

ITCS 6162 Summer 2016
Final Project Deliverable 1

Tools Used:

1. Excel
2. R editor

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. Association Rule – Reason for choosing this technique was to determine patterns that occurred frequently. As the question demanded discovering the pattern hence this technique was used. Apriori algorithm helped find the associations and patterns
2. Time series Analysis – This technique was used to visualize in difference in pattern along the time line on the given data.

What are the limitations of your approach?

The limitations of the approach is if a large number of attribute is considered as a factor for the Apriori algorithm to formulate the association rules, the number of rules or patterns generated becomes significantly high.

Questions

1.a – What are the typical patterns visible in the prox card data? What does a typical day look like for GASTech employees?

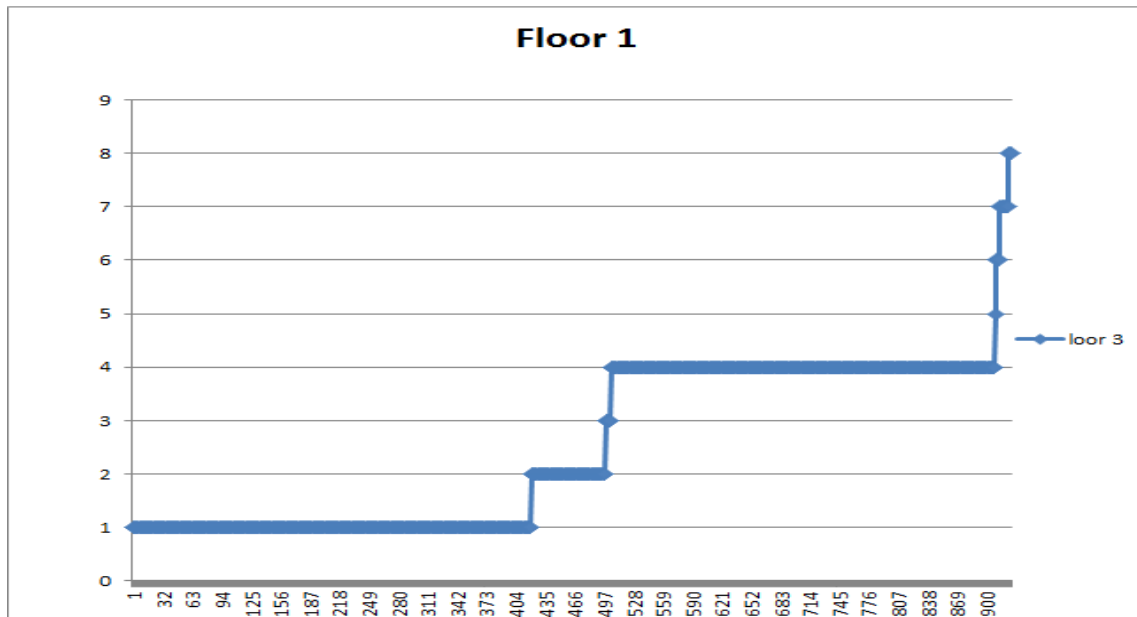
The typical patterns visible in the prox card data are as below:

1. {floor=1=>zone= 4} has 100% confidence which suggests this is one of the usual pattern in the prox card data. It has been found to occurred frequently and has the maximum probability that it is one of most frequently visited zones in the Gastech
2. {floor=2=>zone= 4} has 100% confidence which suggests this is one of the usual pattern in the prox card data. It has been found to occurred frequently and has the maximum probability that it is one of most frequently visited zones in the Gastech
3. {floor=1=>zone= 1} has 100% confidence which suggests this is one of the usual pattern in the prox card data. It has been found to occurred frequently and has the maximum probability that it is one of most frequently visited zones in the Gastech

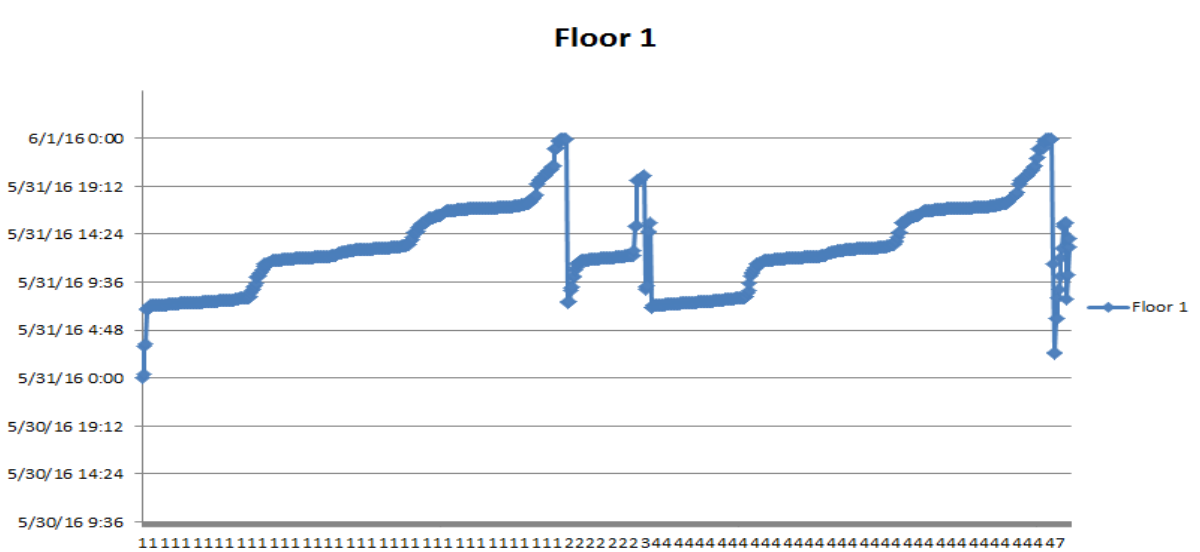
4. {floor=2=>zone= 1} has 100% confidence which suggests this is one of the usual pattern in the prox card data. It has been found to occurred frequently and has the maximum probability that it is one of most frequently visited zones in the Gastech

5. {zone= 7} => {floor=2} has 97.2% confidence which suggests this is one of the usual pattern in the prox card data. It has been found to occurred frequently and has the maximum probability that it is one of most frequently visited zones in the Gastech.

Zones Vs Frequency

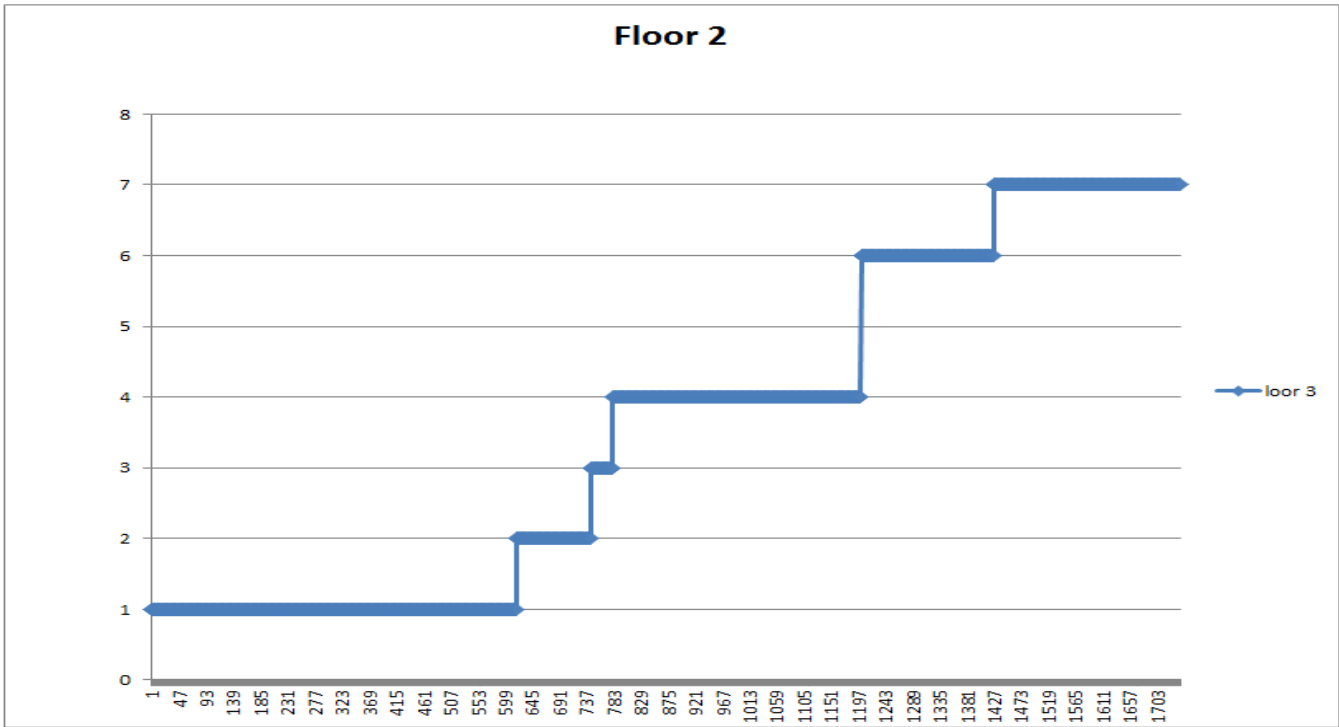


Timestamp Vs Zones

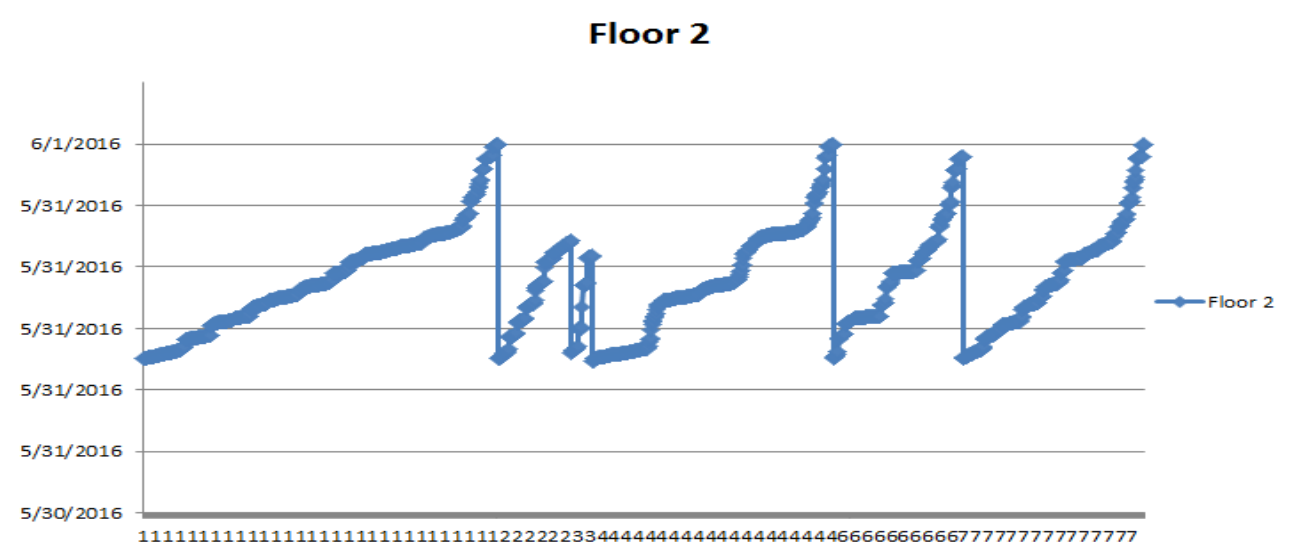


Above graphs represents in a given day the frequency of employees in zones of floor 1. Above clearly indicates that in a given day frequency of employees is much higher in Zone 1 and Zone 4 of floor 1 which ranges in between 1-400 card read a day.

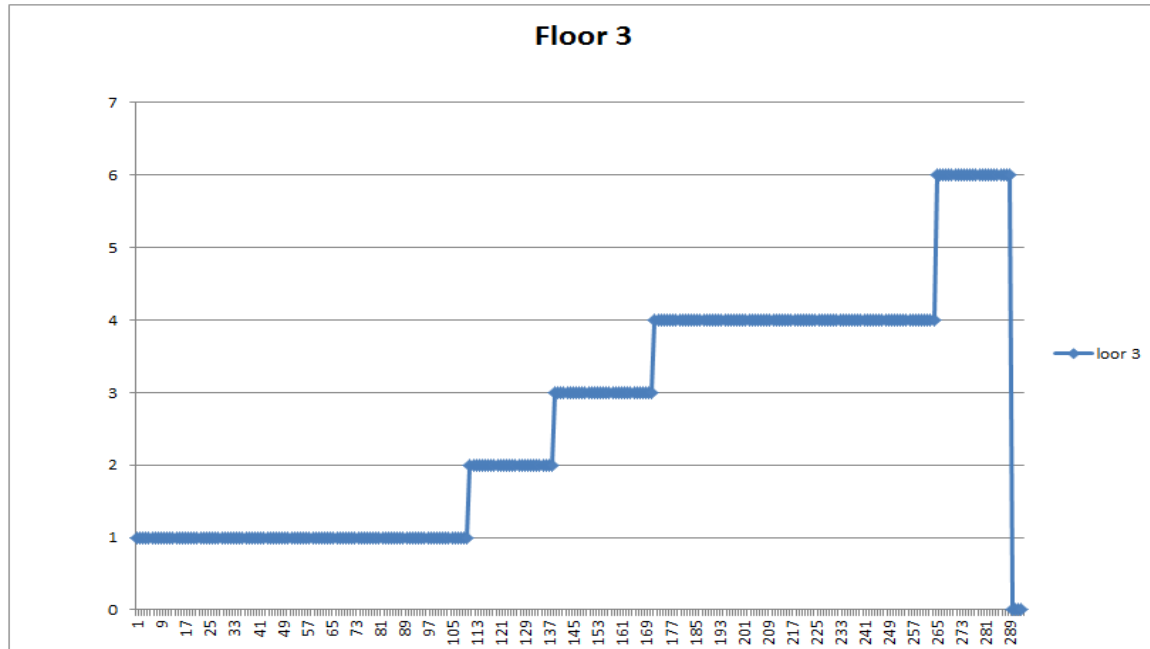
Zones Vs Frequency

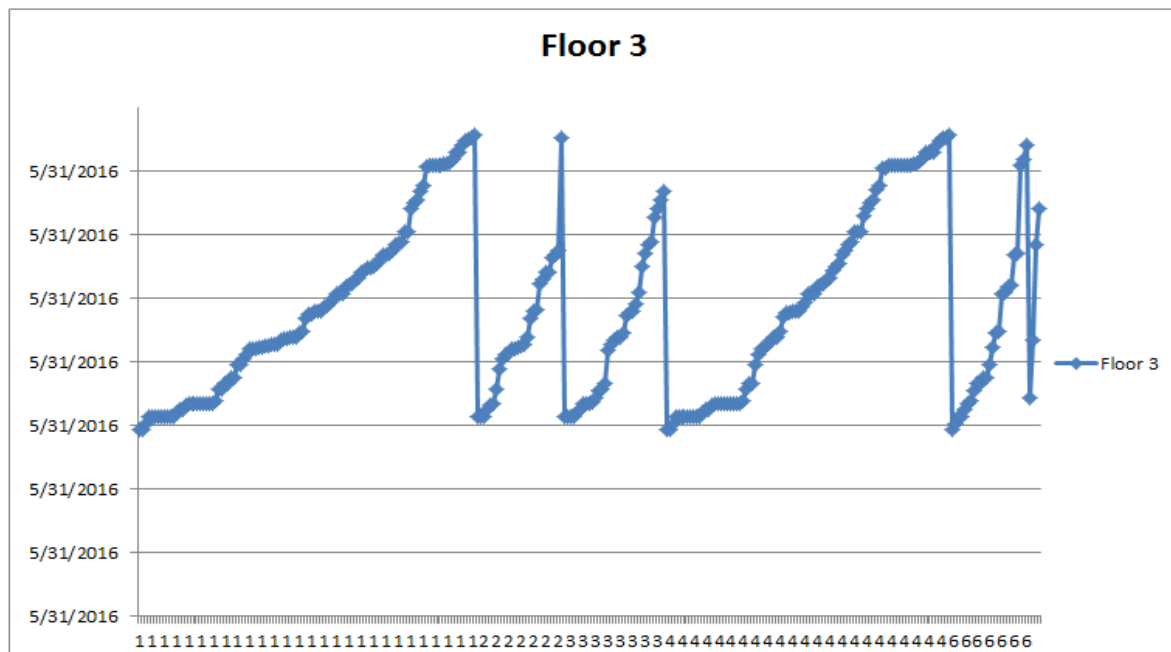


Timestamp Vs Zones



Zones Vs Frequency





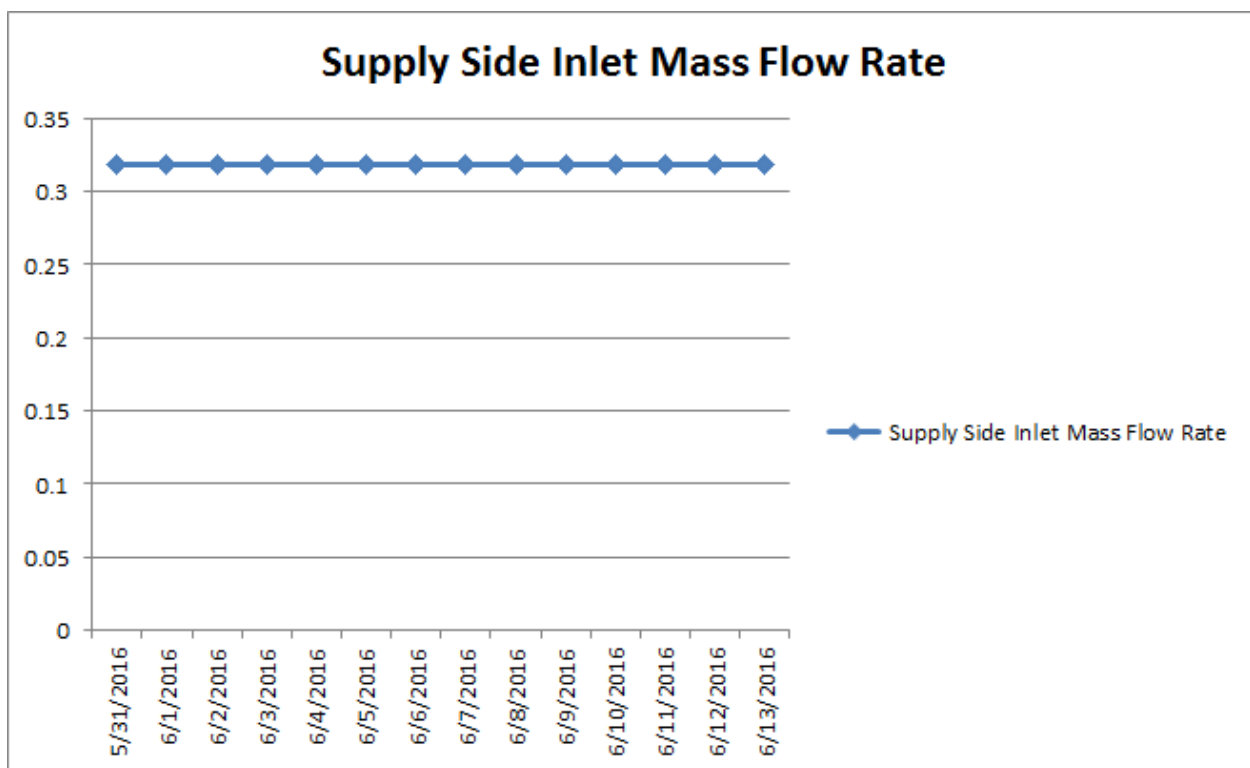
Above graphs represents in a given day the frequency of employees in zones of floor 2. Above clearly indicates that in a given day frequency of employees is much higher in Zone 1 and Zone 4 of floor 3 which ranges in between 1-100 approx card read a day.

In a usual day the employees start the day reaching offices and using the elevators or stairs located in the zone four for each floor to the office space followed by meeting in conference room and break rooms.

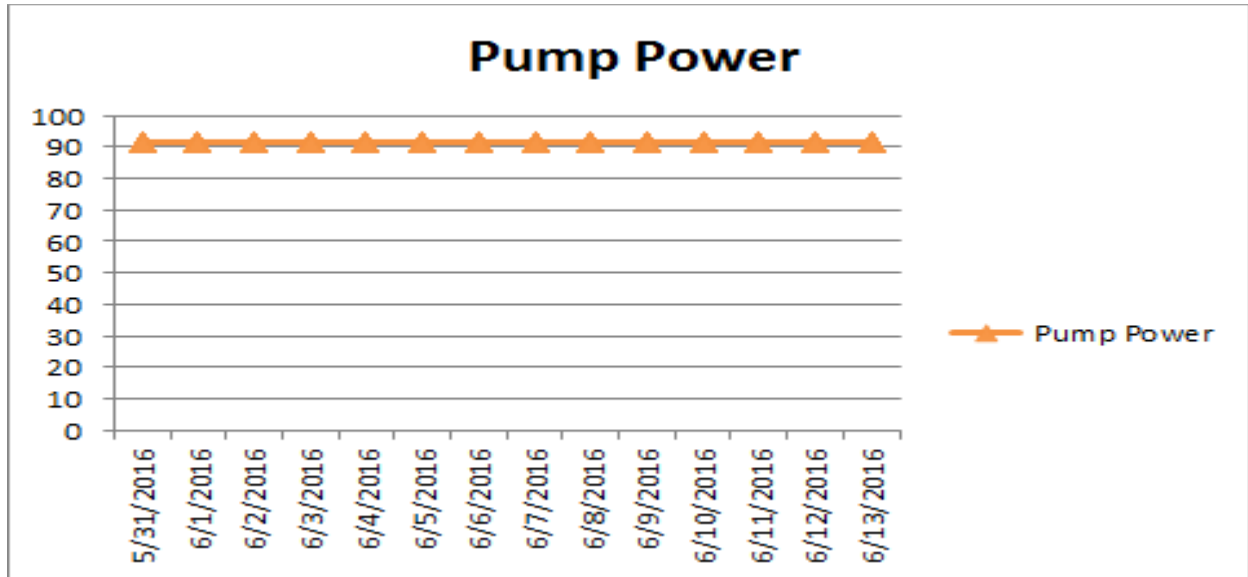
1.b – Describe up to ten of the most interesting patterns you observe in the building data. Describe what is notable about the pattern and explain what you can about the significance of the pattern.

Ten most interesting patterns in the building data are as follows:-

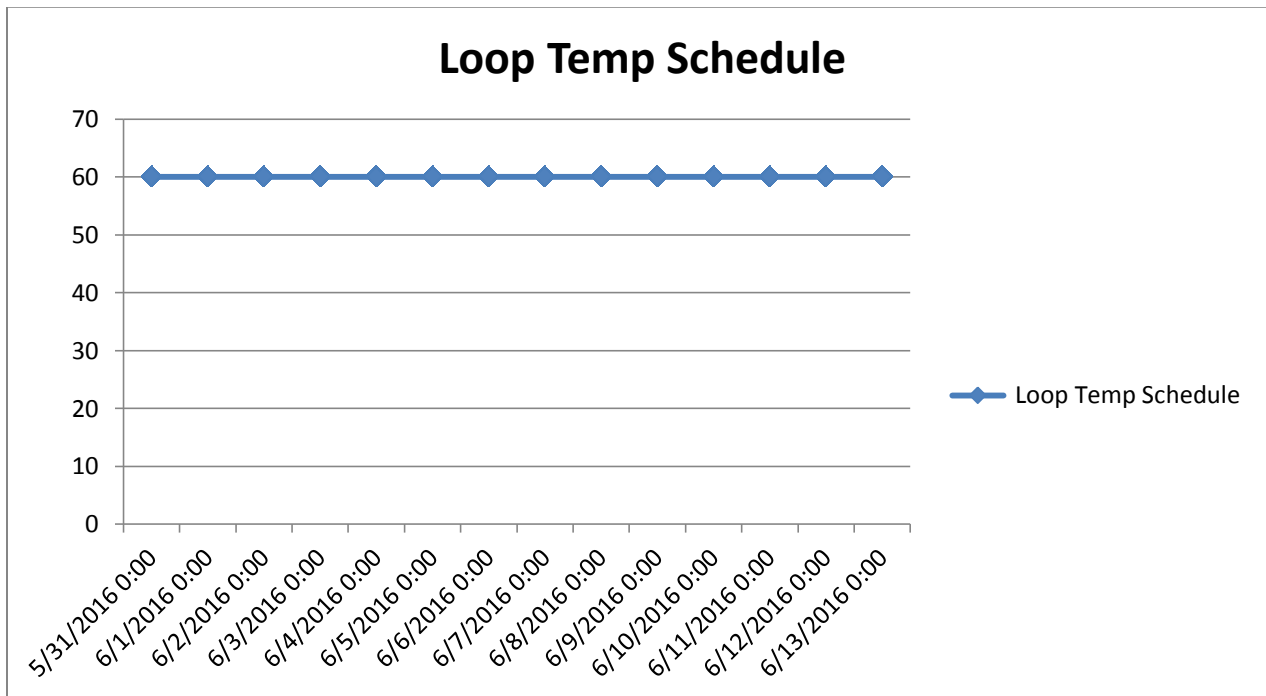
1. Water Heater Gas Rate is 0 most of the time with 87.47 % confidence
2. Supply Side Inlet Mass Flow Rate remains constant at 0.3179 throughout. Temperature of water entering the hot water heater has to be maintained at a specific temperature to get the desired out put



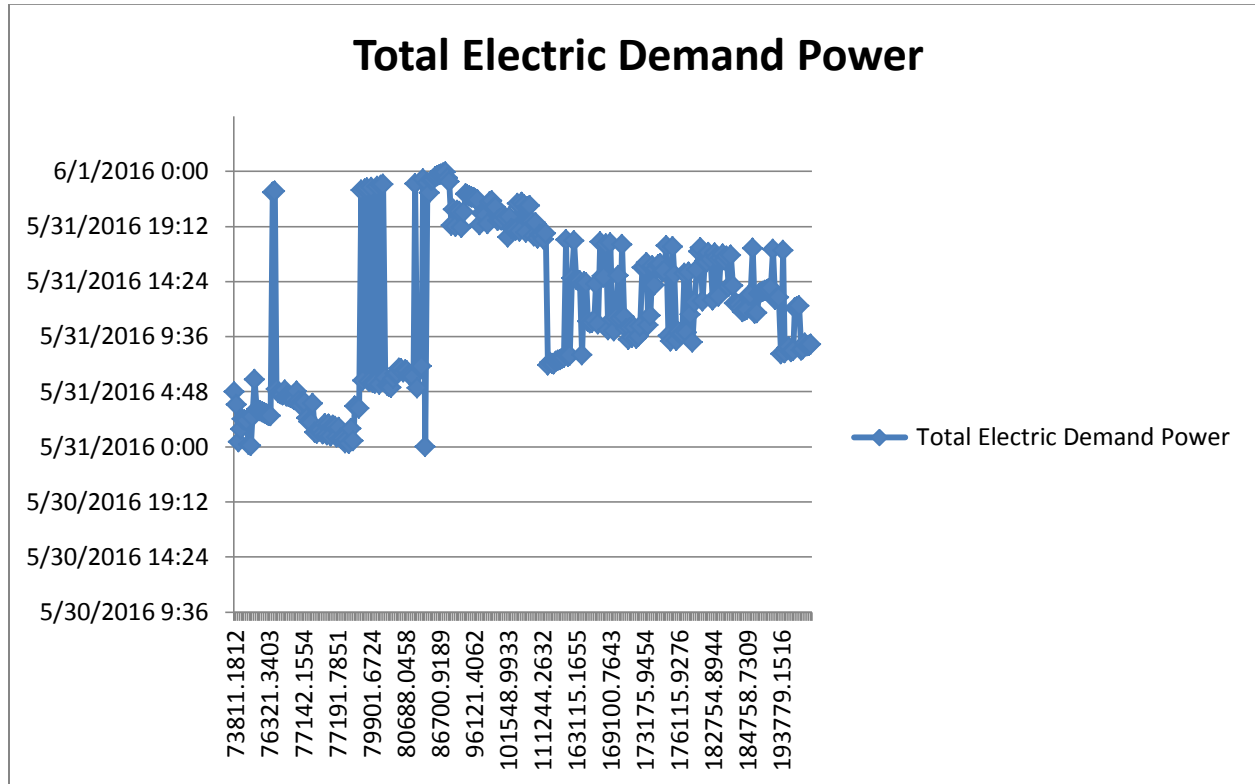
3. Pump Power is constant at 91.3744W every day. As it's the power used by the hot water system pump it has to remain constant.



4. Temperature set point of the hot water loop is 60. This is the temperature at which hot water is delivered to hot water appliances and fixtures for it proper function.

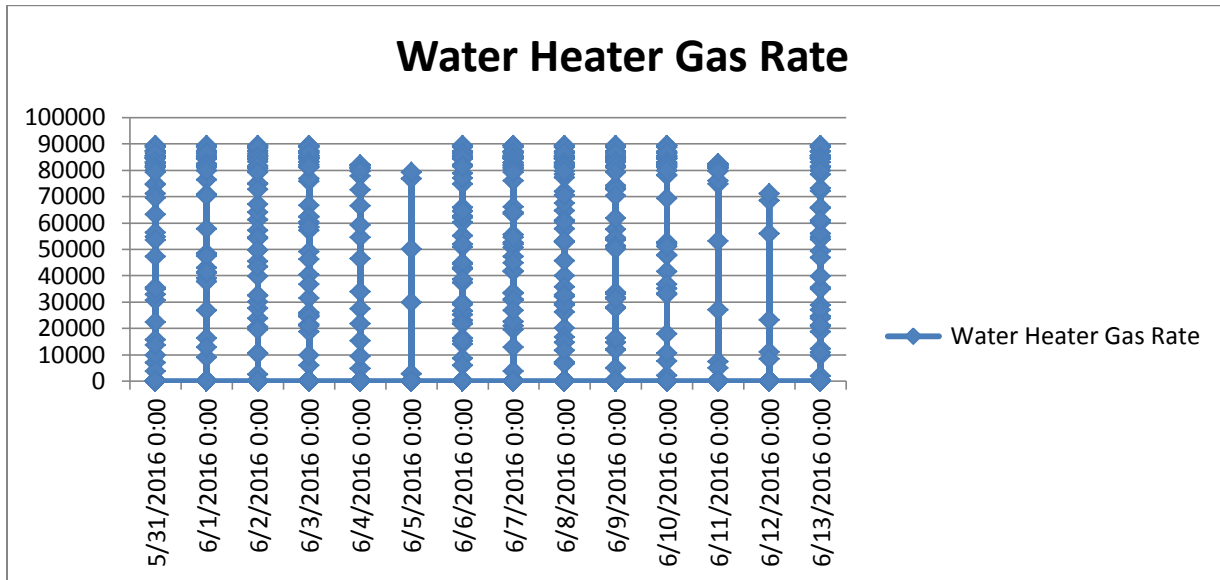


5. Highest total electric demand is between 8:00 am to 19:00 pm daily. As it's the peak working hour hence the demand is more in this time.

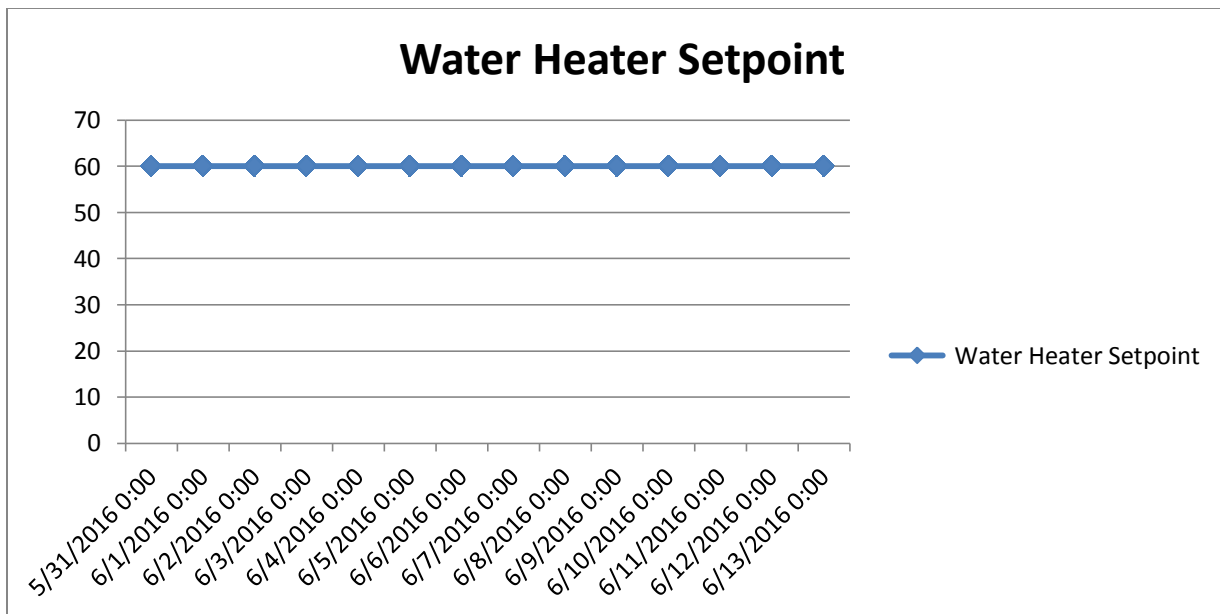


6. The power used by the deli exhaust fan remains 47.1667 W except for morning 6:00 am to 8:00 and 17:05-22:00 on weekdays i.e. Monday to Friday. However on Saturdays the power consumed is 0 from 6:00 am to 6:00 pm and on Sundays throughout the day it uses 47.1667 W. This unusual pattern could be due to weekend work off for employees.

7. Water heater gas rate follows a pattern throughout the weekdays and is minimum on Sunday. This pattern pertains to the working hours.



8. The thermostat heating and cooling set points for all the zones in the floors are constant. Heating set point is at 15.6 C and cooling set point is at 26.7C. this is set to maintain the temperature of air inside the zones, so that the temperature doesnot fall below 15.6C and above 26.7 C.
9. Water heater set point is constant at 60 C. This temperature is maintained as the set point for the water heater.



10. Dry bulb temperature remains highest during 10:00 AM to 3:00 PM. It refers to the temperature of outside air hence it remains highest during this time slot when the temperature is at its peak.

