****

Pooja Gupta (A0163281J)

Seshan Sridharan (A0148476R)

SMART BUILDING ANALYTICS

NUS School of Design and environment

Contents

[**EXECUTIVE SUMMARY** 2](#_Toc512376002)

[**INTRODUCTION** 3](#_Toc512376003)

[**BUSINESS OBJECTIVE** 4](#_Toc512376004)

[**DATA ANALYTICS OBJECTIVE** 4](#_Toc512376005)

[**PLANNING** 5](#_Toc512376006)

[**TECHNICAL STRATEGY** 5](#_Toc512376007)

[**DATA ACQUISITION AND PROCESSING** 6](#_Toc512376008)

[OUTLIER ANALYSIS 6](#_Toc512376009)

[MISSING DATA 6](#_Toc512376010)

[DATA AGGREGATION 7](#_Toc512376011)

[PROBLEMS AND CHALLENGES 7](#_Toc512376012)

[COLLECTION OF NEW DATA 7](#_Toc512376013)

[DATA SECURITY 7](#_Toc512376014)

[**ANALYSIS AND MODELLING** 7](#_Toc512376015)

[EXPLORATORY ANALYSIS 9](#_Toc512376016)

[**VALIDATION OF RESULTS** 13](#_Toc512376017)

[**DASHBOARD DESIGN** 14](#_Toc512376018)

[**CONCLUSION** 16](#_Toc512376019)

[**RECOMMENDATION FOR FUTURE WORK** 16](#_Toc512376020)

[**REFERENCES** 17](#_Toc512376021)

[**APPENDIX A** 19](#_Toc512376022)

# **EXECUTIVE SUMMARY**

Building performance plays a major role in the expectations expressed by owners and occupants and their fulfilment by designers and building operators. The improved match between the expectation and the actual delivery is considered an important target for building industry to provide better value. The need to move the industry in this direction has fostered the introduction of performance-based building methods.

NUS School of Design and Environment is towards conceptualising green building designs, such as harnessing solar energy, hybrid cooling approach, natural ventilation and lighting. The successful delivery of green buildings requires balancing energy and resource efficiency while providing a comfortable, healthy and productive environment within economic means. Occupant comfort and behaviour can have a significant impact on green building performance. In this light, it is desired to design Environmental control systems to intelligently respond and adapt to changing conditions and needs with minimal user engagement.

Integrated IoT sensors were designed and deployed at different locations to measure the environmental parameters like temperature, humidity, VOC (Volatile Organic Compounds), Carbon-di-oxide, Noise, Light and human presence.

Analytics approach has been adopted to slice and dice the data to find various insights. Various team discussions and reference to some research papers helped to come up with an exhaustive list of factors that contribute to changes on the parameters like temperature, humidity which in turn impact the occupant comfortability. A predictive classification model is the need of the hour to predict if the occupant will be comfortable or not in given the ambient conditions. However, lack of feedback data from occupants at this point of time created a gap to meet the requirement. Instead, data pre-processing automation and an interactive dashboard was delivered, for future analytics.

The Data Pre-processing automation tool has been developed which consist of four sub processes like Data Extraction, Abnormal Data Detection, Missing Value Analysis and Data Aggregation. Processed Data is pushed into the database for visualization purpose to understand the pattern across time, among buildings and among sensors. R shiny visualization has been developed for business user to plug and play with data. Business users can view data across time, see relationship between parameters, trends of parameters, compare the values across sensors and buildings.

Data pre-processing and visualization tool helped the business to understand more about their data, buildings in which sensors were deployed, occupant behaviour pattern across time and also to keep track of quality of the sensors.

# **INTRODUCTION**

Buildings sector contributes up to 40% of the global energy consumption, and most of this energy is accounted by existing buildings. While there is a move to replace existing buildings with new-build green buildings, the rate of this transition is only about 1.0 to 3.0% annually, exemplifying the need for rapid enhancement of energy efficiency measures in the existing buildings.

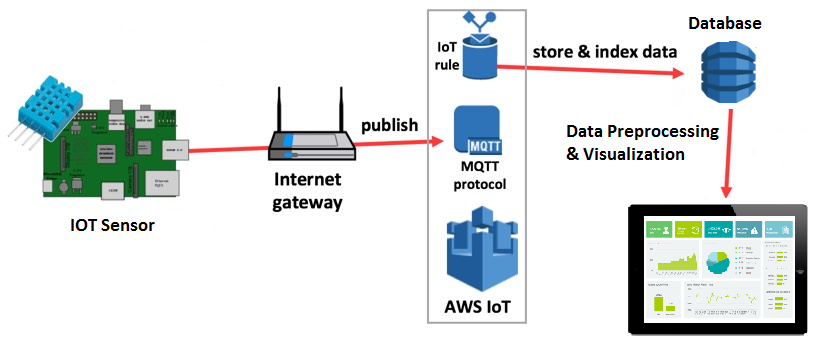
Overall building performance analysis highlights the measurement and evaluation of different performance indicators accommodating the interests of owners, facility operators, and occupants in aspects like energy consumption, thermal comfort, lighting etc.

Building performance comprises not only energy metrics but also other performance metrics such as indoor thermal comfort, visual comfort, indoor air quality, and acoustic performance. There are several parameters to be measured to assess these performance metrics where each parameter affects different building’s aspects in different level.

The research project at NUS School of Design and Environment, headed by Dr. Sekhar Kondepudi, Associate professor, Department of Building, proposes to provide occupant comfortability through intelligent and adaptive Environmental control systems for the building facilities manager to respond to changing conditions to optimise energy consumption and occupant comfortability. This will be achieved through an integrated sensor designed and developed by the Research Assistants of the project Mr. Rahul Ashok and Mr. Obi Gari Reddy.

The sensors measures indoor Temperature, light, humidity, noise, Volatile Organic Compounds, CO2. A number of these sensors have been deployed in various indoor settings inside NUS labs and offices, and extensive data has been collected over 2 years.

The resources provided for this project were reports from previous research projects and data used, and the data that was being continuously measured and pushed to Amazon Web Services Database. Access to the sites where sensors were located was provided to give better understanding of the setting in which sensors were deployed.



*Figure 1: Project Architecture*

# **BUSINESS OBJECTIVE**

The aim of the project is to provide optimal indoor conditions to the occupants contributing towards comfortability impacting their productivity, by taking into account various comfort factors together with moving towards smart building to save energy by identifying occupancy pattern.

An effective measurement of all these parameters requires real time and continuous data collection covering the entire area with minimum information loss and reduced cost w.r.t. number of sensors deployed enabling the facility operator to operate the building efficiently across all the performance metrics. Therefore, an approach to identify the optimum number of sensors and the position for sensor deployment in a given area to provide a holistic picture of the building’s overall performance is an integral part of the project.

# **DATA ANALYTICS OBJECTIVE**

To provide comfortable indoor conditions to the occupants, it is important to find various factors impacting an individual’s comfort. A lot of discussions and research was done to find out what factors other than temperature, humidity, CO2, VOC, noise and light impact comfort.

Also, the factors that impact these parameters were found. An exhaustive list of the parameters is available in Appendix A.

The initial objective was to build a predictive classification model to predict if an occupant will be comfortable or not given the occupant demographics and other factors. Sensor data alone is not sufficient for building the model, it also requires other data points for various factors which could be either collected from internet or derived using standard formulas from research papers. Also, continuous feedback data from the occupants was necessary for the purpose of model building. However, the feedback collection exercise was postponed and other data points were unavailable to build the model.

Initial data exploration demonstrated a requirement of an automated data processing pipeline followed by visualization to understand the trends and patterns to make the users understand the behaviour of the data.

Therefore, at the beginning of Phase III, the objective of the project was

* Data pre-processing pipelining and
* An interactive dashboard for diagnostic and descriptive insights.

The Success criteria of this project was a deliverable in the form of a platform to the business users that pre-processes the data to be fed to the interactive dashboard.

# **PLANNING**

* The initial model building plan required to identify factors affecting occupant comfort and also the factors that affect the parameters being measured.
  + For e.g., a printer in operation produces a significant amount of Volatile Organic Compounds and CO2, the size of the window affects the amount of light and temperature around the window, male prefer lower temperature than women, age of the occupant is a factor,

External weather conditions, External Humidity etc. these factors affect the parameter readings.

* Take a sample of data for all locations. Three months : March 2017 to May 2017
* Perform correlation with sensor data and available external data to filter out the insignificant factors.
* Perform initial data exploration to understand patterns.
  + Divide the data into working hours and Non-Working Hours
  + Compare building performance among the 4 locations.
* Create a high level data pre-processing pipelining plan.
* Create an interactive dashboard.

# **TECHNICAL STRATEGY**

The approach to automating an end-to-end data processing pipeline and visualization was to decompose the task into sub tasks. Firstly each sub task were developed and executed independently. Finally, the sub tasks were integrated into a single sequence process based on business requirement. An ‘Email alert trigger’ is also implemented to acknowledge about completion of the process.

*Figure 2: Data Pre-Process Flowchart*

The initial data engineering requirement was to process the data and store it in multiple csv files. The folder hierarchy was supposed to follow: **Location/Sensor/Year/Month/Day/Hour**

Also, the data for every 10 seconds was to be aggregated for different time intervals like 5 minutes, 15 minutes, 30 Minutes, 1 Hour, 24 Hours. This also had to be pre-computed and stored since aggregating huge volume of data on the fly for visualization was affecting the performance.

The above process was implemented but the consequences were explained to the team. It created at least 500 files every day (Memory space issue) and since it is temporal data, the division of files made it difficult and time consuming to combine for visualization purpose.

Therefore, the data engineering was changed to take data from database, process and push it back to database instead of creating daily excel files.

**VISUALIZATION:**

Initially, the Data exploration step was for the understanding of the behaviour of data to come up with insights like

* ‘Switch-off Analysis’ where based on temperature, light and humidity values on working hours and Non-working hours, weekdays and weekends, it could be found if Air Conditioners and/or lights were on during holidays.
* The analysis helped in determination of time required to cool the room once occupants started coming in and time required for the room to come back the normal temperature once Air Conditioners were switched off. This insight is useful for a facilities operator because the chillers of Air conditioner can be switched off during lunch hours to save electricity and switched back on in time without affecting the comfort of the occupants.

Through these insights, it was considered that this kind of analysis would be useful on everyday basis if provided as tool to the facilities operator. Therefore, it was required to build an interactive dashboard of the processed data as a package.

A few dashboards were designed on Tableau to understand the requirements. However, it was later decided that the dashboard be built on an open source platform such as R shiny.

# **DATA ACQUISITION AND PROCESSING**

The team provided us access to their AWS server and explained to us the different structure, the data, their definitions, the problems, timestamping issues, etc. The previous research work from other students were also provided to get us started.

Initially, only 3 months data was analysed. However, there were a lot of missing data for various sensors due to Wi-Fi issues. A rolling window technique was used to identify the three months with maximum data points for all locations and all sensors for a fair comparison.

## OUTLIER ANALYSIS

There were certain outliers, for example, temperature reading of -12 degrees or 45 degrees. These sudden values were removed. However, outlier analysis is important to the team since using this outlier data, the team can identify if the sensors have gone bad if there are abnormal data points for a long period of time or, for example, if there is fire in that place which is causing sudden high temperature, low humidity and high noise values.

Therefore, the minimum and maximum ranges of each of the parameter was obtained from the team and using this boundary, outlier were removed and stored in a separate file.

## MISSING DATA

The sensors did not post any data or empty timestamp in the event of Wi-Fi issue. This made it difficult to find out the missing data. As a result, an algorithm was implemented to create data points with missing timestamps and NULL values to determine the percentage of missing data in the given time frame. This also helps in determining the mal functioning of sensors.

## DATA AGGREGATION

The data is posted every 10 seconds and gives maximum information about the behaviour of the location. However, it creates a huge volume and practically, the parameters being measured are things that don’t change drastically in a few seconds (Except Noise). It takes at least 5 minutes for temperature to reach 22.5 degrees to 23 degrees in a controlled environment.

Therefore, the data points were aggregated at different time intervals like 5 minutes, 15 Minutes, 30 Minutes, 1 Hour and 24 Hours to provide a higher granularity view.

## PROBLEMS AND CHALLENGES

Handling huge volume of data posed a challenge while performing any processing and visualization. As a result, the data was broken down into smaller chunks.

There were few days and hours where data was missing. There are other various external factors that causes changes in sensor readings, for example, no. of people inside the room, no. of supply and return ducts, room size, etc. However, availability of these data is a challenge and hence, the effect of these factors were unaccounted for.

## COLLECTION OF NEW DATA

Site survey of the sensor location was performed where measurements of the room size, the number of Air conditioner supply and return ducts and vents, number of lights were noted down. However, these data points were not incorporated into the dataset as there was no feedback data to perform any kind of model building.

# **ANALYSIS AND MODELLING**

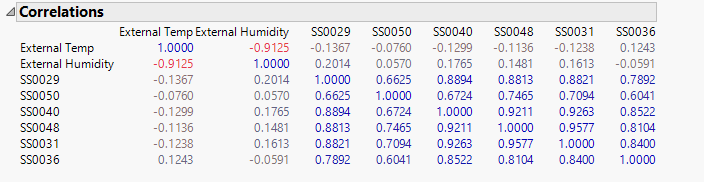
Data for three months - March 2017 to May 2017 was analysed and used to perform various combinations of **correlation test** to see if there is any significant relation between the parameters and various external factors like external temperature and humidity, which are believed to be impacting the temperature and light inside a room.

**Test of ANOVA** was performed on all sensors of one room to confirm that there is variation in mean temperature and light across weekdays and weekend.

**Autocorrelation test** was done to check if the current temperature and light is impacted by the previous temperature and light value.

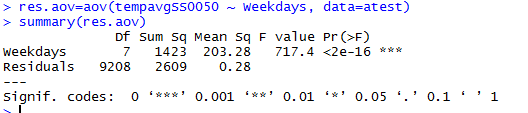
It can be noted from figure 3 that there is no correlation between internal office temperature and outside temperature and humidity.

Correlation was also performed to check if the temperature readings from sensors had correlation with other parameter readings like CO2, humidity, etc. There is moderate negative correlation between temperature and remaining parameters. This is because the air conditioners are switched on which decreases the temperature and as the number of people in the room increases, humidity, CO2, light, noise increases.



*Figure 3-Correlation between Internal Temperature and External Environment*

The result of test of ANOVA in figure 4, shows that there is significant variation in mean temperature and light when weekdays and weekends are compared. Autocorrelation test on temperature and light has not given any meaningful output. So this has been discarded for rest of parameters.



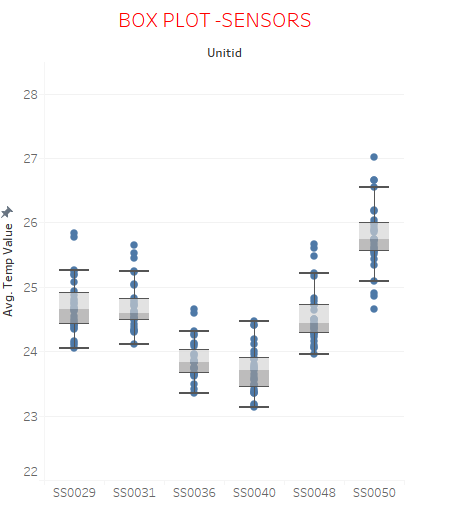
*Figure 4-ANOVA Test Output*

## EXPLORATORY ANALYSIS

The data was divided into working hours (8 am to 8 pm) and non-working hours (8:01 pm to 7:59 am), weekdays and weekend to do further analysis and understand the pattern.

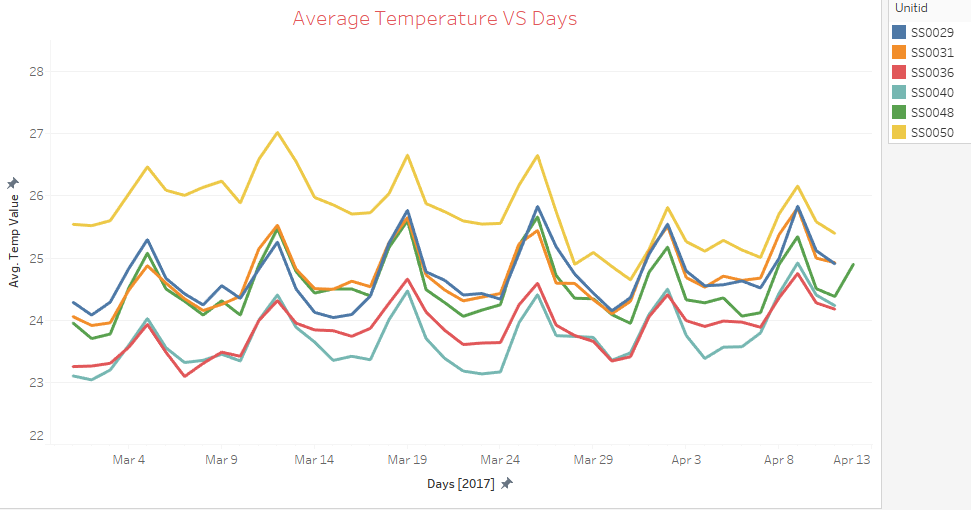
Exploratory analysis of data has been performed using Tableau to understand the pattern and behaviour of data. The will help in understanding the electricity consumption pattern which could be utilised to conserve energy when there are no occupants in the room. The data has been explored from lower granular level (in minutes) to higher level (days) to understand pattern across time and days.

Box plot was plotted using three months data to understand the data distribution of six sensors at study location. It can be seen from the figure 5, that the temperature ranges from 23°C to 27°C.

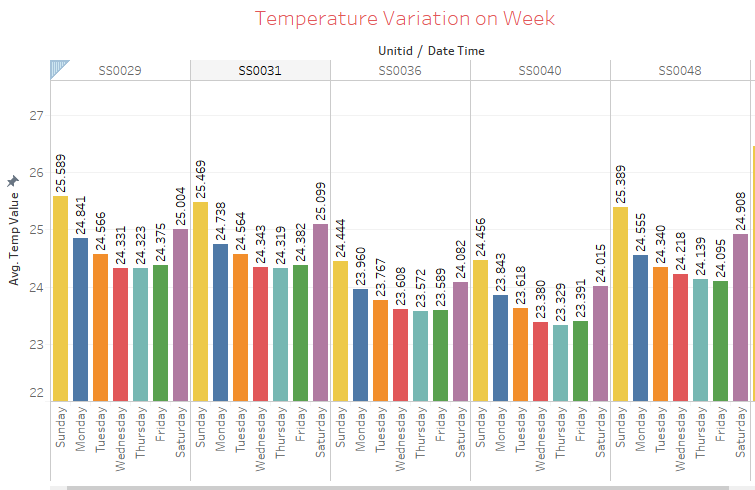


*Figure 5-Box Plot of MD11 Sensors*

Line graph has been plotted using the data from March to April to understand the behaviour of each sensor across days. It is observed from figure 6, that there is an underlying pattern followed by sensors. Sensors SS0036 and SS0040 are recording almost same temperature. This is because they are positioned close to each other. Sensors SS0029, SS0031 and SS0048 are recording almost same temperature as they are positioned close to each other. From this, it can be inferred that sensors at certain distance apart will read almost the same temperature and distance between sensors play an important role in analysis. Sensor SS0050 is reading high temperature value when compared to other sensors because it positioned near the window, whereas sensor SS0040 is reading the lowest value as it is far away from the window and close to and Air conditioner Vent. The position of sensor is another important factor that needs to be considered for building the model. The spikes in the graph appears on weekends which means that air conditioner are switched off causing increased temperature in the room.

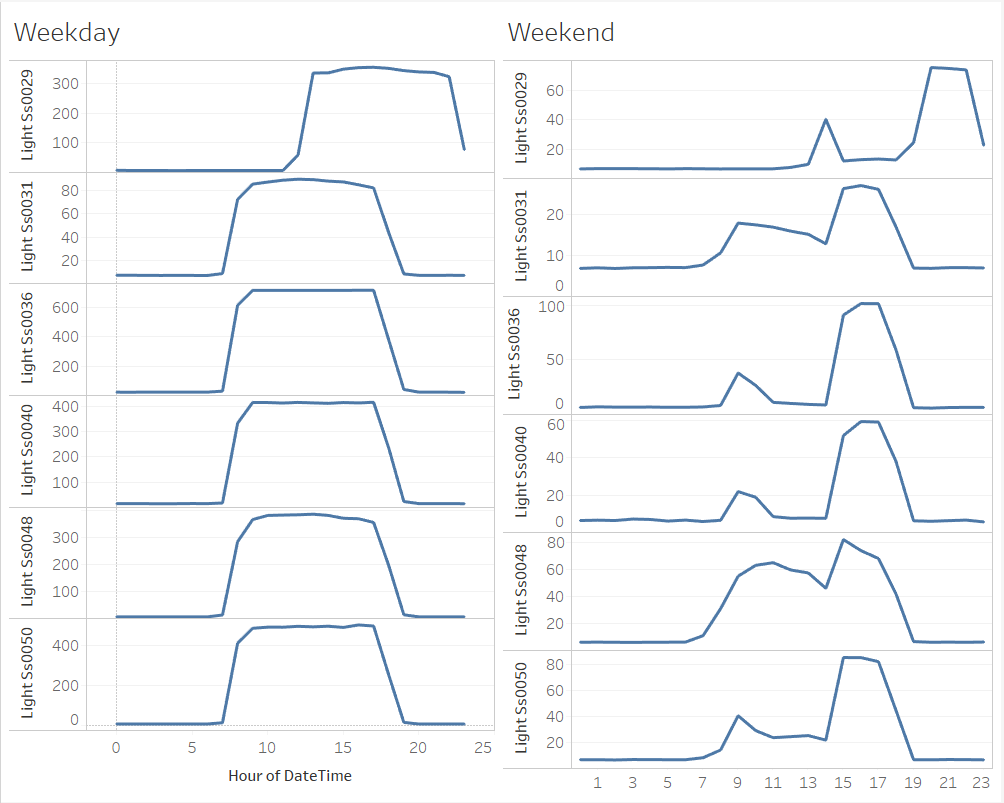
*****Figure 6: Line Graph of Sensors across two months*

To understand the average temperature value maintained on weekdays and weekends, a bar plot was plotted. It seen from figure 7 that average temperature is high on Sunday and Saturday as the air conditioner is switched off and it gradually decrease from Sunday until Monday morning and is almost constant on Tuesday, Wednesday, Thursday and Friday for all sensors.

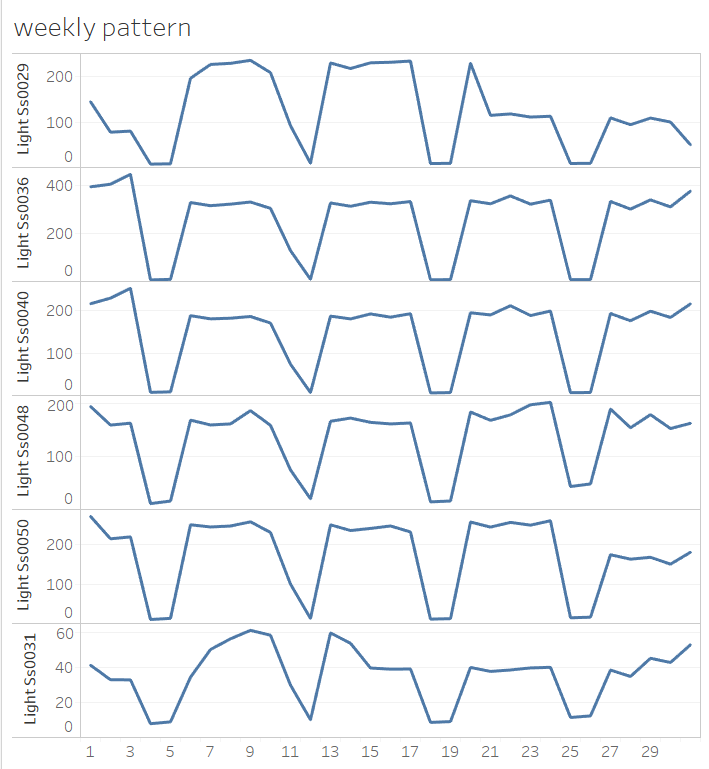
*Figure 7: Bar Graph of Sensors across Days*

The line graph was plotted to understand the behaviour of light values across weekdays and weekends. Figure 8 shows a constant reading of value 0 before 7 am and after 7 pm indicating that light has been switched off during non-working hours and between 7 am and 7 pm the lights are switched on. There is variation in the ‘lux’ (unit for light value measurement) value of each sensors because of the position of the sensors. Sensors which are directly placed under the lights will read more value than sensor away from lights, in this case sensor SS0036 is placed near the lights and its reading up to 600 lux whereas sensor SS0031 is far away from light and its reading up to 80 lux. Sensor SS0029 is positioned in a closed cabin room and there is drift in sensor value as office timing for him/her is different.

It is seen that on weekend the lights are switched off and sensor is reading the ambient light from outside through windows. The lux value is fluctuating with the time of the day and reading more during afternoon. Sensors which are close to windows will read high lux value than sensors away from window.

****

*Figure 8: Line Graph of Sensors across weekday and weekend*

In order to understand the light pattern across an entire month for each sensor, a line graph was plotted for month data as shown in figure 9. It is observed from the graph that the pattern of each sensor is almost same across days, at weekends the lux value is low and on weekdays its lux value is high. The lux value of each sensors depends on position of the sensor from the lights.

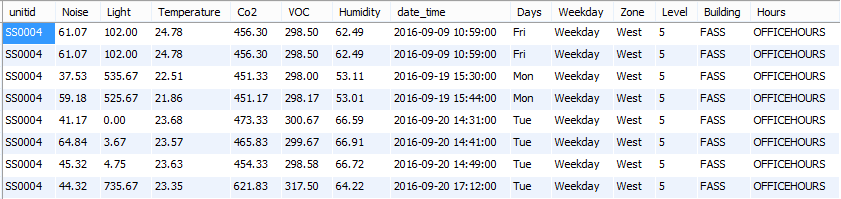
*Figure 9: Line Graph of Sensors across month*

# **VALIDATION OF RESULTS**

**DATA VALIDATION**

The following checks were performed on the data in the pipelining process:

* Proper dynamic classification of working / non-working hours based on light value
* Proper detection of Missing values
* Aggregation was completed properly for a given time interval



*Figure 10: Data Validation*

The following checks were performed in the pipelining process

**PROCESS VALIDATION**

* The sub processes in the data pipeline were running sequentially as expected
* Email Alert system working properly after process completion/ failure
* Time taken to complete the process given the volume of data

**VISUALIZATION**

* Data is being fetched properly
* Interactive enough for business users
* Able to answer the questions of the business users
* Able to weave a story by going from Building level to drilling down to sensor level

# **DASHBOARD DESIGN**

**Building Level**



User selects the Location and the time granularity

Pattern of the sensors across time.

Overall comparison of all locations for all parameters w.r.t mean, min and max values

Range Comparison of all the sensors present at that location

Selection of Parameter

Figure 1: Dashboard Design- Building level



Selection of Sensor

Outlier Analysis w.r.t to time, of the selected parameter/variable

Trend Analysis of the selected parameters/variables

Trend Analysis of the Selected X and Y axis across different time divisions of the selected sensor.

Select X and Y variable, Weekday vs Weekend, Working hours vs non-working hours and time frame for correlation analysis.

Summary statistics and percentage missing data

Figure 1: Dashboard Design - Sensor Level

# **CONCLUSION**

A platform has been created for non-technical/ business users who can plug and play with sensors data to understand more about the data at different building levels, sensor levels across day and time. A roadmap has been created for the future analytics by discovering what others parameters values needs to be collected in order to achieve ultimate goal of business.

Automated data Pre-processing, visualization tool and exhaustive list of factors directly/indirectly contributing occupant comfortability has been delivered to the business

# **RECOMMENDATION FOR FUTURE WORK**

The following are the recommendations that could help in achieving the predictive modelling.

* Collect the feedback data from occupant
* Collect value of missing factors identified through research

Other use case where this idea of continuous environmental monitoring can be beneficial would be to deploy the sensors in areas like agricultural fields to aid higher yield, and logistics and supply chain to reduce spoilage of perishable items.

# **REFERENCES**

1. Success stories in building energy efficiency – Stacey Davis – Centre for clean air policy
2. Zhenjun, M., Cooper, P., Daly, D., and Ledo, L. 2012. Existing building retrofits: Methodology and state-of-the-art. Energy and Buildings. Vol. 55 (Dec. 2012), 889-902.DOI= http://doi:10.1016/j.enbuild.2012.08.018
3. F.Liu, A. S. Meyer, and J. Hogan, Mainstreaming building energy efficiency codes in developing countries: global experiences and lessons from early adopters. World Bank Publications, 2010.
4. S. Wang, C. Yan, and F. Xiao, “Quantitative energy performance assessment methods for existing buildings,” Energy Build., vol. 55, pp. 873–888, 2012.
5. Asadi, E., Silva, M.G., Antunes, C.H., and Dias, L. 2012. Multi-objective optimization for building retrofit strategies: a model and an application. Energy and Buildings. Vol. 41
6. Ma, Z. and Wang, S. 2009. Building energy research in Hong Kong: A review. Renewable and Sustainable Energy Reviews. Vol. 13 (2009), 1870-1883. DOI=http://doi:10.1016/j.rser.2009.01.006
7. Chidiac, S.E., Catania, E.J.C., Morofsky, E., and Foo, S. 2011. Effectiveness of single and multiple energy retrofit measures on the energy consumption of office buildings. Energy and Buildings. Vol. 36 (Aug, 2011), 5037-5052.
8. X.-G. Xia, X. Ye, and J. Zhang, “Optimal metering plan of measurement and verification for energy efficiency lighting projects,” in Energy Efficiency Convention (SAEEC), 2012 Southern African, 2012, pp. 1–8.
9. Augenbroe, G., Park, C. S. 2005. Quantification methods of technical building performance. Building Research and Information 33 (2): 159-72.
10. Dimosthenis Ioannidisa,b, Pantelis Tropiosa, Stelios Krinidisa,∗,George Stavropoulosa, imitrios Tzovarasa, Spiridon Likothanasisb. Occupancy driven building performance assessment. Journal of innovation in digital eco systems 3 (2016) 57-69
11. Natural Resources Canada’s Canmet ENERGY, Overview of different measurement and verification protocols - Report no. 2008-069\_TR\_424-CONOPT
12. Z. Ma, P. Cooper, D. Daly, and L. Ledo, “Existing building retrofits: Methodology and state-of-the-art,” Energy Build., vol. 55, pp. 889–902, 2012.
13. J. Granderson, M. A. Piette, and G. Ghatikar, “Building energy information systems: user case studies,” Energy Effic., vol. 4, no. 1, pp. 17–30, 2011.
14. J. Kummer, “MeasureMent and verification of energy savings,” Inst. Build. Effic., 2011.
15. Ramtin Attar, Ebenezer Hailemariam, Simon Breslav, Azam Khan, Gord Kurtenbach, Sensor-enabled Cubicles for Occupant-centric Capture of Building Performance Data.
16. Martín-Gómez, César1, Vidaurre-Arbizu, Marina1, Eguaras Martínez, María1, Krinidis, Stelios2, Ioannidis, Dimosthenis2 & Tzovaras, Dimitrios 2 by Sensor Placement for BPM Analysis of Buildings in Use to Implement Energy Savings Through Building Performance Simulation. Journal of Engineering and Architecture, Vol. 2(2), December 2014.
17. A. Ahmed, J. Ploennigs, Y. Gao & K. Menzel, Analyse building performance data for energy-efficient building operation. IRUSE, University College Cork, Ireland
18. D. J. Chmielewski, T. Palmer, and V. Manousiouthakis, “On the theory of optimal sensor placement,” AIChE J., vol. 48, no. 5, pp. 1001–1012, 2002.
19. F. E. Udwadia, “Optimal sensor location for geotechnical and structural identification,” in
20. 8th WCEE, 1984, pp. 315–322.
21. D. C. Kammer, “Sensor placement for on-orbit modal identification and correlation of large space structures,” J. Guid. Control. Dyn., vol. 14, no. 2, pp. 251–259, 1991.
22. C. Papadimitriou, J. L. Beck, and S.-K. Au, “Entropy-based optimal sensor location for structural model updating,” J. Vib. Control, vol. 6, no. 5, pp. 781–800, 2000.
23. C. Papadimitriou, “Optimal sensor placement methodology for parametric identification of structural systems,” J. Sound Vib., vol. 278, no. 4, pp. 923–947, 2004.
24. H. Xiao, Q. Wei, and Y. Jiang, “The reality and statistical distribution of energy consumption in office buildings in China,” Energy Build., vol. 50, pp. 259–265, 2012.
25. N. Gaitani, M. Santamouris, G. Mihalakakou, and P. Patargias, “Cluster Analysis in Energy Classification of School Buildings,” in Proceedings of the EPIC 2006 AIVC Conference, Lyon, 2006.

# **APPENDIX A**

|  |
| --- |
| **Parameters** |
| Effect of External Temperature |
| Effect of Office Size |
| Effect of Floor to ceiling height |
| Effect of Ambient Light |
| Presence of window |
| Distance from window |
| Effect of closed office /Open office |
| Closed Office |
| Number of people in closed office |
| Open Office |
| Number of people in open office |
| Effect of Number of Vents - Supply |
| Effect of distance between supply vents and sensor |
| Effect of Air Velocity from Supply Vents |
| Effect of Air Temperature of supply vent |
| Effect of Number of Vents - Return |
| Effect of Air Temperature of return vent |
| Effect of distance between return vents and sensor |
| Effect of Electrical devices - Energy consumed |
| Effect of Distance between Electrical device and sensor |
| Effect of Number People |
| Effect of Time of Day |
| Effect of Day of Week (Weekday vs Weekend) |
| Effect of the floor number of building |
| Effect of flooring type |
| Effect of CupBoard |
| Effect of Plant |
| Effect of External Humidity |
| Effect of Haze |
| Effect of Cloudiness |
| Effect of Pressure |
| Effect of Orientation of building |
| Effect of Office types |
| Effect of Mobile phone |
| Effect of Landline |
| Effect of use of Cleansers, Aerosols, Disinfectants |
| Effect of Mopping |
| Effect of Surrounding |
| Effect of Radiant Temperature |
| Effect of Work Rate / Metabolic Heat |
| Effect of Furniture and Chairs |