ITCS 6162 Summer 2016
Final Project Deliverable 2

Tools Used:

1. Excel
2. Tableau

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

27 hours

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

Data mining techniques used are as below

1. Time series Analysis – This technique was used to visualize in difference in pattern along the time line on the given data. As for this section we had to compare between two different datasets hence applying association was difficult and time series turned out to be help for the establishing a relationship between the two datasets along the lines of time.
2. Clustering based approaches were used to cluster prox data to specific floor and zones. It was used to find a pattern which is frequent in particular floor and zone.

What are the limitations of your approach?

The limitations of the approach is time considered is in the approximation of hours and not the exact time.

Questions:

2.a) 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 or a serious issue for building operations.

Time series analysis using tableau software was done for the following deductions. Specific measures such as power,CO2 consumption etc. were plotted against continuous time to reflect a pattern or any anomaly to the pattern. All of the below results were assembled using Time series Analysis.

Notable anomalies in the data are below:

1) Power consumption is unusually high for the dates 06/07/2016 and 06/08/2016. Below graph represents the maximum electric demand for a particular day. For prior weekend it is observed to have demand of 100k-150k W which is seen to be elevated in the subsequent weekends.

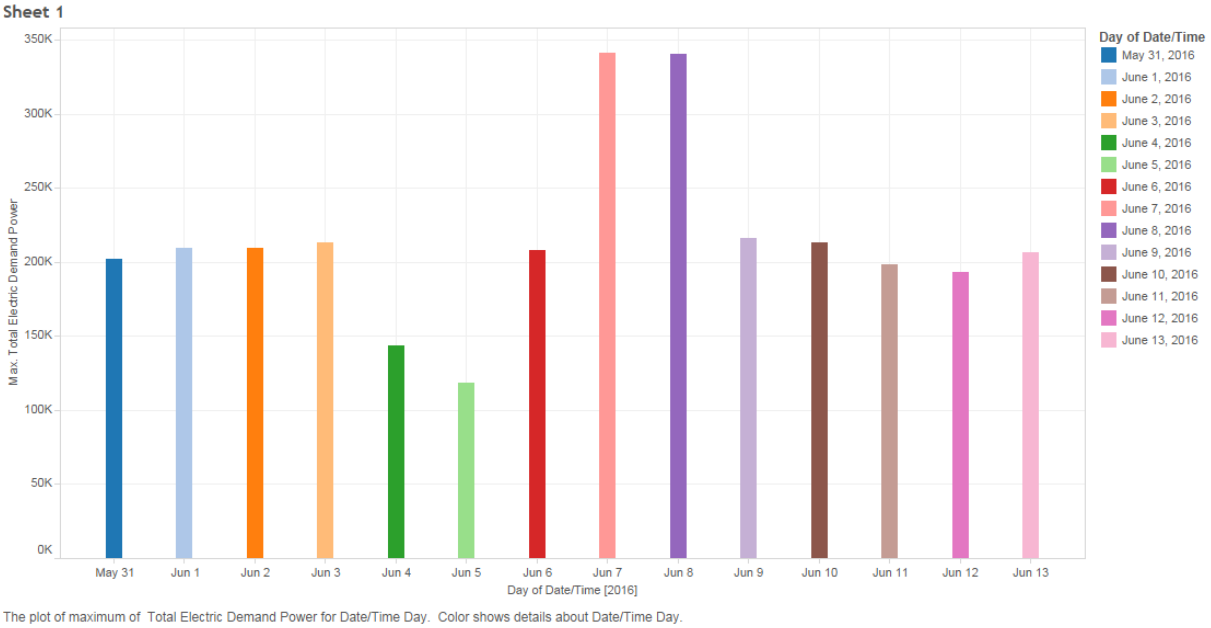


Figure-1

2) CO2 concentration levels have increased for the dates 06/07/2016 and 06/08/2016. Below graph represents the CO2 concentration for Floor 1 zone1 & zone 4 and Floor 2 zone1 & zone 2. These floors were taken in to consideration because of high prox card reads.

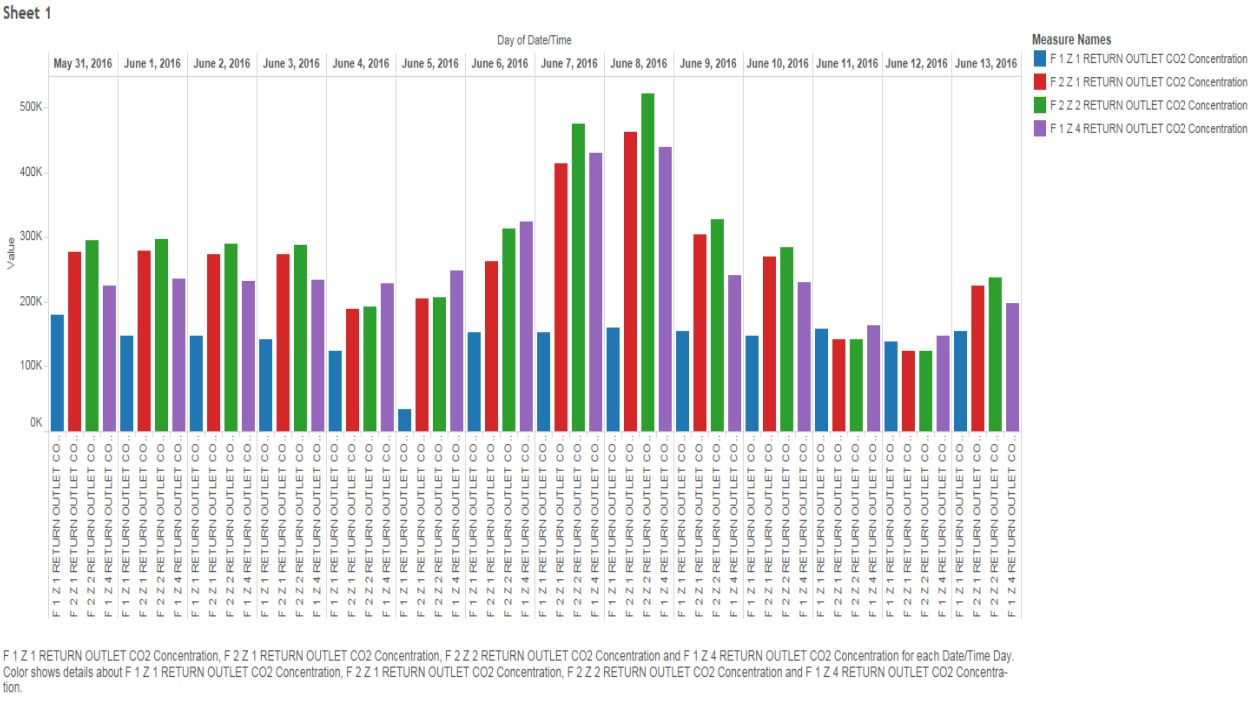


Figure-2

3) Power used by the zone air supply box coil i.e. the Reheat coil power is Significantly elevated for dates 10th, 11th 12th & 13th June which is usually observed to be 0 or less than 500W for other days. Below chart displays the values for Floor 1 zone 1 and Floor 2 zone1&zone2.

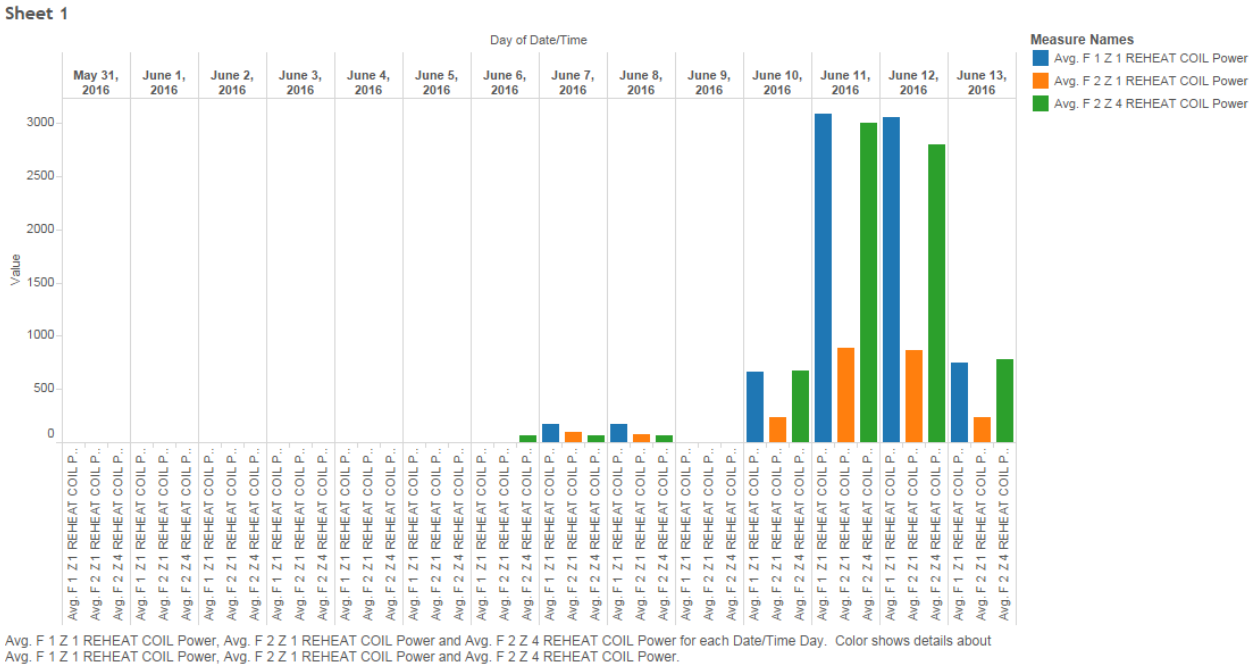


Figure-3

Floor 3 reheat coil power shows unusual pattern where it is elevated from 0 to about 8000W

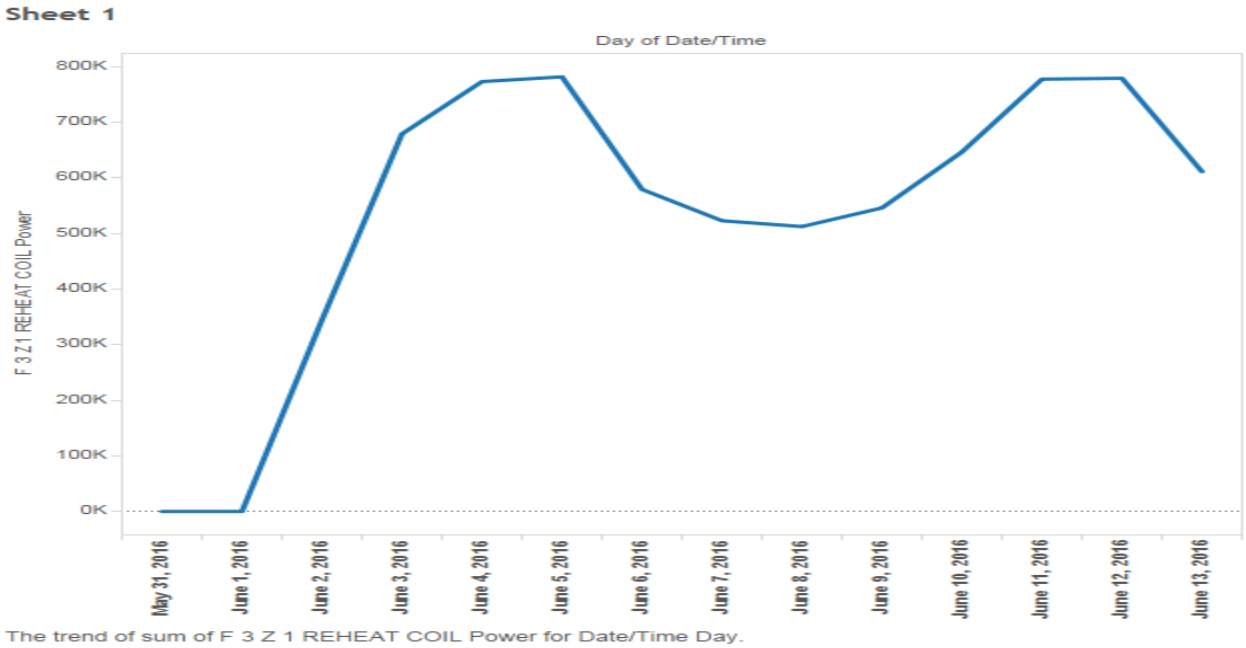


Figure-4

4) The HVAC electric demand is unusually elevated for 7th and 8th of June and remains more than usual for the following days.

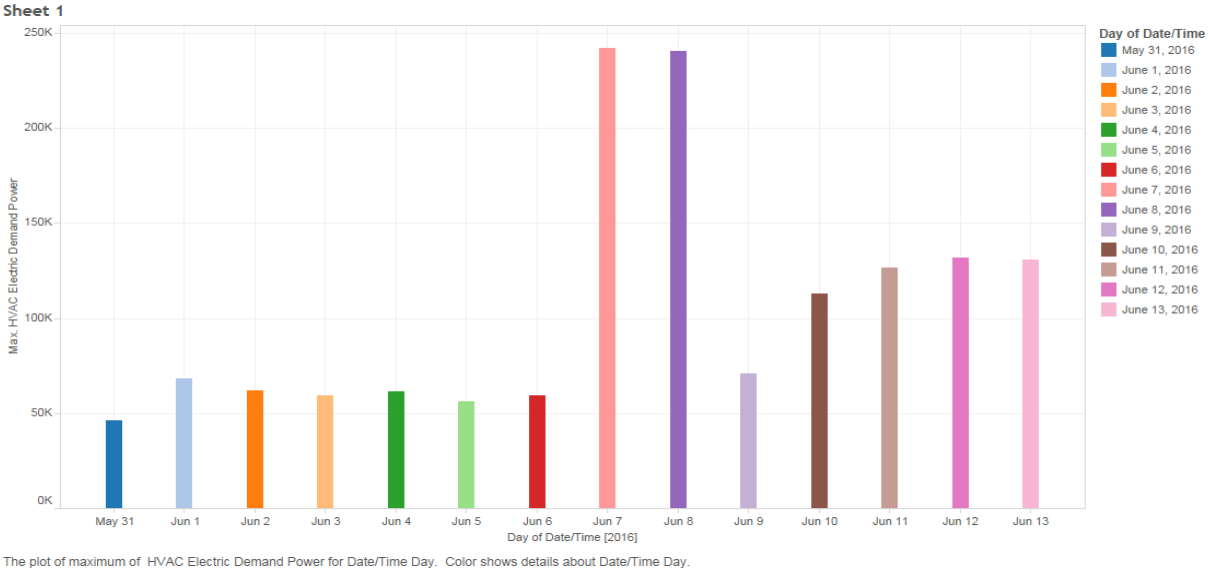


Figure-5

5) Elevated thermostat temperature for 6th, 7th and 8th of June which later is dropped to usual pattern.

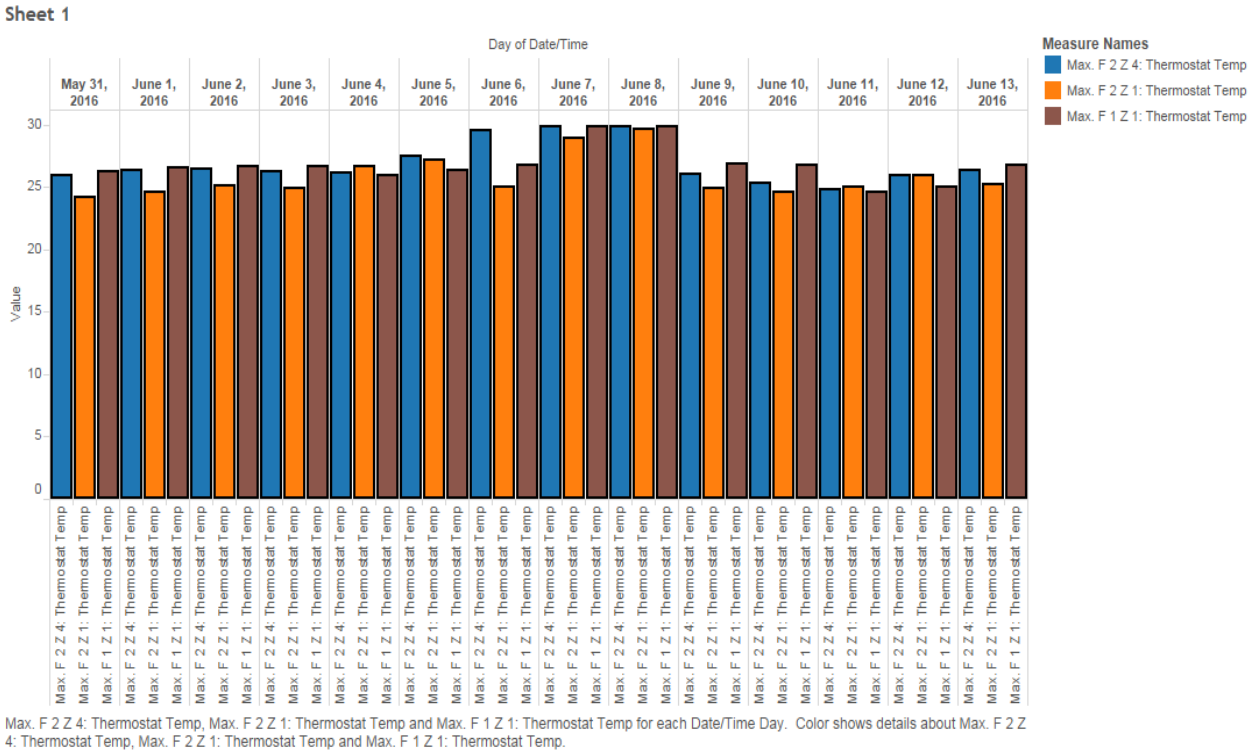
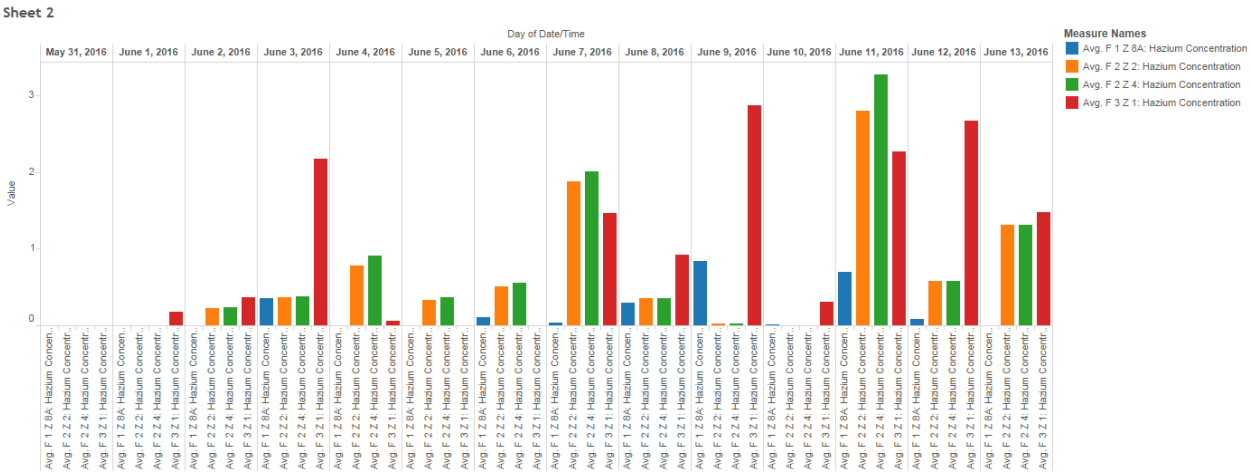


Figure-6

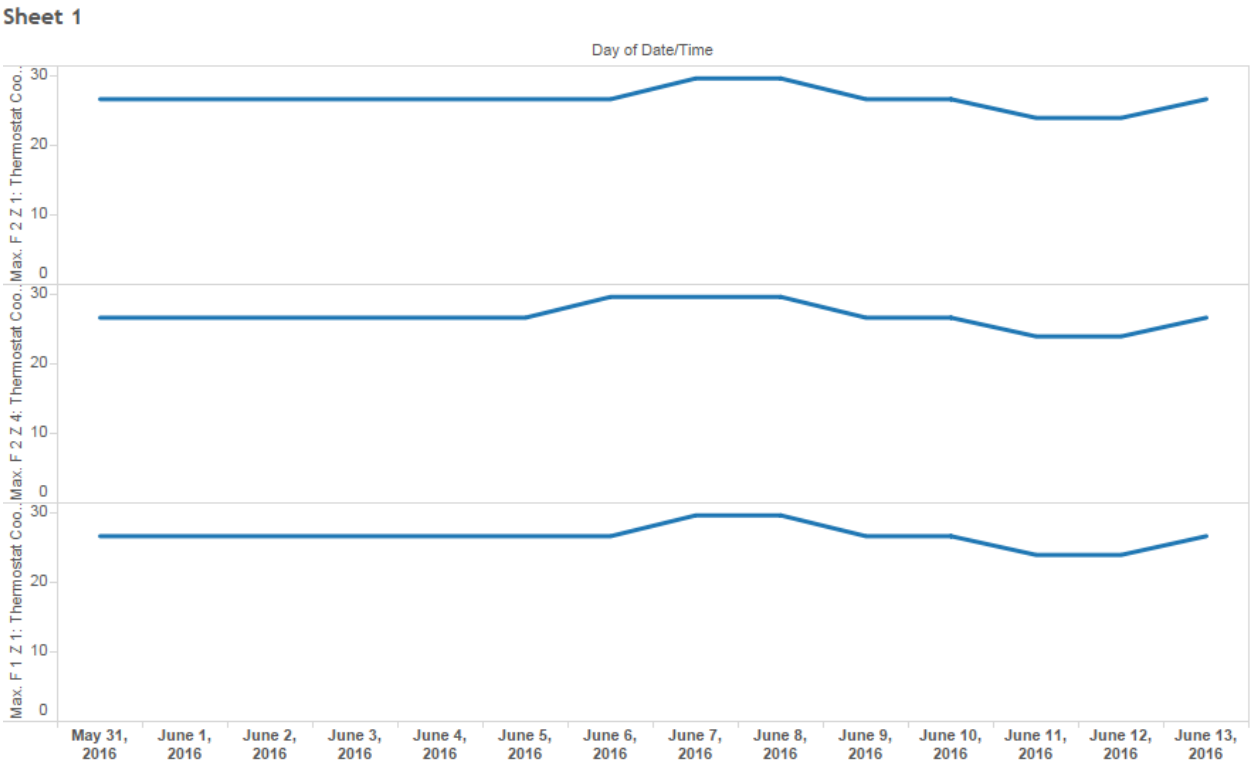
6) Average Hazium concentrations are elevated for 11th & 12th June and remain elevated from normal range even after that. Floor 3 Average Hazium concentration seems to be elevated most of the times.



Avg. F 1 Z 8A: Hazium Concentration, Avg. F 2 Z 2: Hazium Concentration, Avg. F 2 Z 4: Hazium Concentration and Avg. F 3 Z 1: Hazium Concentration for each Date/Time Day. Color shows details about Avg. F 1 Z 8A: Hazium Concentration, Avg. F 2 Z 2: Hazium Concentration, Avg. F 2 Z 4: Hazium Concentration and Avg. F 3 Z 1: Hazium Concentration.

Figure-7

7) Variations in the maximum thermostat cooling and heating set points from 6th of June and remain inconsistent then onwards.



The trends of maximum of F 2 Z 1: Thermostat Cooling Setpoint, maximum of F 2 Z 4: Thermostat Cooling Setpoint and maximum of F 1 Z 1: Thermostat Cooling Setpoint for Date/Time Day.

Top of the priority should be given to increased levels of Hazium and CO2 concentrations in the building are significantly hazardous. Concentrations typical of occupied indoor spaces with good air exchange should be between 350-1000 ppm where as it has been observed that for most of the floors the maximum CO2 concentration is more than 1000ppm.

Rest of the elements looks less prior to be focused as hazardous for the building.

2.b) Various observed relationships between the proxim ity card data and building data elements are as follows

Time series analysis using tableau software was done for the following deductions. Specific measures such as count of prox reads, power, CO2 consumption etc. were plotted against discrete time to reflect any relationship between the prox dataset and building dataset to the pattern. All the below relations were established using time series analysis and clustering the prox data in zones and floors.

1) Lights power for each zone varies according to the prox reads in a particular zone, i.e. the lights are functional only if there are employees in that zone else it would be turned off. For evidence below graph of floor 1 zone 1 is taken into consideration

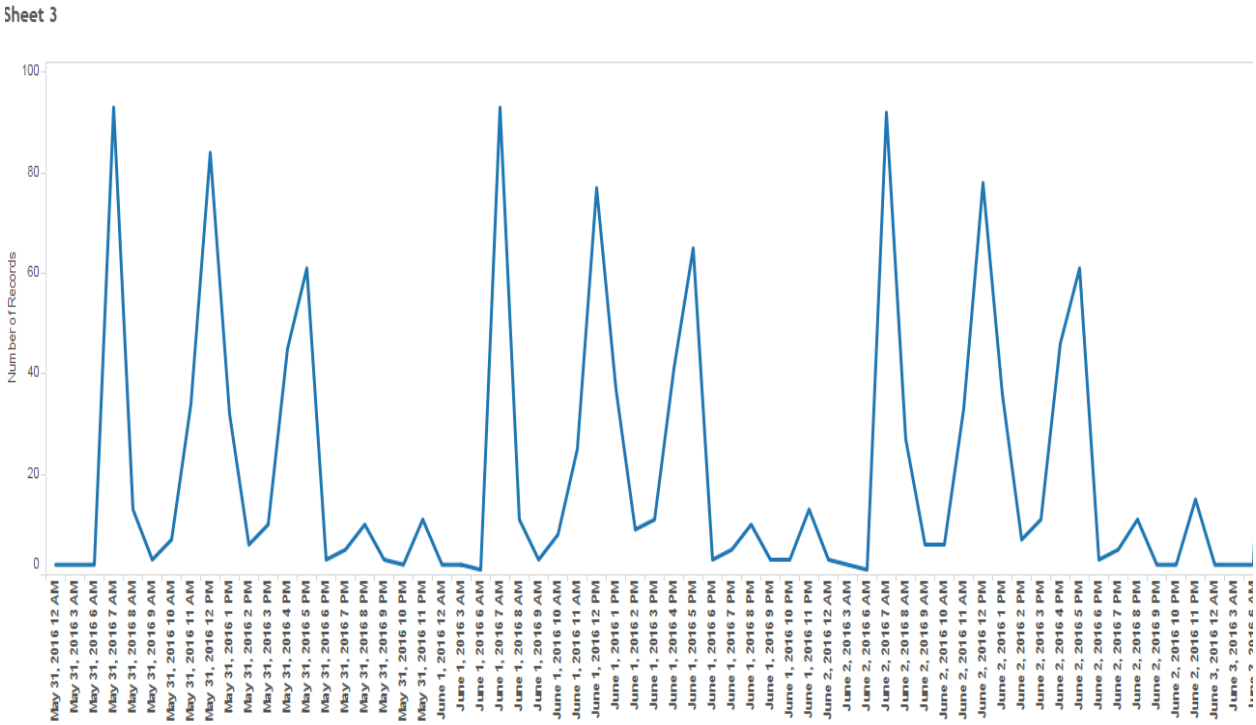


Figure- 9

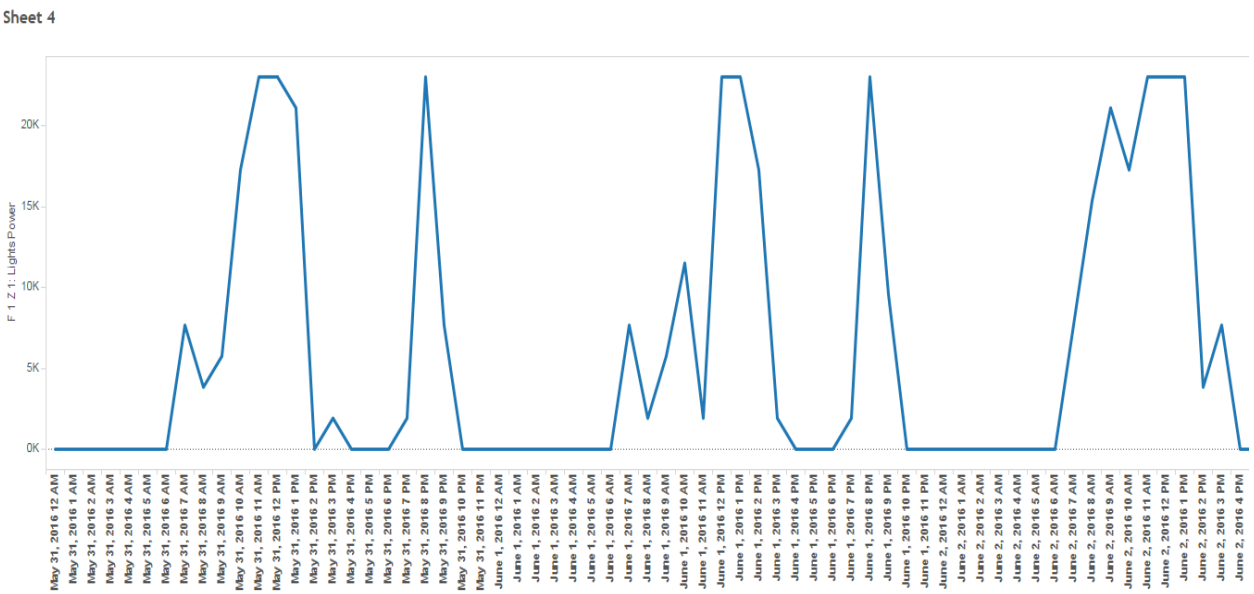


Figure-10

Before 6 AM there is no change in light powers in figure 10 and so is the prox read counts in figure 9. Variation is noticed after 6AM in both the patterns, it comes to a halt in between 10-11 PM. And the pattern continues to repeat for all the floors. From 4th June (June 5 & June 5 are weekends) to 6th June morning there is no activity in lights power, so is for prox reads.

For the above assessment of relationship level of confidence is 100%.

2) Total electric demand for the building follows the general working shift of the employees, i.e. when there is maximum number of employees working in the building. It is observed between 6AM to 7PM there are maximum number of prox reads so does the total electric demand elevates and drops after 6PM and gradually after 7PM. Weekends there isn't much variations seen. Hence the total electric demand is at its maximum during the general shift of operation. Below is the evidence supporting this assessment, Figure 11 represents the total prox read in the building for distributed hourly and figure 12 represents total electric demand for the building distributed hourly.

Level of Confidence for this assessment of relationship is 100%

Sheet 6

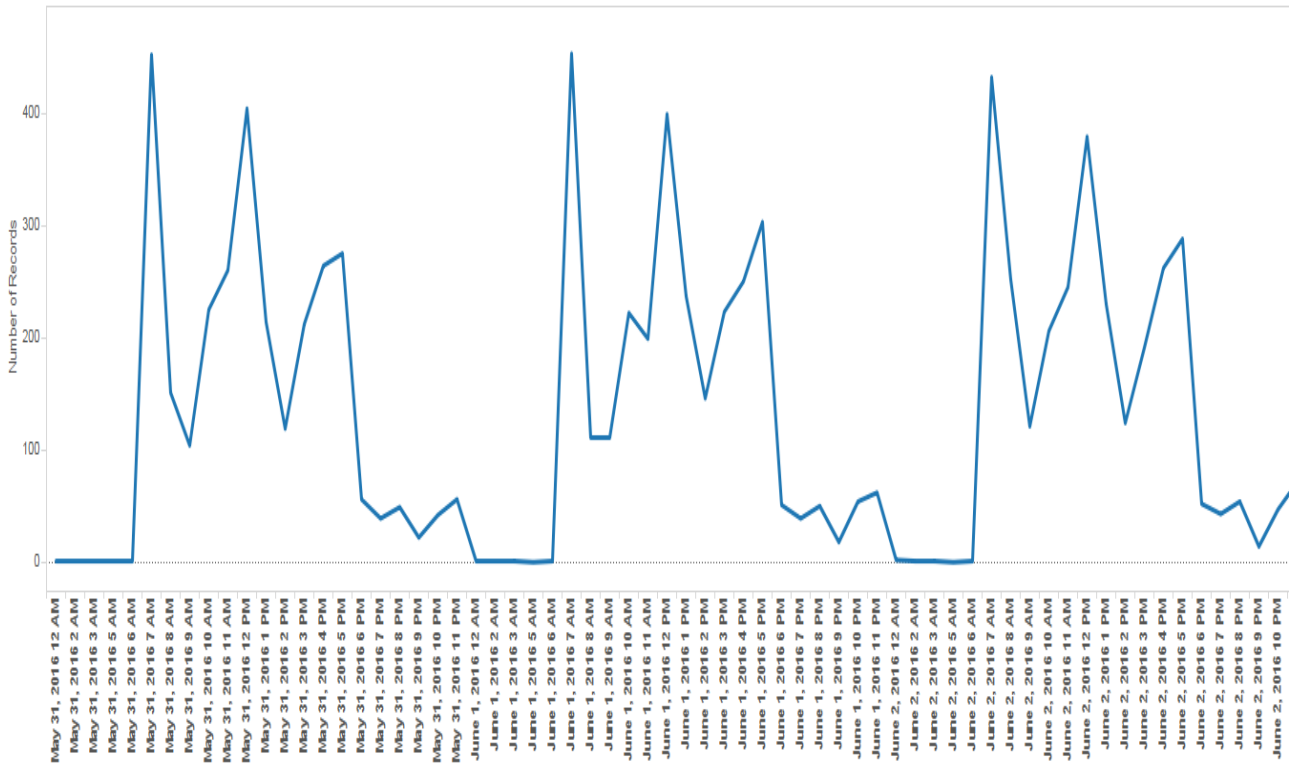


Figure-11

Sheet 7

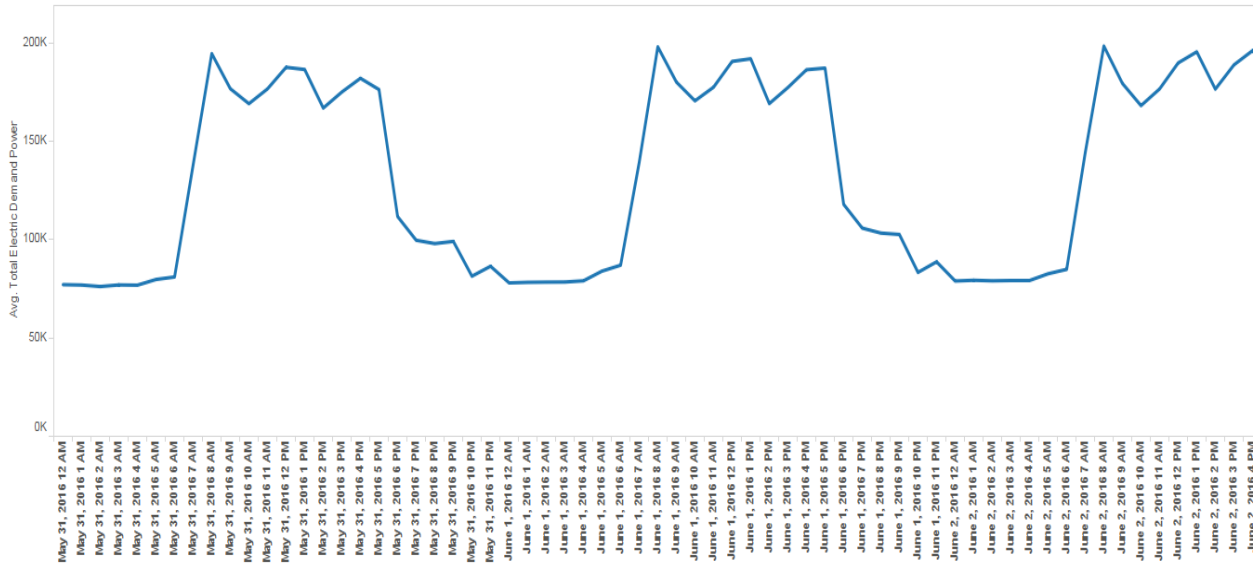


Figure-12

3) The CO2 concentration is directly proportional to the number of prox reads per floor, i.e. higher the number of prox reads in a floor higher the CO2 concentration for that particular floor. Figure 13 shows the number of prox reads per floor per day and figure 14 shows CO2 concentration for F1Z1, F2Z1,F2Z1(These have maximum prox reads with in the floor)

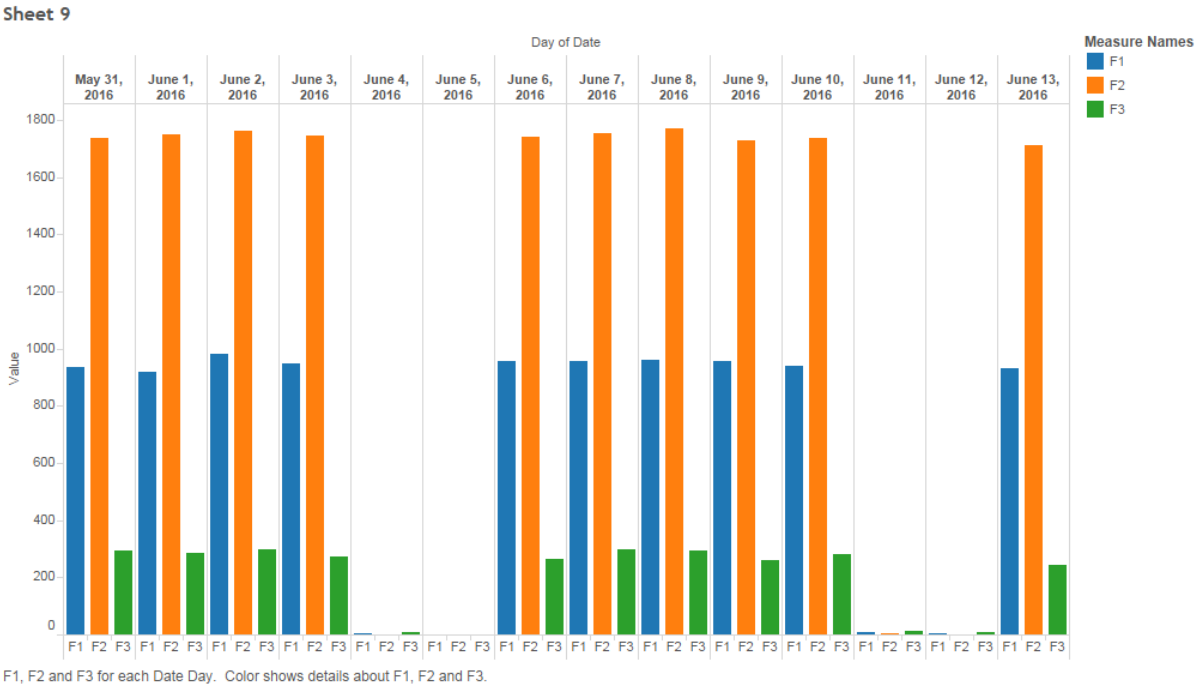


Figure-13

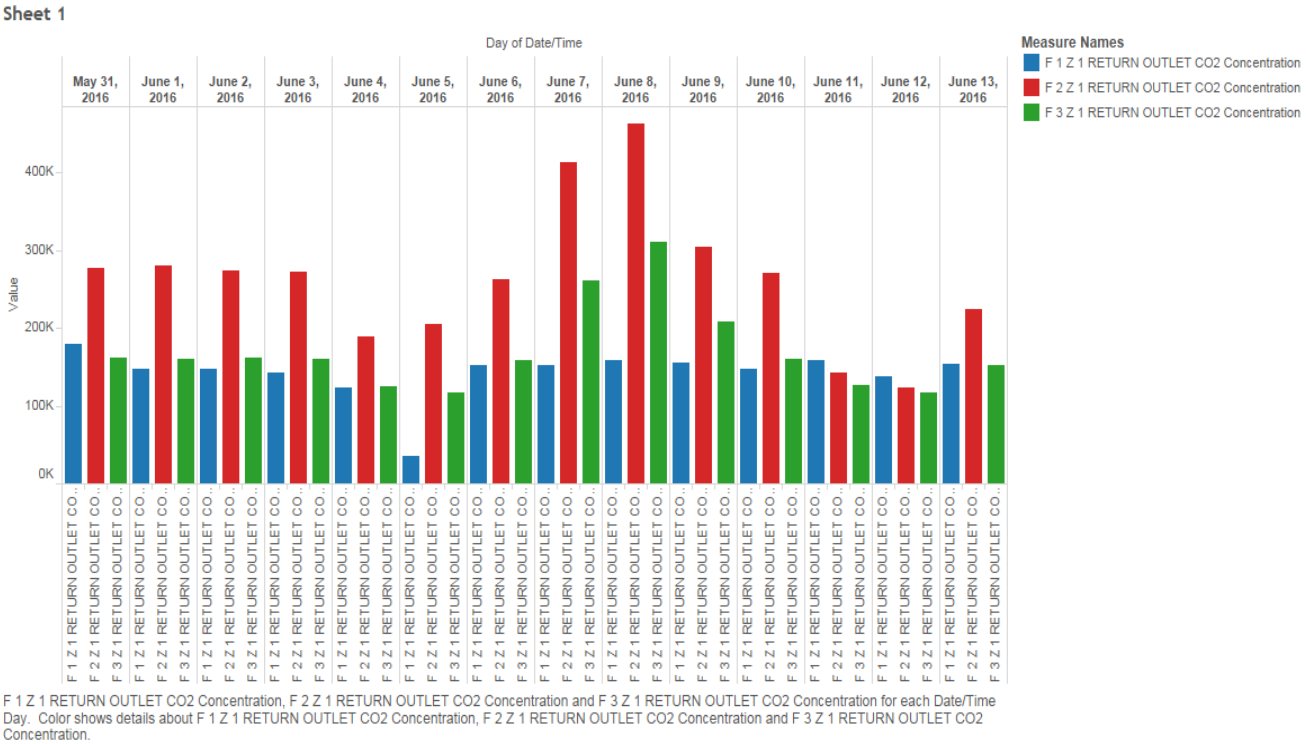


Figure-14

It clearly indicates that Floor 2 has maximum prox reads and maximum CO2 concentration as well, followed by floor 1 which is the 2nd highest in both prox reads and CO2 concentration.

Confidence for this assessment is 80% as for the weekends the data is not coherent.