## Entry Name: Laxman Sai Karthik Mahidhar Gopalabhatla

## Student ID: 800936333

## ITCS 6162 Summer 2016 Final Project Deliverable 2

### Tools Used:

*Tableau, Excel, D3.js*

**Approximately how many hours were spent working on this submission in total?**

*40-50 Hours*

**What are data mining techniques you used and why did you choose them?**

*The most common technique that I have used for my analysis is the visual pattern recognition. One of the most important technique that I have used is the correlation analysis to find out what parameters vary with change in what other parameters and have tried to make a valid correlation among the similar variables.*

**What are the limitations of your approach?**

*As the building data has more than 400 variables for consideration, it becomes very difficult to process them as the comparison of each and every variable with all the other available variables would result in a minimum of 400! (Factorial of 400), which is a very large number and it is highly impossible to correlate every variable with other rather I have depended on visual perception of the human in detecting the obvious patterns in the time series data.*

*I have extensively used Exploratory data analysis techniques in making sensible correlations in the provided dataset. The Multivariate relationships between different listed parameters were uncovered with these techniques.*

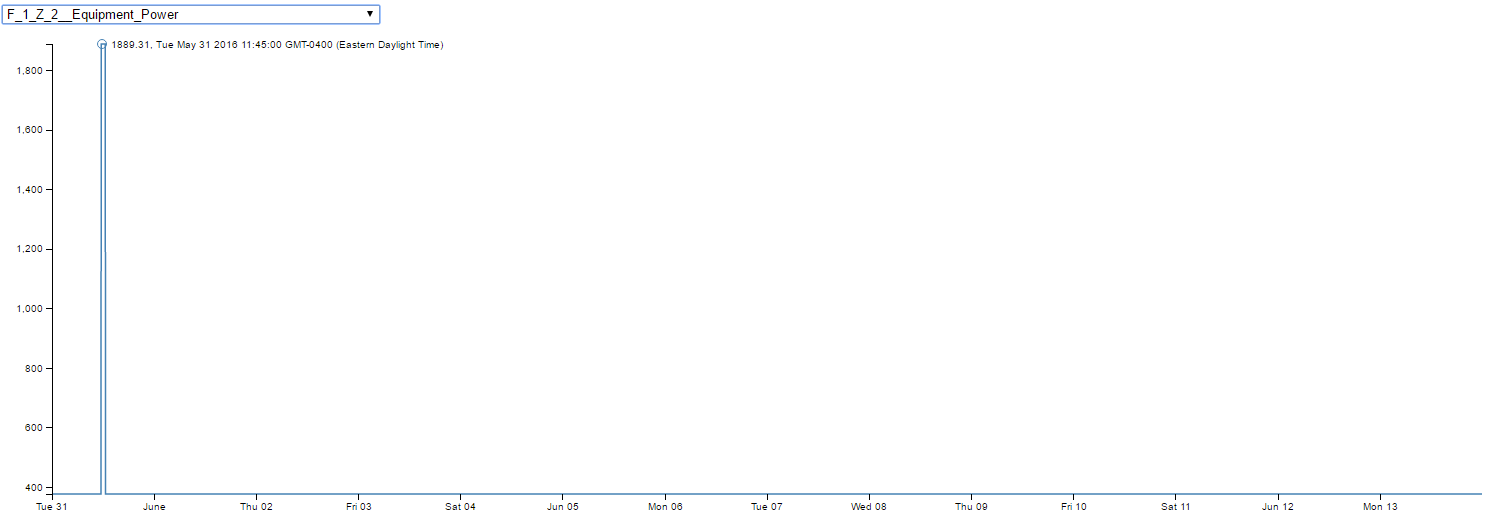
*As mentioned earlier the timestamp data in the proximity data is a point it is hard to understand the duration of an employee at a particular location.*

**Questions**

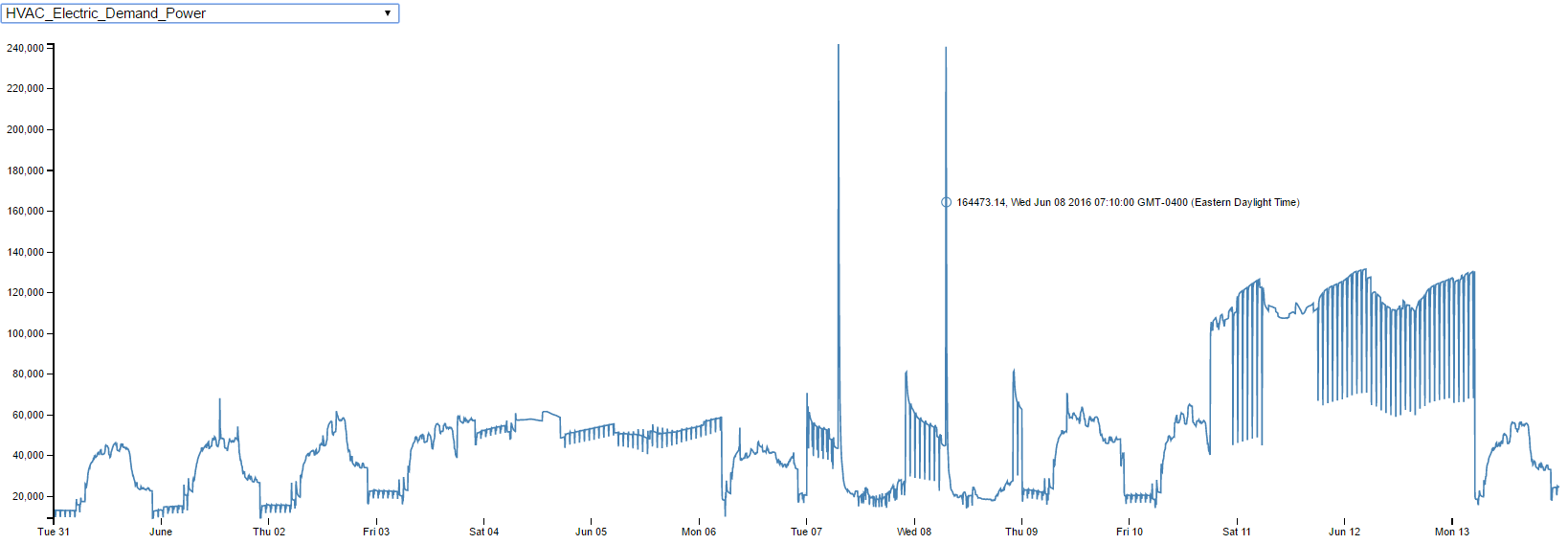
**2.a** – Describe up to ten notable anomalies or unusual events you see in the data. Describe when and where the event or anomaly occurs and describe why it is notable. If you have more than ten anomalies to report, prioritize those anomalies that are most likely to represent a danger or serious issue for building operation.

There were nearly 25 anomalies that surfaced during our analysis of the data. The below mentioned are the top 10 prioritized anomalies which stand out.

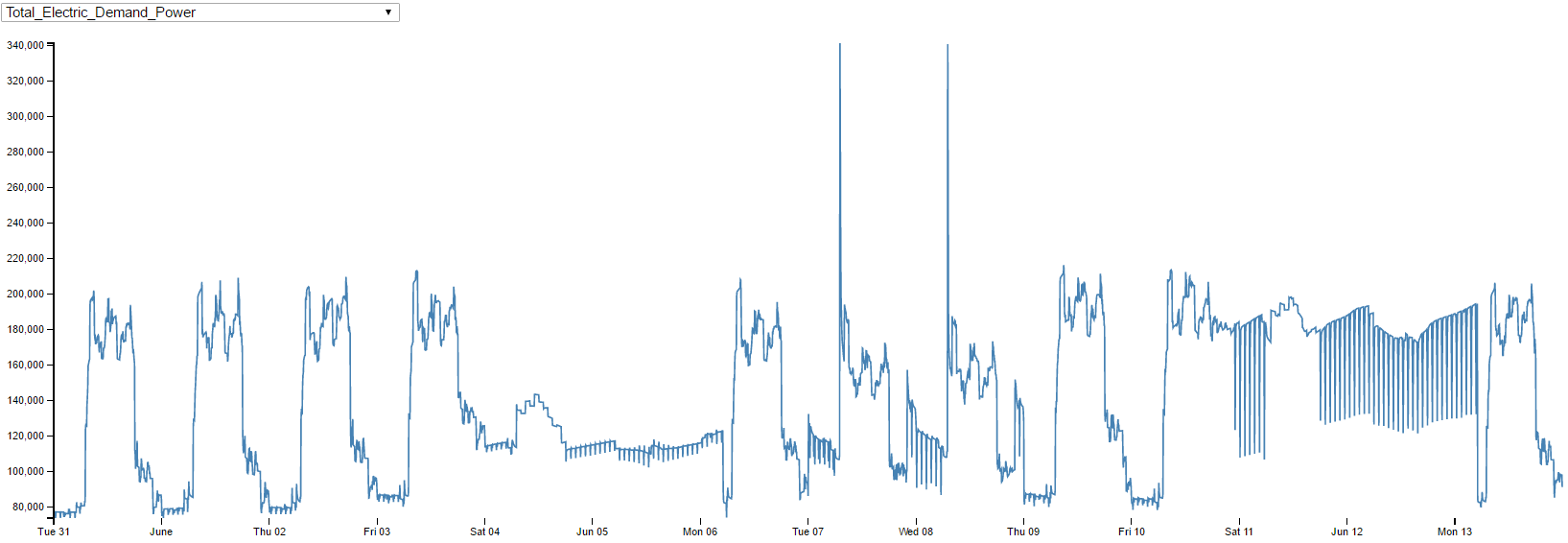
1. Floor 1 Zone 2 power shows a drastic increase during the time 11:45 AM to 12:15 PM on 31st of May. During this period the equipment power is at a maximum of 1889.31 and during rest of the time it is at a constant power of 377.86



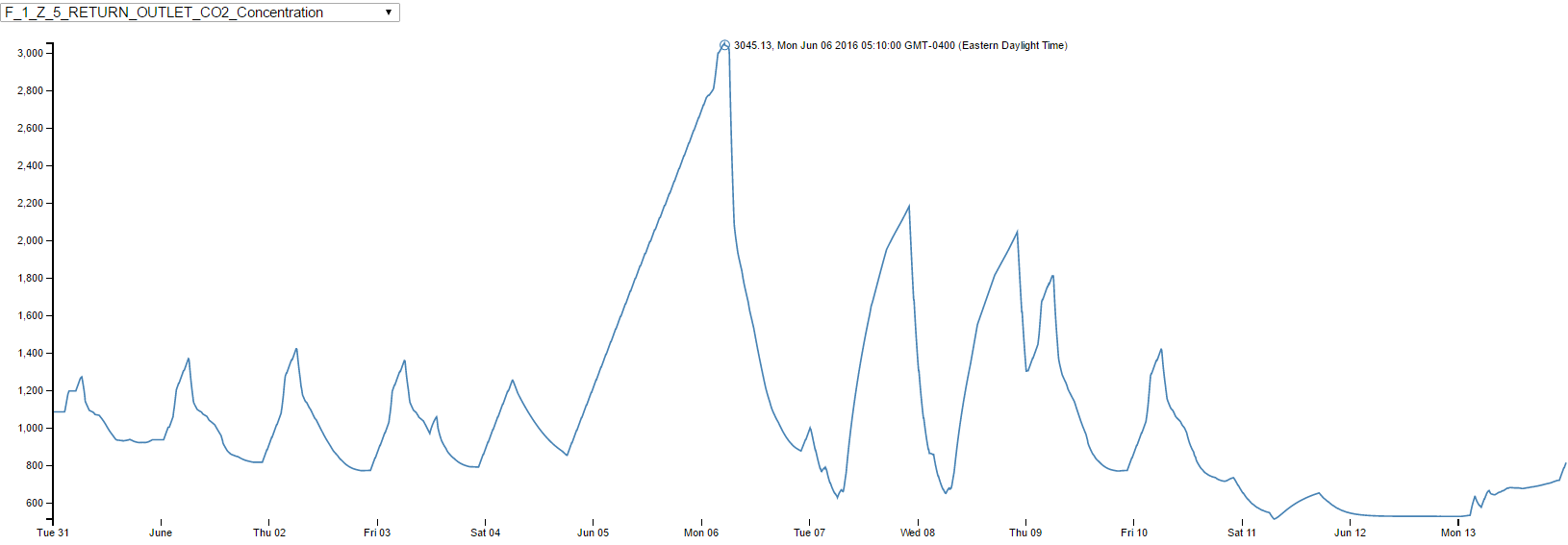
1. During 7thand 8th June from 7:00 PM to 7:45PM, there is a sudden spike in HVAC system’s power consumption.



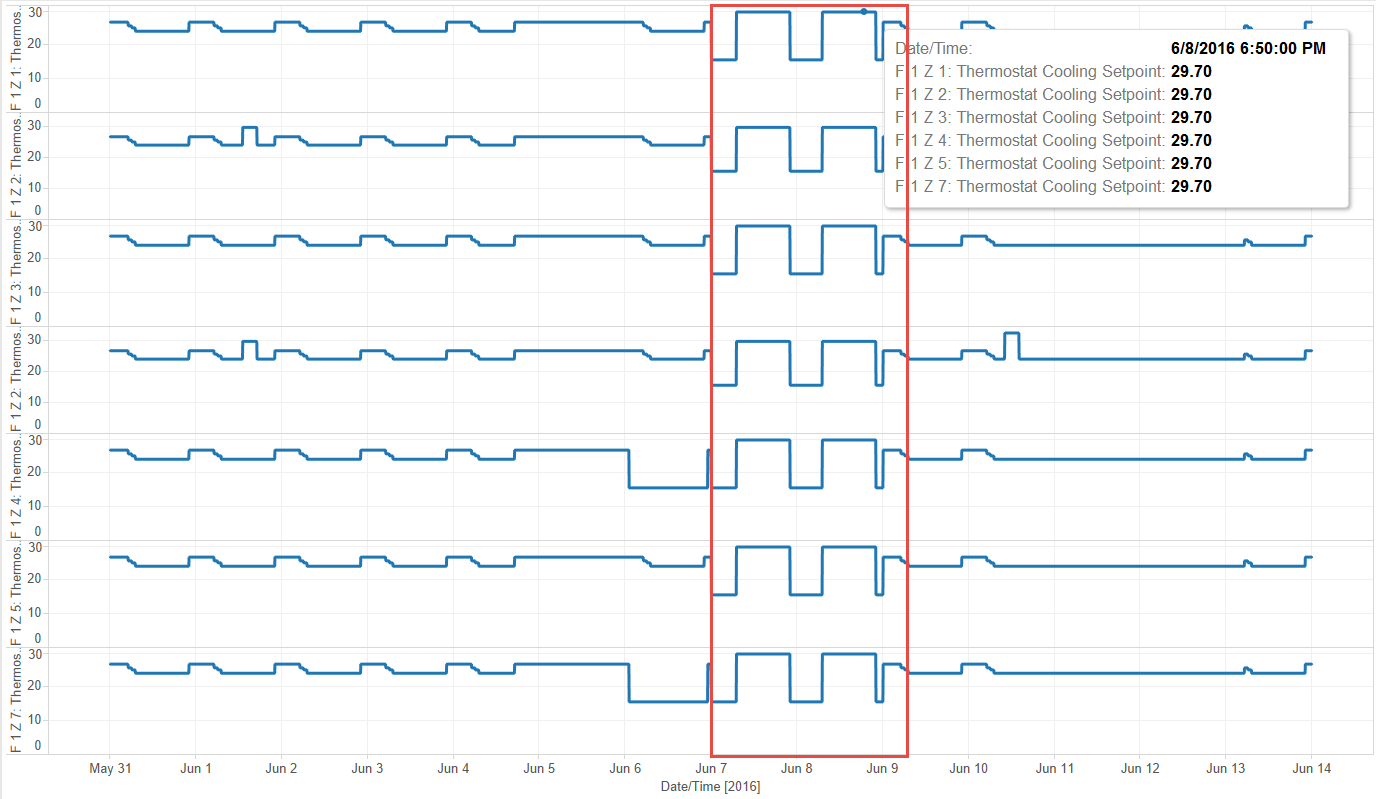
1. The increase in HVAC system’s power consumption has triggered the Total power consumption to shoot up during the same time period.



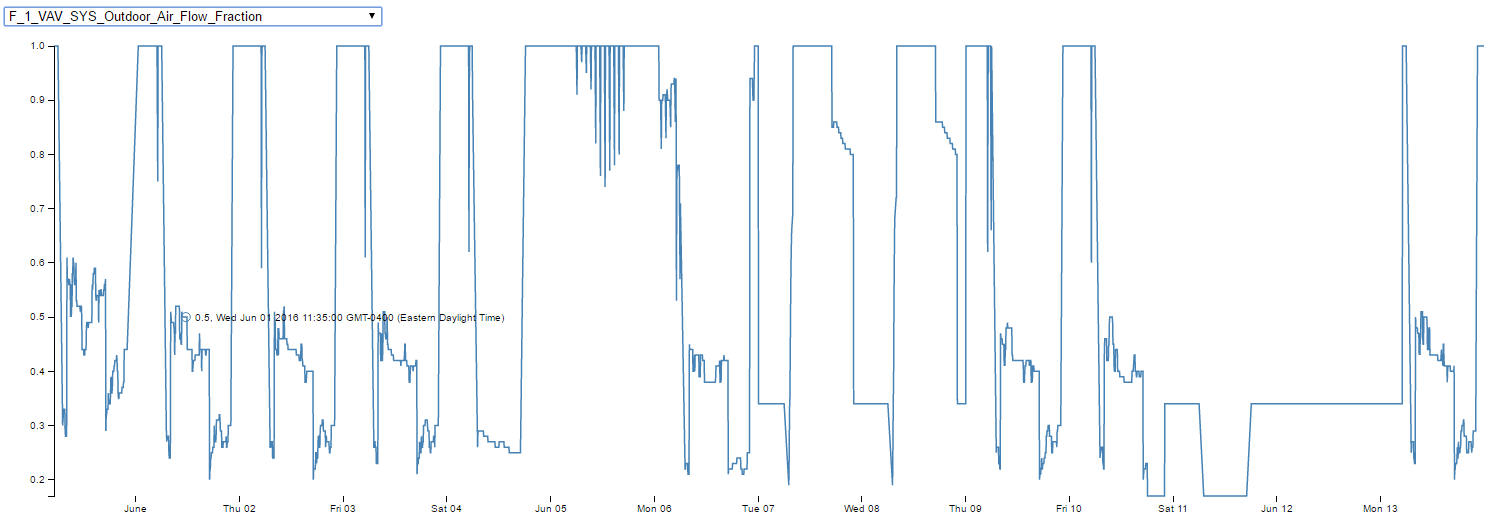
1. Concentration of C02 measured at the Floor 1 Zone 5's return air grille suddenly reached a maximum of 3045.13 during Jun 05 at 5:10 AM and followed an overall downward trend after that till reaching a bare minimum sometime during June 13th. There is absolutely no employee activity near the zone and this activity cannot be explained with certainty.



1. The overall cooling and heating set points of all the zones were set to a different temperature rather than being set to the regular temperature during the 7th June 12:00 AM to 9th June 12:00 AM.

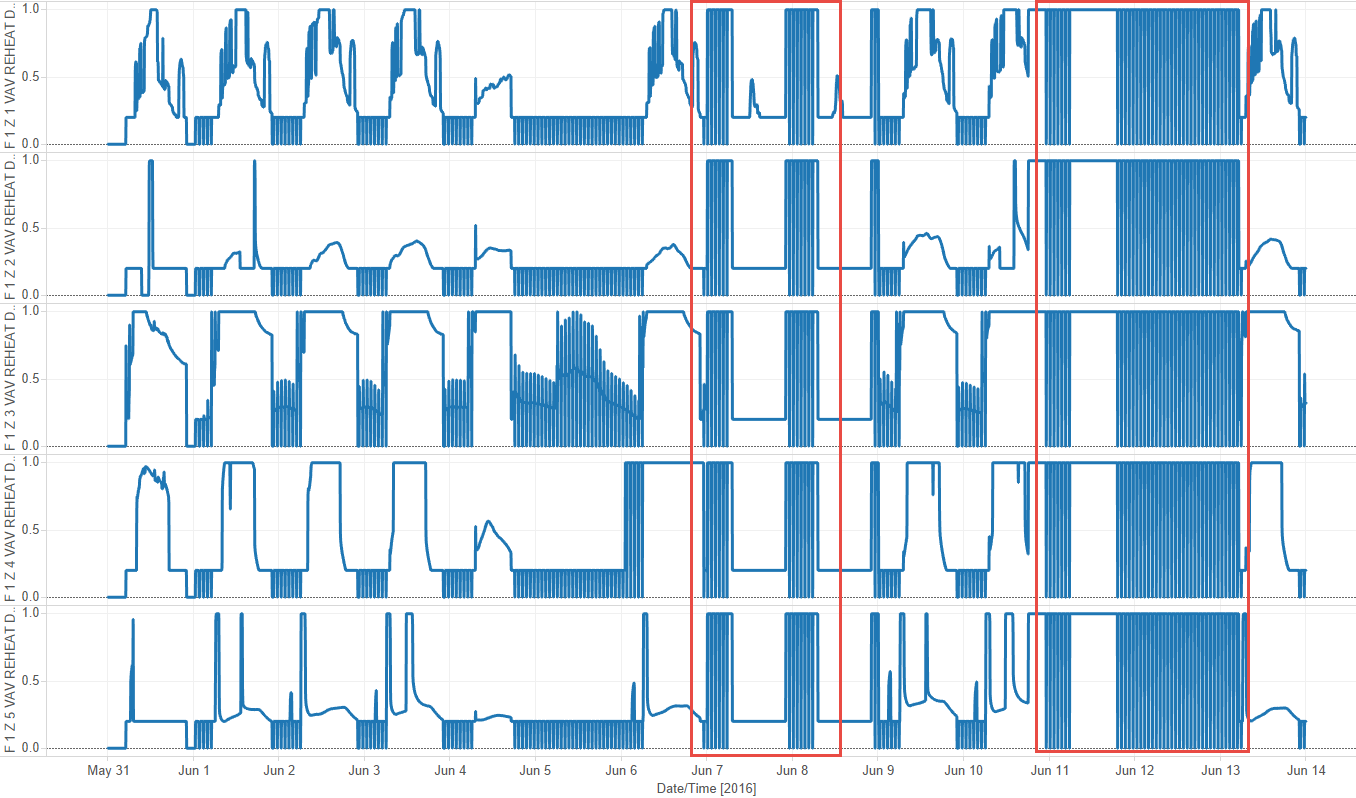


1. Percentage of total air delivered by the HVAC system that is from the outside very low from 11th June to 13th June. But the Flow rate of outside air entering the HVAC system didn’t show any fluctuations from usual as the damper position is set to 1 indicating maximum air flow which compensates for reduced outside air circulation.

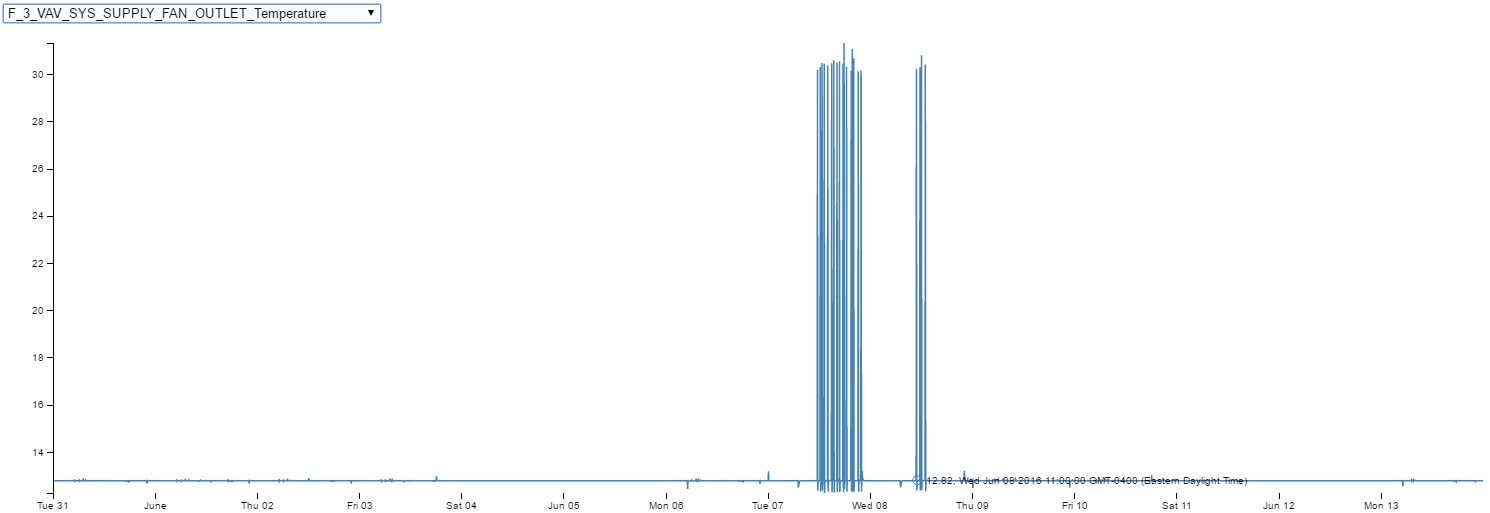


1. The Damper positions are set to 1 for all the zones during the following times
   1. June 7th 12:00 AM to 7:00 AM
   2. June 7th 9:00 PM to June 8th 7:00 AM
   3. Late hours of June 10th to June 13th 4:00 AM

As a result the airflow increases and thus the HVAC fan consumes more power during this time (As mentioned in Deliverable 1)



1. There is an anomaly observed in temperature of the air exiting the HVAC system fan on Floor 3. A sudden spike is observed after 11:00 PM on 7th June.



1. Some anomalies were also observed in the naming convention. The pattern usually followed is to have first letter of First name followed by Last Name. The ID’s for following users doesn’t follow this pattern and can be considered as special cases.

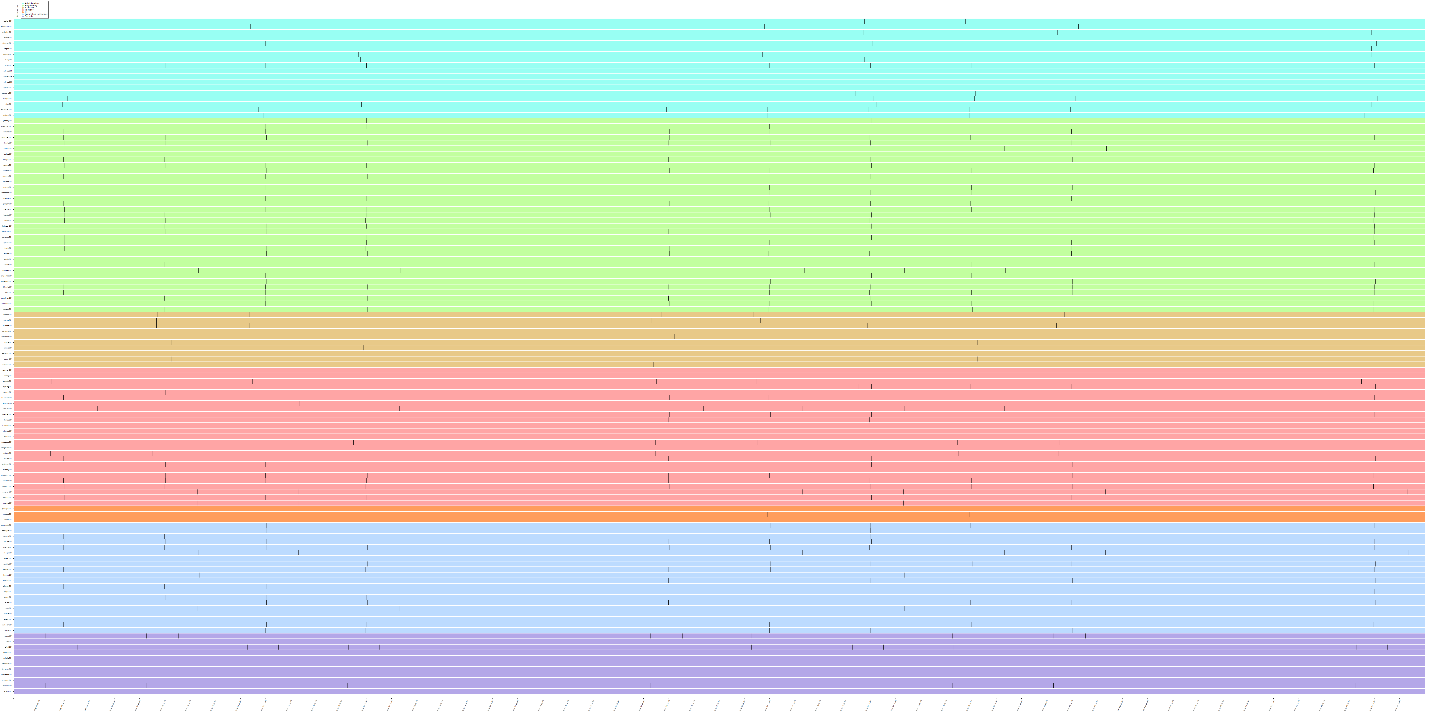
|  |  |  |  |
| --- | --- | --- | --- |
| **Last Name** | **First Name** | **Prox ID** | |
| Sanjorge Jr. | Sten | jsanjorge | |
| Morlun | Valeria | morlunv | |
| Alinari | Marcella | malinai | |
| Mies Haber | Ruscella | rmieshaber | |
| Paredes | Raye | rparade |
| Faraldo | Raphale | rfaralso |

1. There is no data available for Elevators, Lifts and stairs area.
2. The proximity card data for certain employees is recorded with different prox-ID’s. This can be logical if the employee has lost a card and have got a new card replacement which reads as new ID. But for a user named **Florez Geneviere** there were as many as 5 different Prox-ID’s in a span of 6 days, which poses a problem of uncertain data collection.

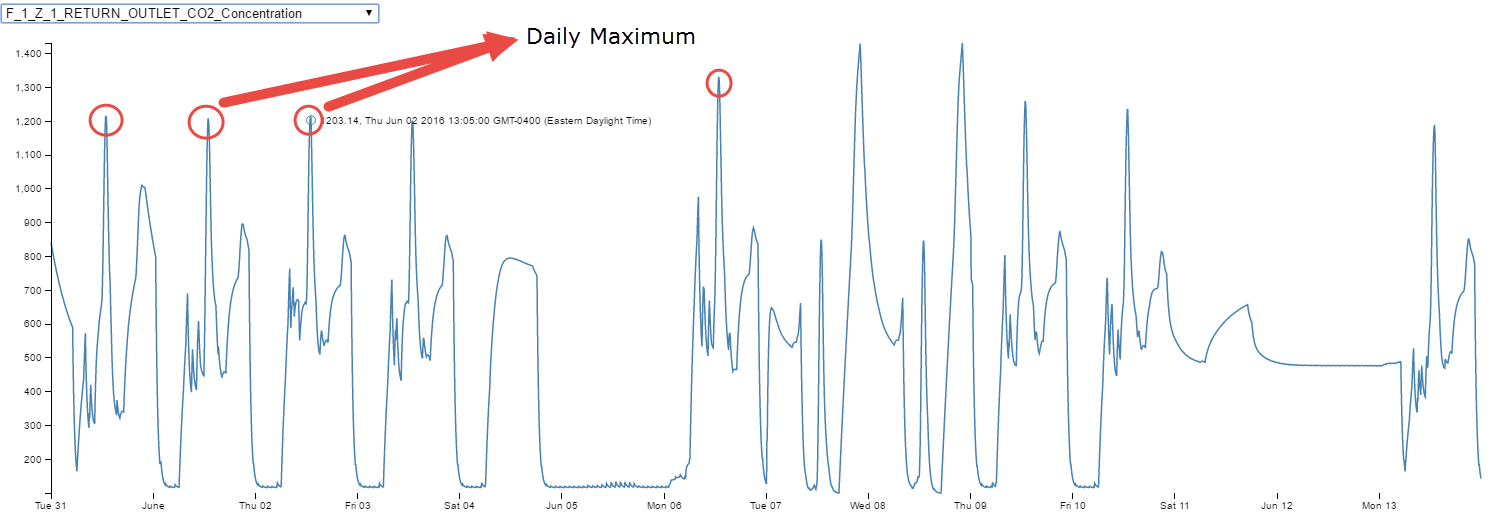
**2.b**–– Describe up to five observed relationships between the proximity card data and building data elements. If you find a causal relationship (for example, a building event or condition leading to personnel behavior changes or personnel activity leading to building operations changes), describe your discovered cause and effect, the evidence you found to support it, and your level of confidence in your assessment of the relationship.

I was able to dig out two relationships after correlating proximity card data and building data and they are listed as below.

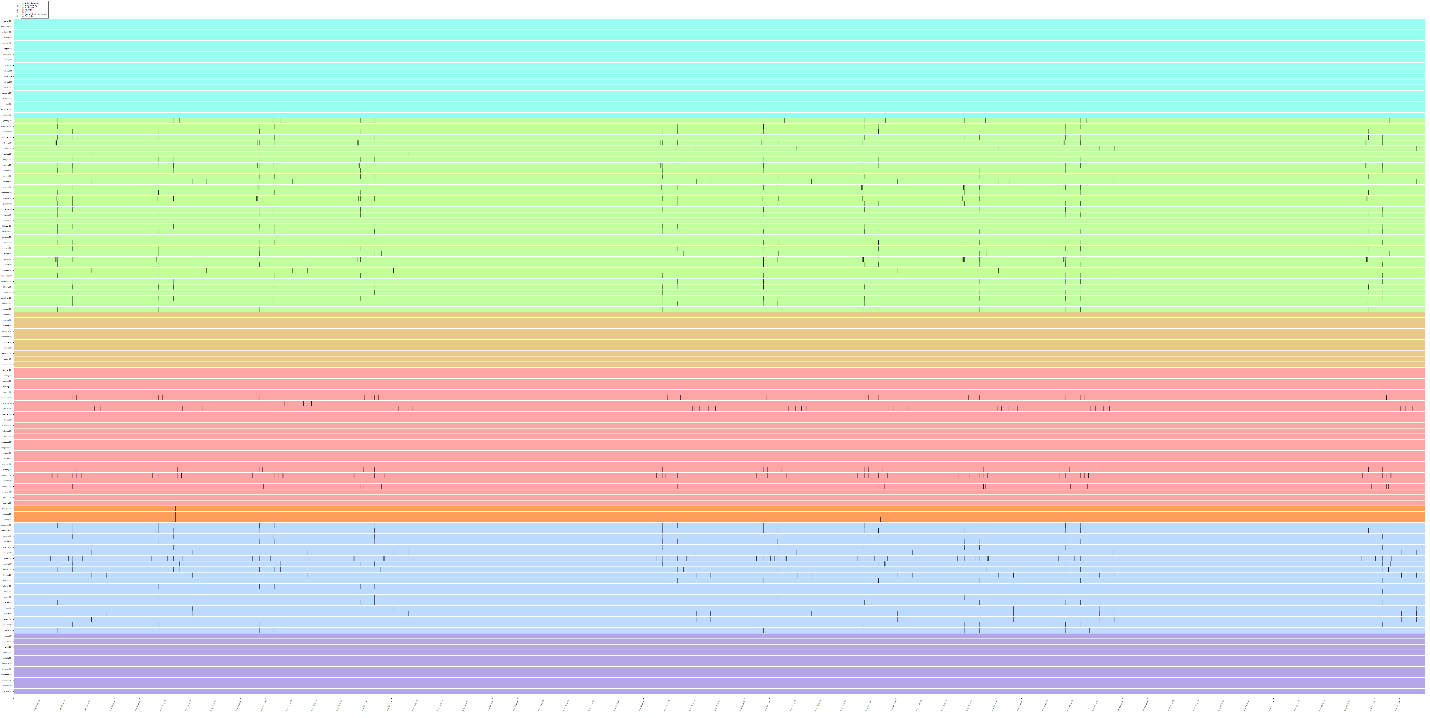
1. Relationship between Deli CO2 and Employee count at that area.
   1. Floor 1 Zone 1 in HVAC zones and Floor 1 Zone 2 in Proximity zone correspond to Deli on first floor.
   2. When a Gantt chart was rendered with just Deli as the only place and interesting pattern has turned up.
   3. According to this pattern Most of the employees from Engineering, Facilities and IT Department are located at Deli around 11:30 AM and 12:30 PM daily (the same can be observed in the Gantt chart)



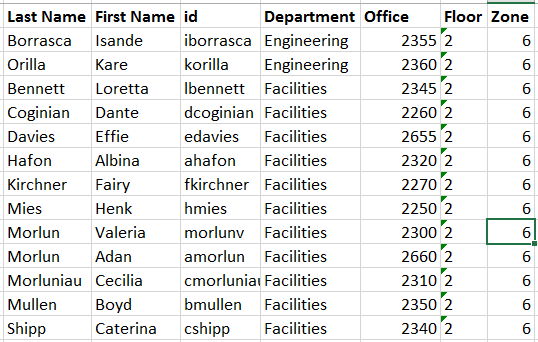
* 1. During this time around every day the Concentration of C02 measured at the Floor 1 zone 1's return air grille is at a daily maximum during this time and the same can be observed from the graph below.



1. Relationship between Floor 2 Zone 6 of Proximity data, Floor 2 Zone 14 of HVAC system with CO2 and Employee count at that area.
   1. A regular pattern was observed from Gantt chart where most of the Engineering and IT employees are located at a strikingly at the same proximity zone during two time intervals during the day i.e. around 10:30 AM and again around 2:00PM in the afternoon.



* 1. According to Proximity zone data the Zone 6 on Floor 2 predominantly has offices of employees from Facilities department and the same can be observed from the table below.



* 1. The one logical conclusion can be that all the employees from all the Departments which are picked up by proximity sensors at Floor 2 Zone 6 might be using the Meeting/Training room (2700).
  2. To support this fact the Return CO2 concentration of Floor 2 Zone 14 of the HVAC system which corresponds to the Meeting/Training room shows a rise after these timings daily. And the same can be seen in the graph below thus suggesting that there is a daily meeting attended by most of the Engineering and IT employees including employees from HR. This meeting occurs at the time intervals mentioned above i.e. around 10:30 AM and again around 2:00PM in the afternoon.

