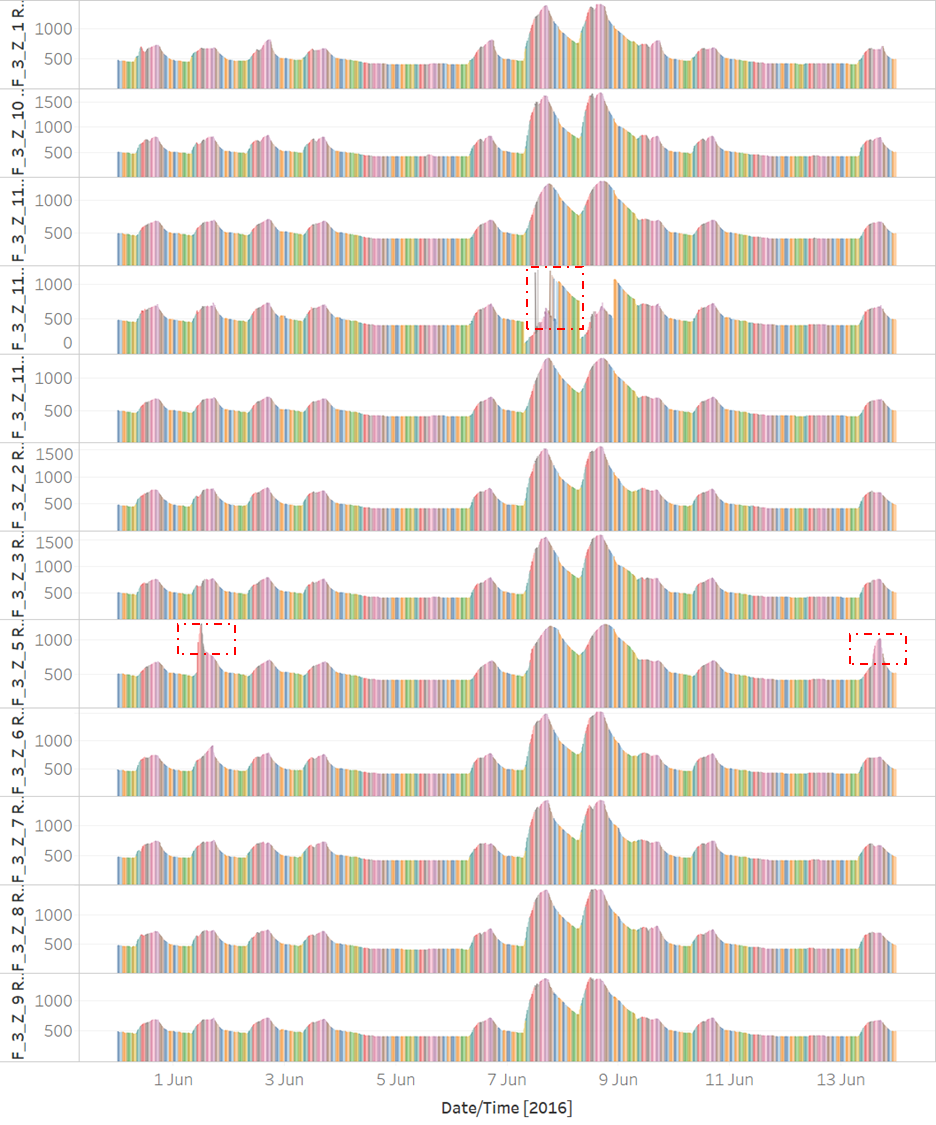


Figure 9: Return outlet Co2 concentration (Floor 3) and annotated in red are noticed anomalies for question 3.

1. Total Electric Demand Power:
2. Total Electric Demand Power for each day of the two weeks is shown in the figure below. As annotated in blue, there seems to be a pattern happening with three peaks during the day around 9am, 1pm, and 5pm. On the 5th day which is suspected to be a day off from other results, the electric demand power is relatively low with no peaks occurring. These mentioned patterns in blue are assumed to indicate the everyday normal pattern of electricity demand.

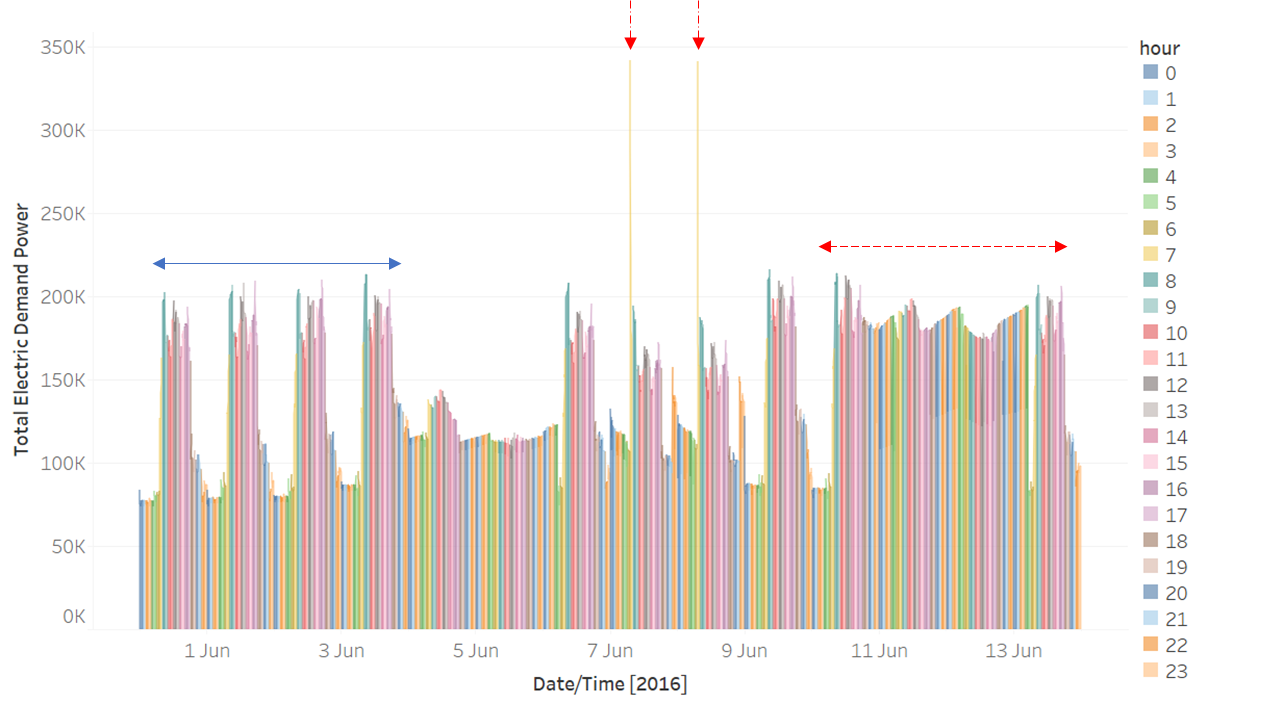


Figure : Total Electric Demand Power for each day of the two weeks. annotated in blue are noticed patterns, and annotated in red are noticed anomalies for question 3.

1. HVAC Electric Demand Power:

9. Taking into consideration the electrical demand power coming from the HVAC system gives a better idea of patterns happening. As visualized in the figure below, the HVAC systems seem to be the source of the suspected to be normal patterns mentioned on the point above.



Figure : Total and HVAC Electric Demand Power for each day of the two weeks. annotated in blue are noticed patterns, and annotated in red are noticed anomalies for question 3.

1. Thermostat Temp:

Thermostat Temp recording is the temperature of the air inside the zone.

1. Thermostat temp of floor 1,2 and 3, visualized in figures below, are showing a pattern in which there are peaks at when the temperature of the zones increases to high values (maximum of 29 degrees Celsius), then goes back to its steady-state.

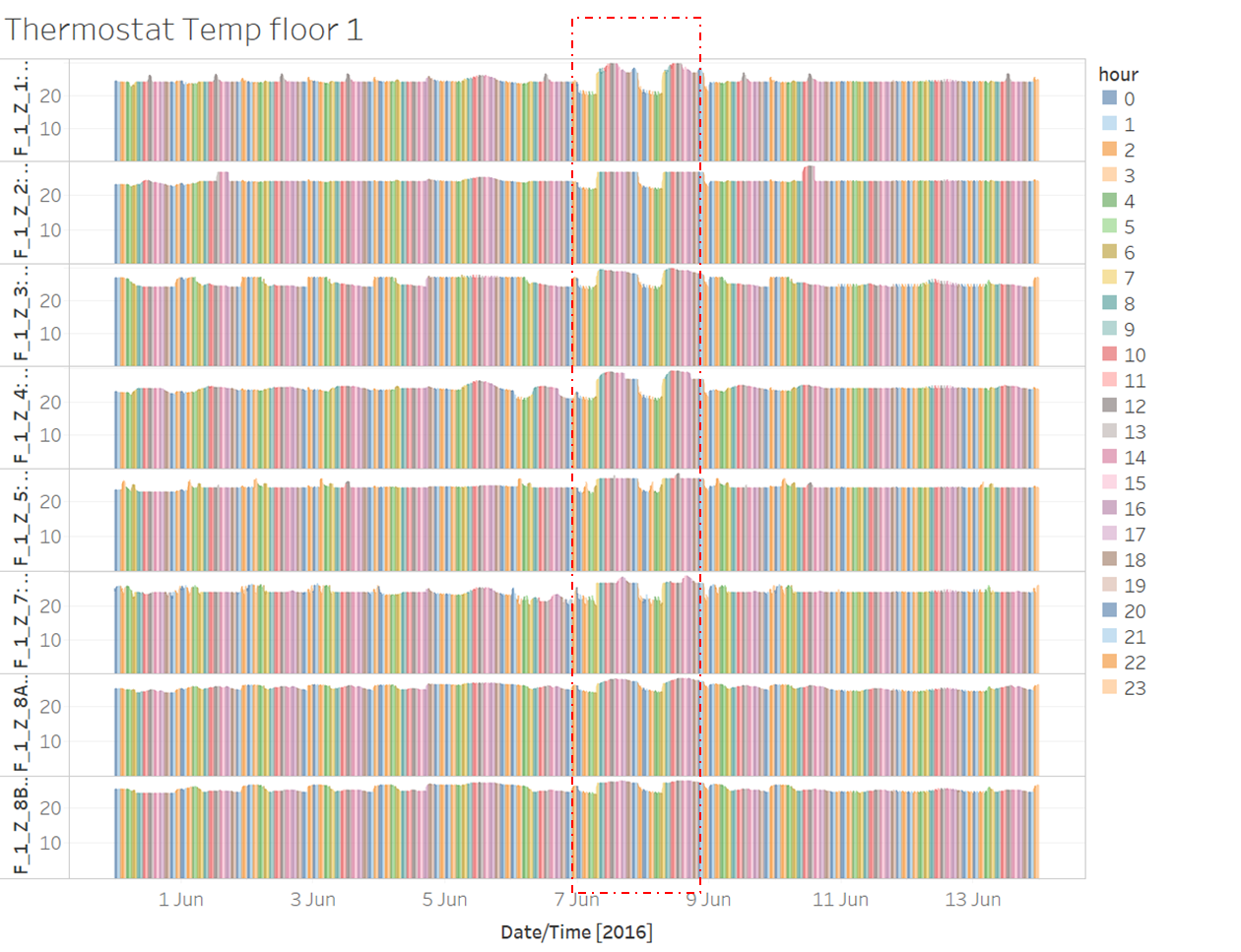


Figure : Thermostat Temp floor 1 for each day of the two weeks. Annotated in red are noticed anomalies for question 3.

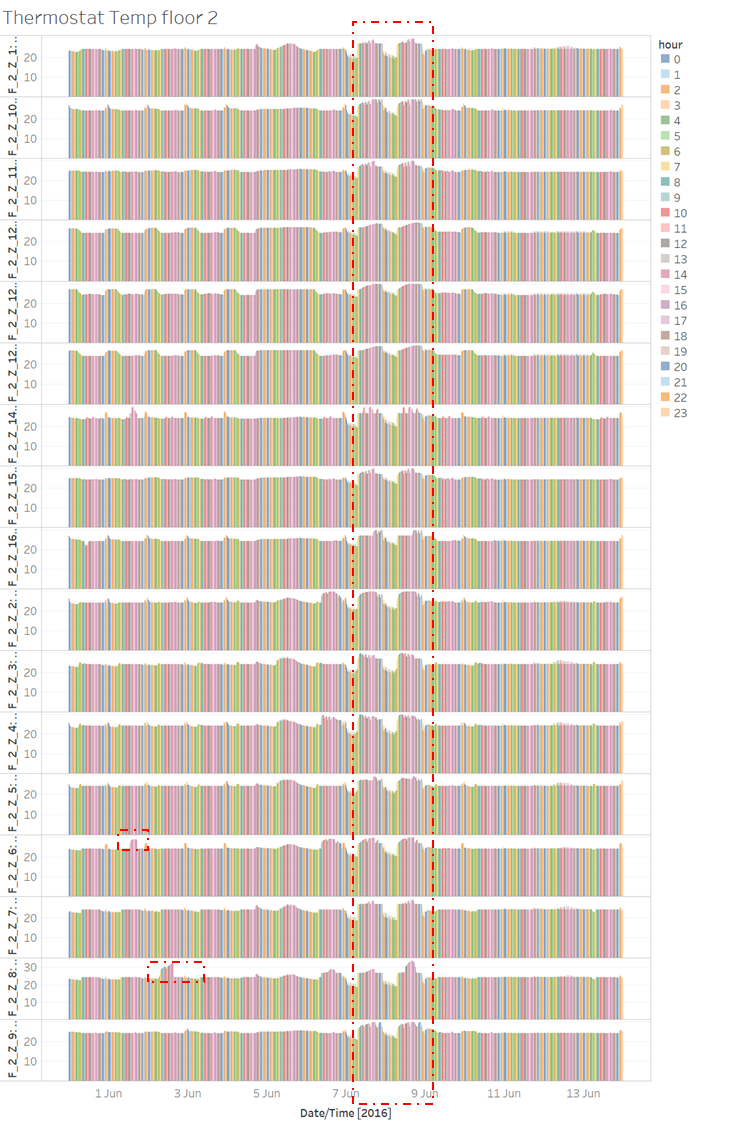


Figure : Thermostat Temp floor 2 for each day of the two weeks. Annotated in red are noticed anomalies.

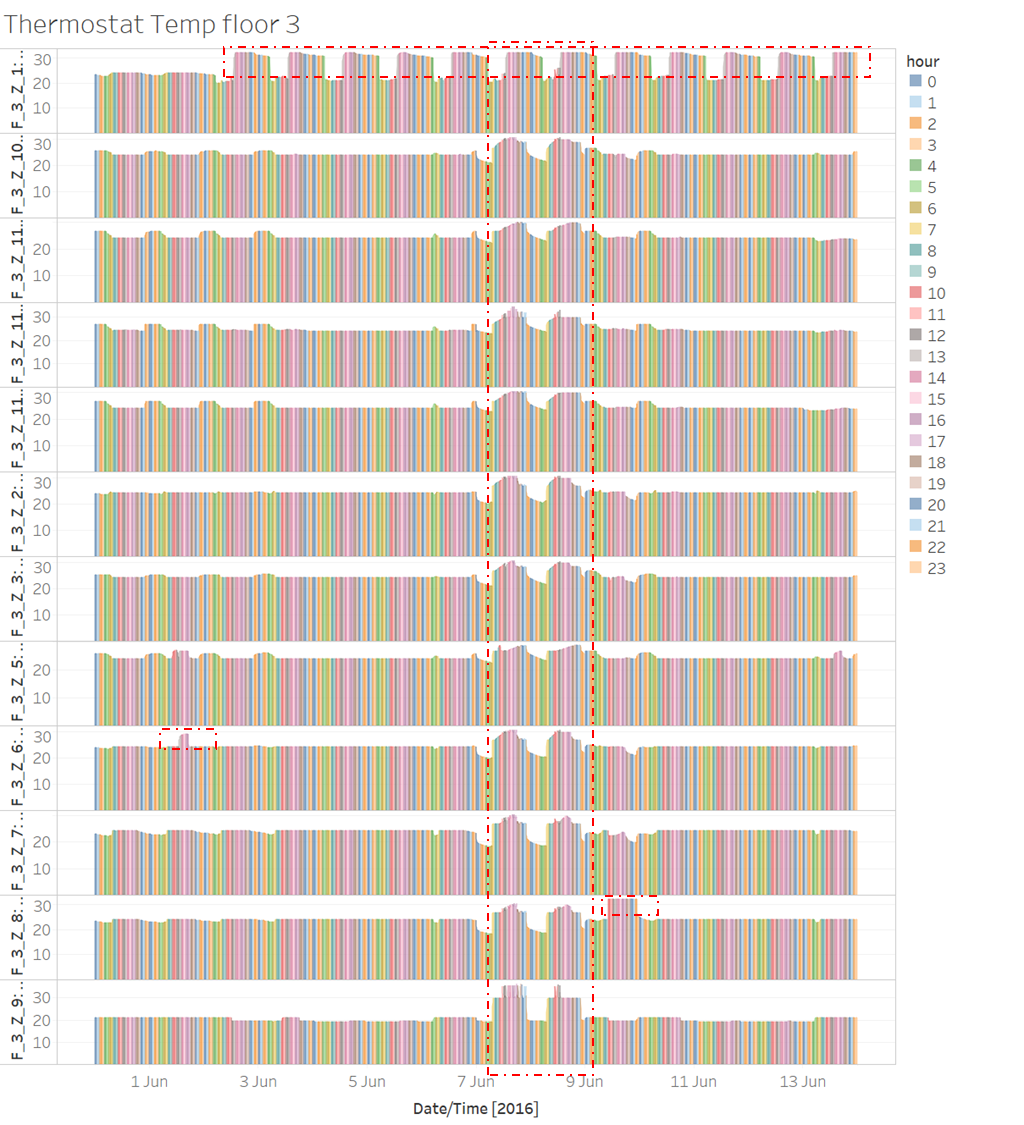


Figure : Thermostat Temp floor 3 for each day of the two weeks. Annotated in red are noticed anomalies.

**Question 4:**

Describe up to ten notable anomalies or unusual events you see in the data. Prioritize those issues that are most likely to represent a danger ( hint: see hazium ) or a serious issue for building operations.  
  
Hazium concentration:   
The following figures show a visualization using heatmaps of the 3 floors hazium concentration measurements average per hour for each the 2 weeks days.

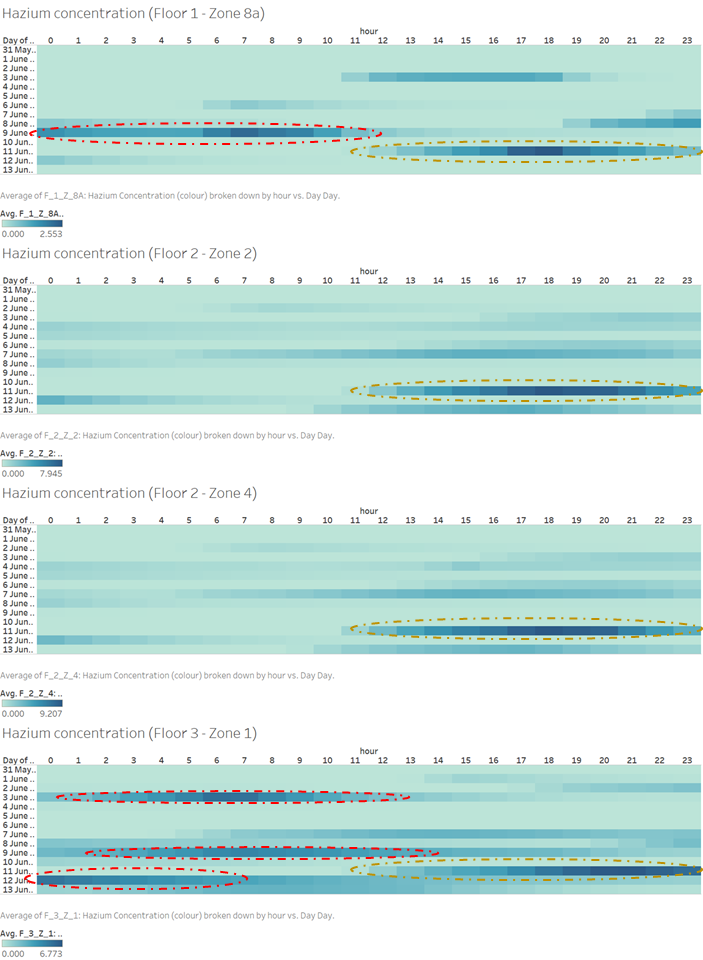


Figure : Hazium concentration annotated with gold for common patterns, and red for potential security threats: a)Floor 1 - Zone 8a b) Floor 1 - Zone 8a c) Floor 2 - Zone 4 d) Floor 3 - Zone 1.

**Anomalies or unusual events from Hazium data:**

1. Figure 16 shows the hazium concentration on the three floors of the building. It is noted that on the 11th of June 2016 and on all of the floors, the hazium concertation starts increasing from 12 pm and reaches to its highest level around 7 pm, and then it reduces a bit. This could be an indication of some issues happening in the building on that specific day.
2. As annotated in red, in Figures 11 (a) & (d), it seems that Floor 1 zone 8a and floor 3 zone 1 are presenting high concentrations on the 9th of June 2016. The high concentration first appeared on zone 8a on the first floor from 12 am at midnight and it continued until around 10 am before the concentration level decreased. Around 3 am the concentration on floor 3 increased. This could be an indication of some leakage issues happening between the building floors which could be an issue that could put employees' safety in danger.

**Anomalies or unusual events from Building data:**

1. Figure 7, shows the co2 concentration in floor 1, and unlike all other patterns which show resemblance indicating that on the 5th, 6th, and 11th of june 2016, the building has a very small count of people inside, the co2 in the floor on the 5th of June number shows high values of concentration. This indicates unusual events happening on the 5th of June in floor 1 zones 3,5,7, and 8b.
2. Figure 8 shows the concentration in floor 2, there seems to be a high co2 concentration pattern happening among all of the zones on the 7th and 8th of June. This is a strong indication of an unusual event happening on floor 2 on these two days.
3. Figure 9 shows the co2 concentration in floor 3, this floor has patterns occurring among all zones and days. However, as annotated in red in the figure, there seem to be some unusual co2 high peaks on the 2nd of June on zone 4 and 5, also on the 7th of June on zone 11. These high peaks are standing out in the patterns and might be an indication of unusual activities or building issues.
4. Total electricity demand, visualized in figure 10, shows that on the 7th and 8th of June 2016, there were high peaks both happening at the exact same time 7:05 am, on different dates. This could either be due to some building machine operational issue or due to some unusual activity that has happened on those two consecutive days.

**Anomalies or unusual events from Prox data:**

1. From figure 1 and Figure 2 of the mobile prox data, it is noticed that the robot made its rotation at 9 am and 2 pm during each day of the two weeks. Having the robot rotations timing known, or easily predictable, can be considered as a security issue.
2. Figure 1 (a) also shows that only on the date of 6th June 2016, the robot has faced a few seconds delay, and recorded two employees. This delay can be considered as a possible indication of a security threat that might have occurred.
3. Figure 3 shows the count of employees recorded by the fixed prox sensors. It is noticed that most of the employees are present between 7 am-6 pm. Unexpected recordings are present from 6 pm-7 am, considering these are not usual working hours, while the employees count recorded is less than 40, and further less than 10 between 12 am-6 am. This raises the question of whether these late recordings were authorized by the company or not.
4. From Figure 1, Figure 2, and Figure 3 it is noticed that the three days 5th,11th and 12th of June 2016 have very small employees count. In the case where these days were holidays, the presence of those employees on these three off days could be questioned if it led to any security issues.
5. From figure 5 of the Security Department Employees tracking through the two weeks (Fixed-prox), it is noticed that some zones are less visited by the security, which might indicate a security threat.

**Question 5:**

The major hazium incidences happened on 3rd, 6th and 11th June. Although the major incident happened on 11th June, there is not much proximity card data available for analysis as none of the mobile prox card data was collected on 11th and 12th whereas fixed prox card data for those days are rather limited. Prox card data collected does not seem sufficient to prove if any employee was responsible for the hazium levels.

On the other hand, I am more concerned as to the variation in HVAC settings mentioned above, which I believe is unlikely to be the case. Perhaps someone was able to manipulate it over the server/remotely to affect changes to induce the increase in Hazium by affecting the air flow and temperatures.