Data Analytics Practicum Report

City Digital Smart Building Data Services

Industry Partners:

UI Labs-City Digital, Accenture, Siemens, Microsoft, Tyco

Professor: Dr. Shlomo Argamon

Students: Sachet Misra, Feixiang Xi, Yufei Wu

08/15/2016

Problem domain and the Task

Background

Cities, universities and large scale building owners need to manage a dynamic and multifaceted real estate portfolio, typically full of untapped and uncorrelated data sources. Aim to make an effort to reduce energy and operational expenses related to portfolio management, and seeks solutions from data analysis perspective, by testing smart buildings data to discover new methods of applying large-scale data analysis to impact maintenance improvements and energy efficiencies.

Objective

The Practicum should capture and analyze a defined set of data inputs to gauge the ability to deliver recommendations and algorithms that will produce insights on targeted energy savings and maintenance efficiency and design statistical algorithms to deliver these insights wherever feasible. Give feedback of whether the data provided yield enough information to derive recommendations on energy efficiency and preventative maintenance of operational systems at IIT.

Tasks

1 Data profiling: Get the raw data steam, clean the data, Integration the data into the form that can be used for further analysis.
2 Performance Clustering: Identify patterns and trends and identify poor performing assets based on some criteria from the data.
3 Performance Prediction: Prediction of future performance base on the aggregation of historical data, identify additional data sets that helps the prediction.

1. Data Profile

1.1 Report on methods for retrieving data, including understanding of data formats currently available

This report is limited to the data we have received and the data we have seen. It will only concentrate on the databases we have had access to and the file formats we have received to work on. Plus the dataset is limited to information available to us only of the IIT tower (IIT Research Tower)

1.1.1 Siemens Data:

This dataset had 5111 text files which were daily updates of the devices (AHU01, AHU02, AHU03, AHU05, AHU06, AHU07 ,AHU08, AHU10, AHU11 ,AHU12, AHU13, AHU14, AHU15, AHU16, AHU17, AHU18, AHU19, AHU20, CHW, RTU, Exhaust Fans). The files were converted to excel format, categorised into each of its names which added up to a total of 22 files in the dataset. We used excel and VB macros for joining the files and making them one of each category.

Equipment ID	*Area Served*
AHU-S1	Tower Basement
AHU-S2	Flrs 2-10 S&W Perimeter
AHU-S3	Flrs 2-10 N&W Perimeter
AHU-S5	Tower Shipping/Receiving
AHU-S6	Flrs 2-10 W Interior
AHU-S7	Lobby W
AHU-S8	Flrs 2-10 E Interior
AHU-S10	Flrs 2-10 N&E Perimeter
AHU-S11	Flrs 2-10 S&E Perimeter
AHU-S12	
AHU-S13	Lobby E
AHU-S14	Flrs 11-19 W Interior
AHU-S15	Flrs 11-19 N&W Perimeter
AHU-S18	Flrs 11-19 E Interior
EF-14 & 19	Toilet Exhaust

Table 1 Siemens Data Area Served

1.1.2 Work order data:

The dataset had work order data of various buildings which were arranged based on when a call was made. The IIT Research Tower data was extracted from here. The full data was got in the xls format which was not fully usable on the softwares from Microsoft (Excel 2016 and PowerBI). We used python for data analysis and extraction of the data.

1.2 Identify difficulties in the data – cleanliness, missing values, integration issue

1.2.1 Work Order Dataset

The dataset - table is divided into Wo Number, Description, Wo Number, Description, Assigned To, Building, Floor, Room, Craft, Enter Date, Wo Close Date, Hours SUM(estimate)

- Most of the descriptions are standard but some of them have an overlap in the function but are worded differently.
- The floor column is not complete but it could be completed while looking into the Description column and column number is looked for.
- Hour SUM, Room columns have a lot of NULL values.

The following graph helps to find which type of work order takes most labor hours for each buildings, from the figure below, it shows that Room Turns takes most labor hours(show in red)

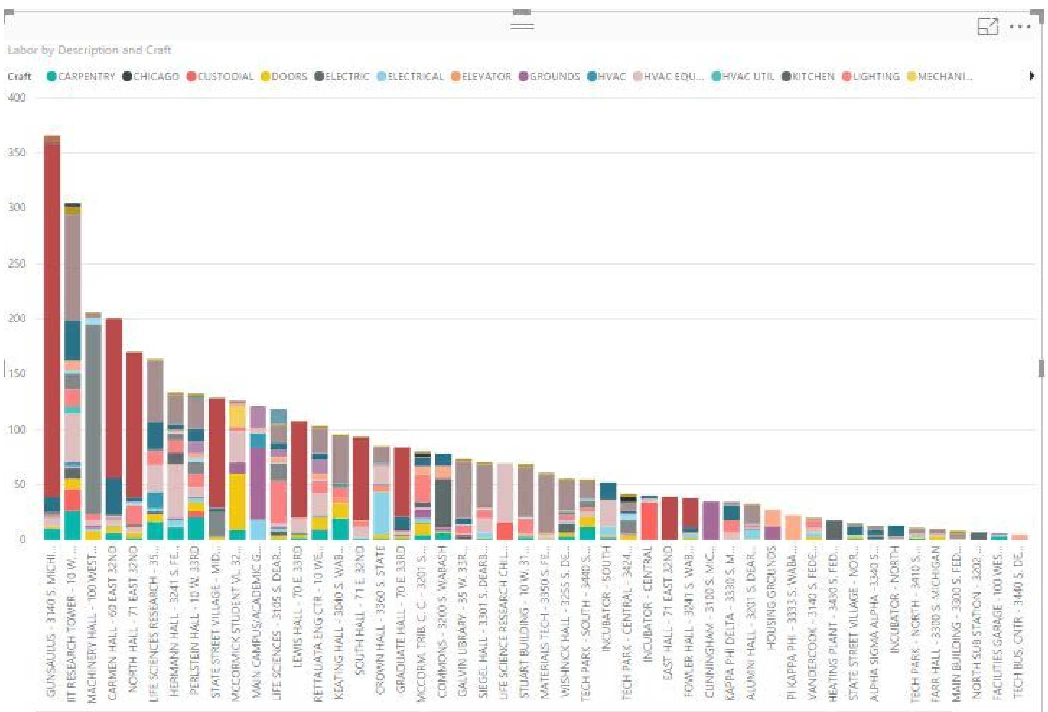


Figure 1 Buildings Work Order Stacked Bar Chart

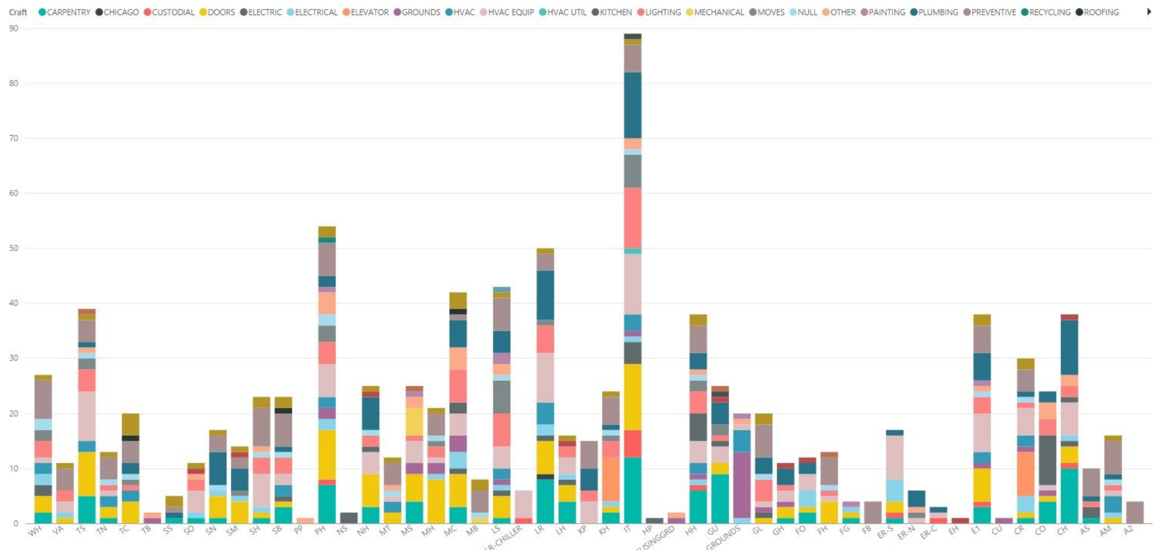


Figure 2 Buildings Work Order Distribution

Result: The above graphs are work order Stacked-Bar Chart. The IIT Research Tower was chosen because it has a variety of work order data and that we can get more value out of this data.

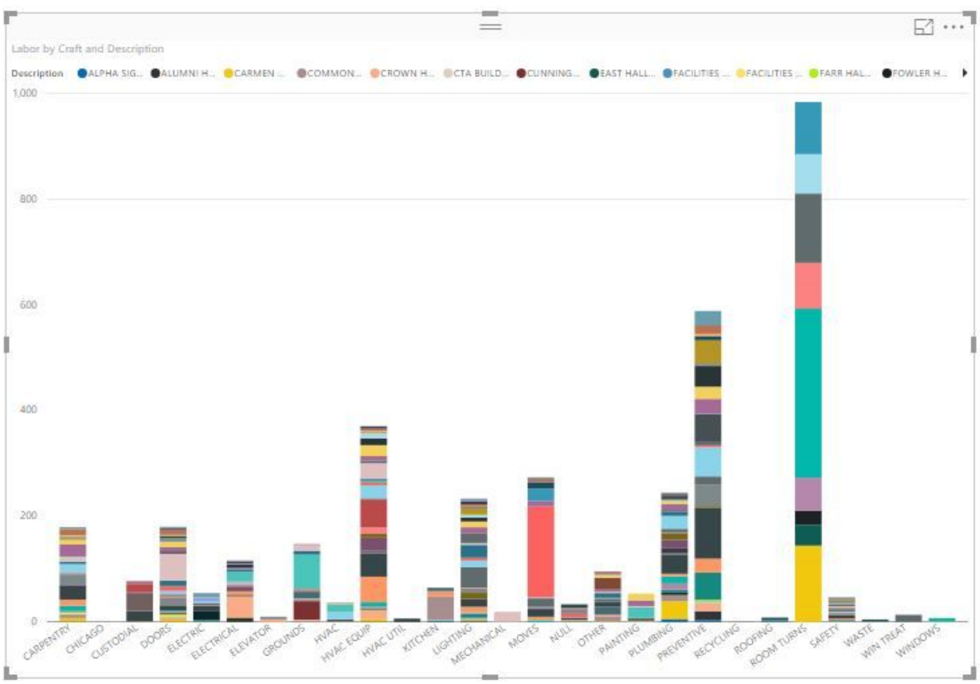


Figure 3 Work Order Kinds Distribution

The above graph is the work order plot across various datasets. This gives us an idea with the calls from various kinds of calls. *The assumption here is that there are no custom calls (all calls types are predefined - note: This assumption is removed later).*

Heat Map

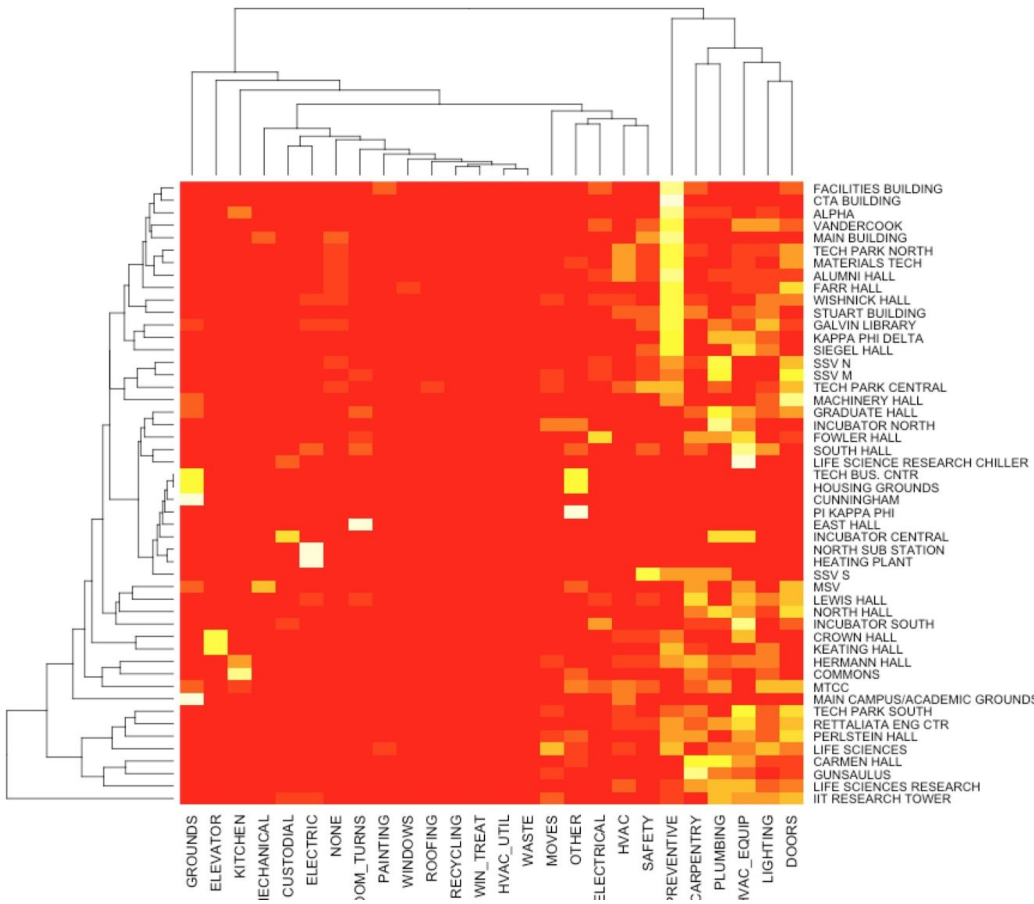


Figure 4 Heat Map for Work Order Data

The heat map shows the distribution of the numbers of 26 kinds work orders in all buildings, and light color means larger number. On the upper side of the graph is the clustering result of the work order numbers. The figure shows the structure of buildings work order distributions. With help of this map, we can understand which buildings are more similar with some other buildings.

Keyword analysis:

Some of the keywords were identified to find the most prominent kind of job. These were used to identify the jobs from various floors also. The floor column has a few entries which aren't entered but it's possible to retrieve the floor number from the description. This analysis also helps us find and analyse the description that were not pre-defined.

The following are the keywords performed using frequency analysis over the words in the Description column:

Keyword	Frequency (IITResearchTower)
HVAC	7
AHU	1
FILTER	32
FAN	32
COIL	16
TOO HOT	85
TOO COLD	125
HEATING	3
COOLING	17
TEMPERATURE	6
CHILLER	83
DAMPER	8
STEAM	6
COLD	129
REPAIR	369
BASEMENT	64
SUPPLY	30
FIRE	44
URGENT	67
NOISE	43
GOLF	8
ROOM	548
BULB	147
LEAKING	46
A/C	10
20th	25

19th	54
18th	6
17th	16
16th	23
15th	27
4th	56

Table 2 IIT Research Tower Frequency

1.2.2 Energy dataset:

The dataset has energy data (steam, water, electricity (IIT research Tower))

The following graph shows the various levels of energy data we have based on the levels of energy we have. The more the amount of entries / energy data the lower the level of the building data.

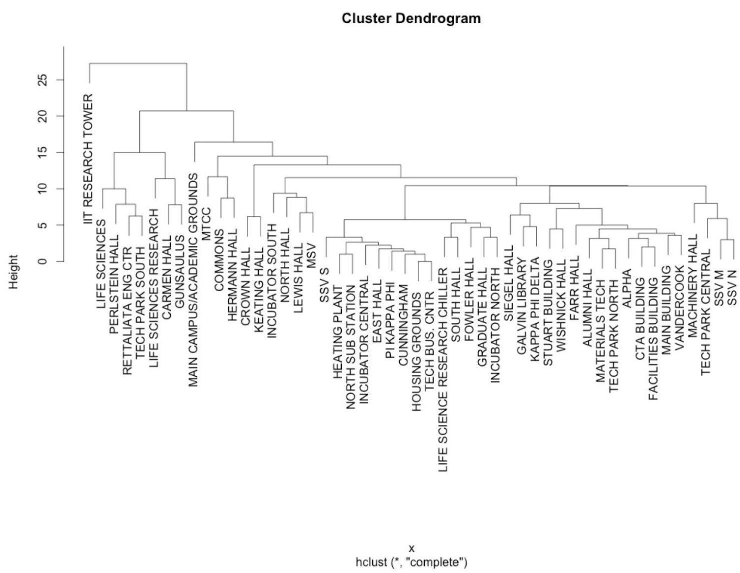


Figure 5 Cluster Dendrogram for Building Energy Consumption

Result: The data at the IIT Research Tower seems to be the most amount of data.

By plot the energy consumption, it helps to find the potential cause for earlier figure of the energy consumption Clustering Dendrogram.

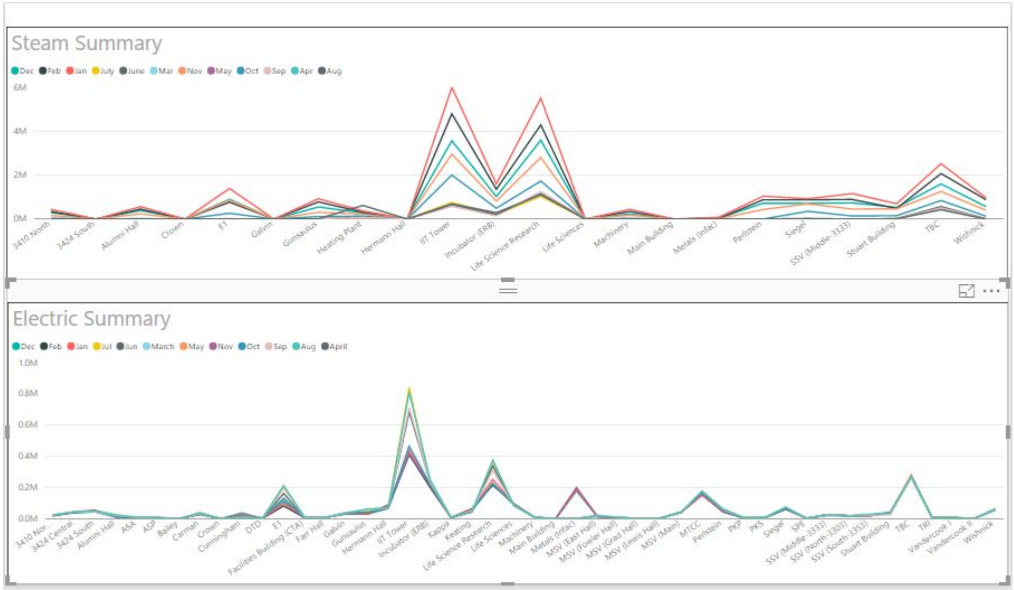


Figure 6 Steam and Electric Summary Visualization

Result: From the Figure 6, we can find IIT tower and Life science building consume most steam and electric, that's why this two buildings away from other buildings in the Clustering Dendrogram

1.2.3 AHU dataset:

All the AHU data have null values which store data only when there is a change in value. Forward fill was used to fill in the data The data is updated at an average of every 15 minutes. There are a few datasets which have missing values (apart from the null values). Some of them are, AHU15, AHU16, AHU16, AHU17, AHU18, AHU19, AHU20. The rows are filled with their average values. The data is relatively clean and easy to use.

1.2.4 RTU dataset

The ones are below are the unique values in the each of the columns in the table. The unique data values highlighted (***bold, italic***) are the errors in that column. This dataset needs better storage of data or a better ETL process.

Economiser position %

'0' ***'TIME (0)' 'TIME (1)'*** '10' '41' '45' '46' '48' '49' '47' '18' '40' '42' '43' '17' '29' '15' '16' '44' '31' '50' '13' '30' '34' '35' '37' '38' '28' '12' '25' '23' '27' '32' '33' '26' '11' '39' '21' '22' '24' '19' '36' '14' '51' '64' '20' '60' '59' '52' ***'TIME (21)' 'TIME (-19)' 'TIME (-20)' 'TIME (-21)' 'TIME (22)' 'TIME (23)' 'TIME (-22)'*** '68' '73' '66' '62' '71' '53' '63' '58' '72' ***'TIME (-23)' 'TIME (24)' 'TIME (25)' 'TIME (-24)' 'TIME (26)' 'TIME (-25)'*** '100' '93' '99' '89' '61' ***'TIME (-26)' 'TIME (27)' 'TIME (3601)'***'74' '55' '70' '54' ***'TIME (-28)' 'TIME (29)' 'TIME (-27)' 'TIME (28)' '56' 'TIME (30)' 'TIME (-29)' 'TIME (31)' '90' 'TIME (-30)'*** '97' '83' ***'TIME (-31)' 'TIME (32)' '65' 'TIME (-32)' 'TIME (33)' '57' 'TIME (-33)' 'TIME (34)'*** '81' '86' '91' '75' '82' '67' '96' ***'TIME (35)' 'TIME (-34)' '95' '76' 'TIME (-35)' 'TIME (36)' 'TIME (-36)' 'TIME (37)'*** '84' '88' ***'TIME (-37)' 'TIME (38)'*** '92' '98' ***'TIME (-38)' 'TIME (39)'*** '69' '78' '79' '80' '85' '87' '94' ***'TIME (-39)' 'TIME (40)' 'TIME (-40)' 'TIME (41)' 'TIME (-1)' 'TIME (-3600)' 'TIME (2)' 'No Data'***

Supplier Temperature

'59.5' '60.4' '61' '61.1' '61.2' '60.9' '61.4' '61.6' '61.3' '61.7' '61.9' '60.8' '60.7' '60.3' '60.6' '58' '59.7' '58.6' '58.5' '59.2' '60' '58.8' '56.5' '59.8' '62.4' '65.5' '66.5' '64.8' '65.6' '65.8' '67.2' '67.9' '68.2' '68.7' '68.6' '68.9' '68.8' '69.6' '69.4' '68.4' '68.3' '65.7' '65.1' '64.9' '64' '64.1' '62.7' '57.4' '54.5' '58.4' '60.1' '59.1' '58.1' '61.5' '62.1' '56' '53.8' '54.7' '57.8' '55.9' '56.9' '57.2' '53.6' ***'TIME (0)'*** '54.6' '55.2' '55.8' '54.1' '52.6' '53.9' '51.4' '50.3' '52.2' '51.8' '63.3' '65' '65.3' '66.4' '66.7' '66.3' '66.6' '67.1' '67.3' '67' '67.4' '67.6' '67.8' '67.5' '68.1' '67.7' '66.1' '61.8' '62.6' '62.3' '62.2' '62' '57.7' ***'TIME (1)'*** '56.4' '56.8' '54' '55.3' '57.1' '55.7' '60.2' '59.9' '58.9' '57.6' '66.9' '68' '69.7' '69.1' '69.3' '69.2' '69.8' '69.5' '70' '70.2' '70.1' '70.3' '70.7' '70.4' '62.9' '64.3' '63.6' '66.2' '56.3' '55.1' '59.4' '58.2' '56.6' '56.2' '53.3' '48' '46' '43.3' '45.1' '46.8' '45.7' '46.5' '48.3' '48.8' '49.3' '46.9' '48.2' '46.4' '43.6' '44.4' '38.4' '40.2' '37.8' '39.7' '41.3' '40.9' '41.8' '40.1' '39.8' '42.1' '43' '41.6' '42.2' '40.6' '43.2' '42.9' '43.4' '38.3' '37.2' '38' '39.1' '50.5' '57.3' '63.9' '66' '64.7' '64.4' '63' '62.8' '64.2' '64.5' '65.4' '65.2' '63.4' '63.5' '59' '54.8' '53.7' '48.6' '48.9' '54.4' '54.3' '47.5' '47.9' '48.7' '49.5' '49.9' '49.6' '48.5' '48.4' '51' '53.4' '58.3' '55.6' '55' '55.5' '55.4' '46.2' '51.7' '52.3' '51.3' '45.4' '45.5' '45.9' '50.8' '44.5' '44.8' '44.7' '45.2' '50.1' '56.7' '51.1' '56.1' '50' '51.2' '52.4' '53' '54.2' '59.3' '54.9' '46.7' '47' '48.1' '57.5' '63.2' '58.7' '63.1' '62.5' '60.5' '59.6' '63.7' '64.6' '65.9' '52.8' '53.5' '49.8' '52.7' '52' '51.5' '43.8' '42.8' '47.4' '52.5' '53.1' '47.1' '44.9' '51.6' '50.9' '47.2' '52.1' '49' '42.3' '42.6' '41.7' '41.9' '50.6' '49.2' '45.8' '44' '47.6' '45.6' '38.7' '50.2' '49.4' '46.3' '45' '38.9' '44.6' '41.4' '47.7' '43.9' '41.5' '46.6' '40' '41.2' '47.3' '50.7' '42' '49.1' '44.1' '57' '41' '50.4' '40.7' '43.5' '40.5' '40.4' '45.3' '42.4' '44.2' '47.8' '37.5' '38.1' '42.7' '43.7' '57.9' '52.9' '63.8' '49.7' '53.2' '46.1' '39.6' '39.5' '38.6' '36.1' '36.3' '38.2' '38.5' '44.3' '37.1' '36.4' '36.9' '35.3' '36.5' '38.8' '41.1' '39.4' '40.3' '39.9' '36.6' '37.3' '37.4' '43.1' '51.9' '73.6' '73' '72' '71.3' '71.4' '71.5' '71.1' ***'TIME (21)' 'TIME (-19)' 'TIME (-20)'*** '40.8' '42.5' ***'TIME (-21)'*** '39' ***'TIME (22)' 'TIME (23)' 'TIME (-22)'*** '70.5' '71.6' '71' '69' '68.5' '70.9' ***'TIME (-23)'*** 'TIME (24)' '79.8' '72.6' '75.4' '66.8' ***'TIME (25)' 'TIME (-24)'*** '69.9' '71.8' '70.8' '85.5' '75.9' '73.1' '73.4' '74.1' '74' '75' '74.5' '74.4' '73.9' '74.3' '72.8' '72.1' '71.9' ***'TIME (26)' 'TIME (-25)'*** '71.2' '70.6' '72.9' '72.4' '72.7' '74.2' '74.7' '73.5' '75.2' '76.9' '76.5' '75.5' '75.3' '75.6' '75.8' '77.2' '76.4' '74.6' '76.1' '76.2' '76.8' '77' '77.1' '77.3' '77.7' '78.4' '78.5' '79.1' '78.7' '78.6' '79' '77.9' '78.3' '76.3' '72.5' '71.7' '72.3' '73.3' '72.2' '73.2' ***'TIME (-26)' 'TIME (27)' 'TIME (3601)' 'TIME (-28)' 'TIME (29)' 'TIME (-27)' 'TIME (28)'*** '73.8' '74.8' '75.1' '74.9' '***TIME (30)' 'TIME (-29)' 'TIME (31)'*** '73.7' '75.7' '76' ***'TIME (-30)' 'TIME (-31)' 'TIME (32)'*** '80.3' '37.7' '37' '36.7' '36.2' '37.6' ***'TIME (-32)' 'TIME (33)' 'TIME (-33)' 'TIME (34)' 'TIME (35)' 'TIME (-34)'*** '76.6' ***'TIME (-35)' 'TIME (36)' '76.7' 'TIME (-36)' 'TIME (37)'*** '79.3' ***'TIME (-37)' 'TIME (38)' 'TIME (-38)' 'TIME (39)'*** '77.5' '78' '78.1' '77.4' ***'TIME (-39)' 'TIME (40)'*** '77.8' '78.9' '78.2' '78.8' ***'TIME (-40)' 'TIME (41)'*** '79.5' '79.6' '80.2' '79.4' '77.6' '79.2' ***'TIME (-1)' 'TIME (-3600)' 'TIME (2)' 'No Data'*** '83.7'

Room Temperature

'80.1' '80.2' '80.4' '80.5' '80.6' '80.7' '80.8' '80.9' '81' '81.2' '79.6' '78.6' '79.1' '79' '78.5' '78.3' '78.1' '77.9' '77.7' '77.6' '77.5' '77.3' '77.4' '77.2' '77.1' '77' '76.9' '76.8' '76.6' '76.5' '76.7' 'TIME (0)' '76.4' '76.3' '76' '76.1' '75.9' '76.2' '78.9' '79.4' '79.7' '79.8' '79.9' '80' '80.3' ***'TIME (1)'*** '81.1' '79.2' '78.8' '78.7' '78.4' '78.2' '78' '75.8' '75.7' '75.5' '75.3' '75.2' '75' '75.1' '75.4' '75.6' '74.8' '74.7' '74.6' '74.5' '73.9' '72.4' '73.2' '72.8' '72.1' '71.7' '71.4' '74.1' '72.9' '73.3' '73.5' '72.2' '72.3' '73.8' '74.2' '74.3' '71.2' '71.6' '71.9' '72' '71.8' '72.7' '71.3' '71.5' '74.9' '74.4' '73.6' '73.1' '73' '71.1' '74' '73.7' '72.5' '72.6' '73.4' '70.9' '70.5' '70.7' '70.8' '70.1' '70.6' '70.3' '71' '69.8' '69.6' '70' '69.9' '68.8' '70.2' '70.4' '79.5' '79.3' ***'TIME (21)' 'TIME (-19)' 'TIME (-20)' 'TIME (-21)'*** '69.3' ***'TIME (22)' 'TIME (23)' 'TIME (-22)' '77.8' 'TIME (-23)' 'TIME (24)' 'TIME (25)'***

'TIME (-24)' '81.7' '81.6' '81.8' '82' '81.9' '82.1' '82.2' '82.3' '82.4' '81.3' '81.4' ***'TIME (26)' 'TIME (-25)'*** '81.5' '82.5' '82.6' '82.8' '82.7' '83' '83.1' '83.2' '83.3' '83.4' '83.5' '83.6' '83.7' '83.8' '83.9' '84' '84.1' '84.2' '84.3' '84.4' '84.5' '84.7' '84.9' '85' '85.1' '85.2' '85.4' '85.5' '85.3' '85.6' '85.8' '85.9' '86' '85.7' ***'TIME (-26)' 'TIME (27)' 'TIME (3601)' 'TIME (-28)' 'TIME (29)' 'TIME (-27)' 'TIME (28)' 'TIME (30)' 'TIME (-29)' 'TIME (31)' '82.9' 'TIME (-30)'*** '69.7' '69.5' '69.4' ***'TIME (-31)' 'TIME (32)' 'TIME (-32)' 'TIME (33)' 'TIME (-33)' 'TIME (34)' 'TIME (35)' 'TIME (-34)'*** '69' '68.5' '69.2' ***'TIME (-35)' 'TIME (36)' 'TIME (-36)' 'TIME (37)'*** '68.9' '68.7' '68.3' '67.9' ***'TIME (-37)' 'TIME (38)' 'TIME (-38)' 'TIME (39)' 'TIME (-39)' 'TIME (40)' 'TIME (-40)' 'TIME (41)' 'TIME (-1)' 'TIME (-3600)' 'TIME (2)' 'No Data'***

<u>1.2.5 Tower 20th KWH dataset</u>

This dataset is not usable as there is a lot of negative energy data which does not make sense. There was some exploratory analysis done on the dataset to get the dataset ready to fill it up with average values and stuff.

1.3 Inventory the data calculating the source and frequency of the data sets

AHU01	Overall	S01SAF	S01CCO		S01DAT	S01EFCAWA									S01SFSPD
	0:04:53	11:52:29	0:14:59		0:14:59	12:06:08									0:14:59
AHU02	Overall	S2	S2CCV		S2DAT	S2E2		S2MAD	S2MAT	S2PHT	S2PHV		S2RH1T/2T	S2R1V/2V	
	0:01:32	9:49:26	0:14:59		0:14:59	12:00:00		0:14:59	0:14:59	0:14:59	0:14:59		0:14:59	0:14:59	
AHU03	Overall	S3	S3CCV		S3DAT			S3MAD	S3MAT	S3PHT	S3PHV	S3RAT	S3RH1T/2T	S3R1V/2V	
	0:01:23	9:50:27	0:14:59		0:14:59			0:14:59	0:14:59	0:14:59	0:14:59	0:14:59	0:14:59	0:14:59	
AHU05	Overall	S5	S5CCV		S5DAT					S5PHT	S5PH1V/2V			S5RHV	
	0:14:19	7:03:25	0:14:53		0:14:53					0:14:53	0:14:53			0:14:53	
AHU06	Overall	S6SAF	S6CCV	S6CDT		S6E1	S6HDT	S6MAD	S6MAT	S6PHT	S6PHV	S6RAT		S6RHV	S06SFVFD
	0:14:09	6:40:06	0:14:53	0:14:53		9:38:26	0:14:53	0:14:53	0:14:53	0:14:53	0:14:53	0:14:53		0:14:53	0:14:53
AHU07	Overall	S7	S7CCV	S7CDT		S7E3	S7HDT	S7MAD	S7MAT			S7RAT		S7RHV	
	0:02:11	9:42:23	0:14:59	0:14:59		11:11:49	0:14:59	0:14:59	0:14:59			0:14:59		0:14:59	
AHU08	Overall	S8	S8CCV	S8CDT		S8E6	S8HDT	S8MAD	S8MAT	S8PHT	S8PHV	S8RAT		S8RHV	
	0:01:42	10:34:21	0:14:59	0:14:59		10:58:50	0:14:59	0:14:59	0:14:59	0:14:59	0:14:59	0:14:59		0:14:59	
AHU10	Overall	S10	S10CCV		S10DAT	S10E7		S10MAD	S10MAT	S10PHT	S10PHV	S10RAT	S10RH1T/2T	S10R1V/2V	
	0:01:23	10:14:55	0:14:59		0:14:59	10:51:18		0:14:59	0:14:59	0:14:59	0:14:59	0:14:59	0:14:59	0:14:59	
AHU11	Overall	S11	S11CCV		S11DAT			S11MAD	S11MAT	S11PHT	S11PHV		S11RH1T/2T	S11R1V/2V	
	0:01:32	9:50:27	0:14:59		0:14:59			0:14:59	0:14:59	0:14:59	0:14:59		0:14:59	0:14:59	
AHU12	Overall	S12	S12CCV		S12DAT	S12EF1		S12MAD	S12MAT		S12PHV	S12RAT			
	0:01:54	9:37:28	0:14:59		0:14:59	9:35:31		0:14:59	0:14:59		0:14:59	0:14:59			
AHU13	Overall	S13	S13CCV	S13CDT		S13E8	S13HDT	S13MAD	S13MAT			S13RAT			
	0:02:11	10:35:32	0:15:03	0:15:03		9:43:23	0:15:03	0:15:03	0:15:03			0:15:03			
AHU14	Overall	S14	S14CCV	S14CDT		S14E13	S14HDT	S14MAD	S14MAT	S14PHT	S14PHV	S14RAT		S14RHV	S14SVD
	0:01:23	9:45:23	0:14:59	0:14:59		9:36:29	0:14:59	0:14:59	0:14:59	0:14:59	0:14:59	0:14:59		0:14:59	0:15:06
AHU15	Overall		S15CCV		S15DAT	S15E15		S15MAD	S15MAT	S15PHT	S15PHV	S15RAT	S15RH1T/2T	S15R1V/2V	
	0:01:16		0:14:59		0:14:59	12:10:48		0:14:59	0:14:59	0:14:59	0:14:59	0:14:59	0:14:59	0:14:59	
AHU16	Overall		S16CCV					S16MAD	S16MAT	S16PHT	S16PHV		S16RH1T/2T	S16R1V/2V	S16SFVFD:INPUT
	0:01:24		0:14:59					0:14:59	0:14:59	0:14:59	0:14:59		0:14:59	0:14:59	0:14:59
AHU17	Overall	S17	S17CCV					S17MAD	S17MAT		S17PHV	S17RAT			S17SVD
	0:01:23	8:11:45	0:14:59					0:14:59	0:14:59		0:14:59	0:14:59			0:29:30
AHU18	Overall	S18	S18CCV	S18CDT		S18E23	S18HDT	S18MAD	S18MAT	S18PHT	S18PHV	S18RAT		S18RHV	S18SVD
	0:01:14	9:49:26	0:14:59	0:14:59		2:03:20	0:14:59	0:14:59	0:14:59	0:14:59	0:14:59	0:14:59		0:14:59	0:15:03
AHU19	Overall		S19CCV		S19DAT	S19E21		S19MAD	S19MAT	S19PHT	S19PHV	S19RAT	S19RH1T/2T	S19R1V/2V	S19SFVFD
	0:01:16		0:14:59		0:14:59	12:06:08		0:14:59	0:14:59	0:14:59	0:14:59	0:14:59	0:14:59	0:14:59	0:14:59
AHU20	Overall		S20CCV		S20DAT			S20MAD	S20MAT	S20PHT	S20PHV		S20RH1T/2T	S20R1V/2V	S20SFVFD
	0:01:35		0:15:33		0:15:33			0:15:33	0:15:33	0:15:33	0:15:33		0:15:33	0:15:33	0:14:59
CHW	Overall	CWRT	CWST)3capacity												
	0:04:26														
RTU	Overall	AV_7	CMP1	SAT	SF	HS1	RMT								
	0:11:45	0:15:17	2:40:57	0:15:17	6:46:17	1:28:37	0:15:17								

Figure 7 data source frequency

Result: The above are the changes calculated over every variable and their average time is noted here (data is stored only when there is a change of value). The table below shows the average time update for every dataset when the table is full.

Dataset	Average time per record
AHU01	14:59
AHU02	14:59
AHU03	14:53
AHU05	14:53
AHU06	14:59
AHU07	14:59
AHU08	14:59
AHU10	14:59
AHU11	14:59
AHU12	15:03
AHU13	14:59
AHU14	14:59
AHU15	14:59
AHU16	14:59
AHU17	14:59
AHU18	14:59
AHU19	14:59
AHU20	15:33
CHW	14:59

RTU	15:17

Table 3 Average Time Per Record

From Figure 1, Most numeric data capture each 15mins
Unit On/Off data mostly capture twice each day, at beginning and end of each day: 0:00:00 and 23:59:59, average capture frequency each 9hrs~12hrs.

2. Data Integration

2.1. Design (simple) SQL database schema to integrate the data on a prototype basis
2.2. Load the data into an integrated database
We did not have a chance to connect with any of the databases to work out on problems of the same.

3. Analytics

Perform exploratory data analysis (summary statistics and visualizations)

3.1 Energy Consumption:

To define “distance” between buildings based on consumption condition. Smaller “distance” means more similar conditions. Finally, with help of R, we plot Cluster Dendrogram to show the similarity of energy consumption conditions. The variables used for the clustering are the consumption of chilled water, hot water, electricity and gas. IIT Research Tower stands out with the highest consumption in all kinds of resources. The second group has lower consumption, and we can find out that they are research or technique related buildings. The last group contains the buildings will less consumption or lack of data. To improve this algorithm, we may need to consider the sizes of the building. Using consumption per square can be better. During the analysis, we only scale the data instead of standard with area.

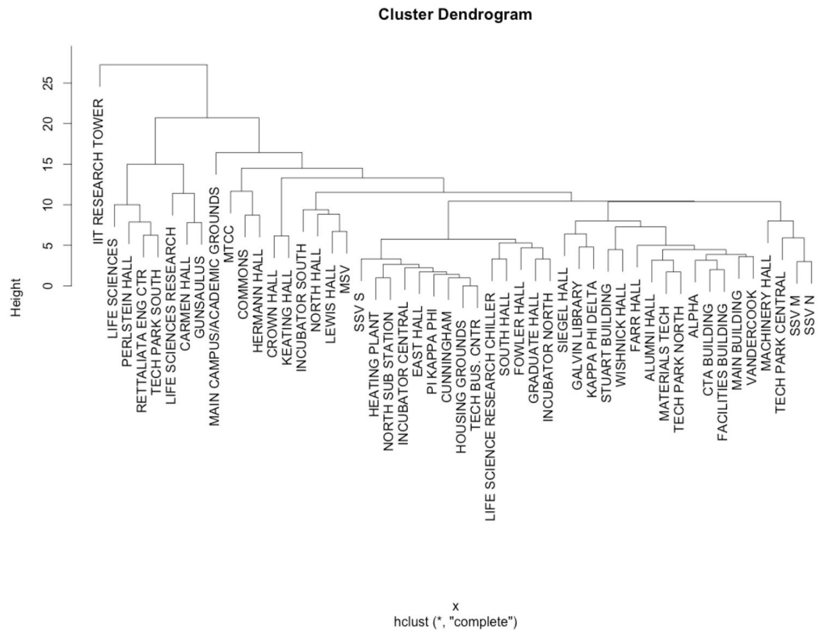


Figure 8 Cluster Dendrogram of Energy Consumption

3.2 AHU:

We designed a method to detect outliers of 18 AHUs using in IIT Research Tower. Each AHU data can be seen as a matrix. After shaping the matrices into same size, we calculate the eigenvectors of the matrices to stand for the feature of the AHU. The last step is to apply Clustering algorithms to the eigenvectors to figure out the abnormal AHU. The method fails because of the lack of the data.

The boxplots show the outliers (in black) in each of the variables. The reports show three types of representation namely: all the values, sup ON (when the supply is ON), sup OFF (when the supply is OFF).

The following are boxplots and and a bivariate relation between every numerical variable in the dataset.

Boxplots -

Q3 + whis*IQR

where

IQR = interquartile range

Whis = Q3-Q1

Q3 - 75th percentile

Q1 - 25th percentile

Objective: To find the Outliers in each variable across all datasets.

AHU 01:
Temperature:

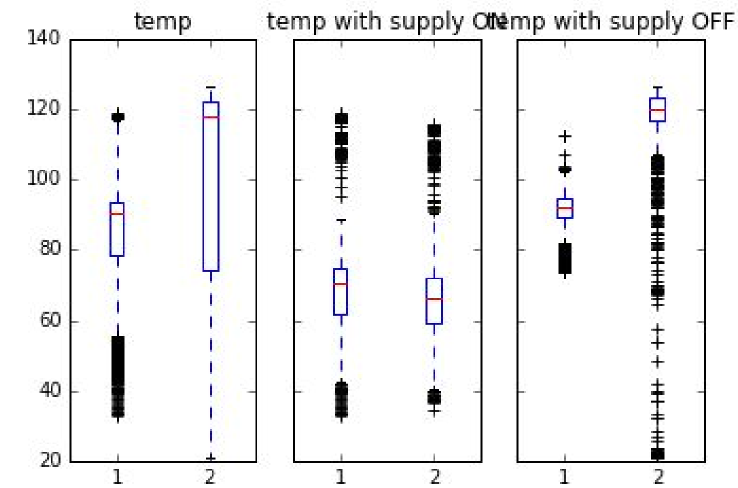


Figure 9 AHU 01 Temperature Box Plot

1. Discharger temperature
2. Heating coil discharge temperature

Result: The outliers in the Heating coil discharge temperature is a lot more than the discharge temperature when the supply is OFF but when the supply is ON, there are outliers at the extremes out of each of the temperature values.

.

Values:

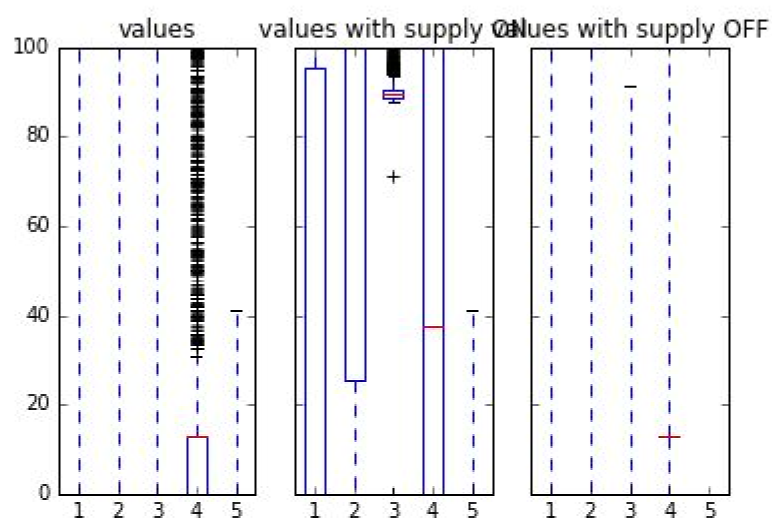


Figure 10 AHU 01 Values Box Plot

1. Cooling coil %open
2. Phase bypass damper

3. Supply fan
4. Heating coil valve
5. Humidity valve

Result: The Heating coil value has outliers all over. The Supply fan value has outliers to the higher than the values at 90. The values of cooling coil value, Phase bypass damper and Heating coil valve are varied over the percentages .

Bivariate:

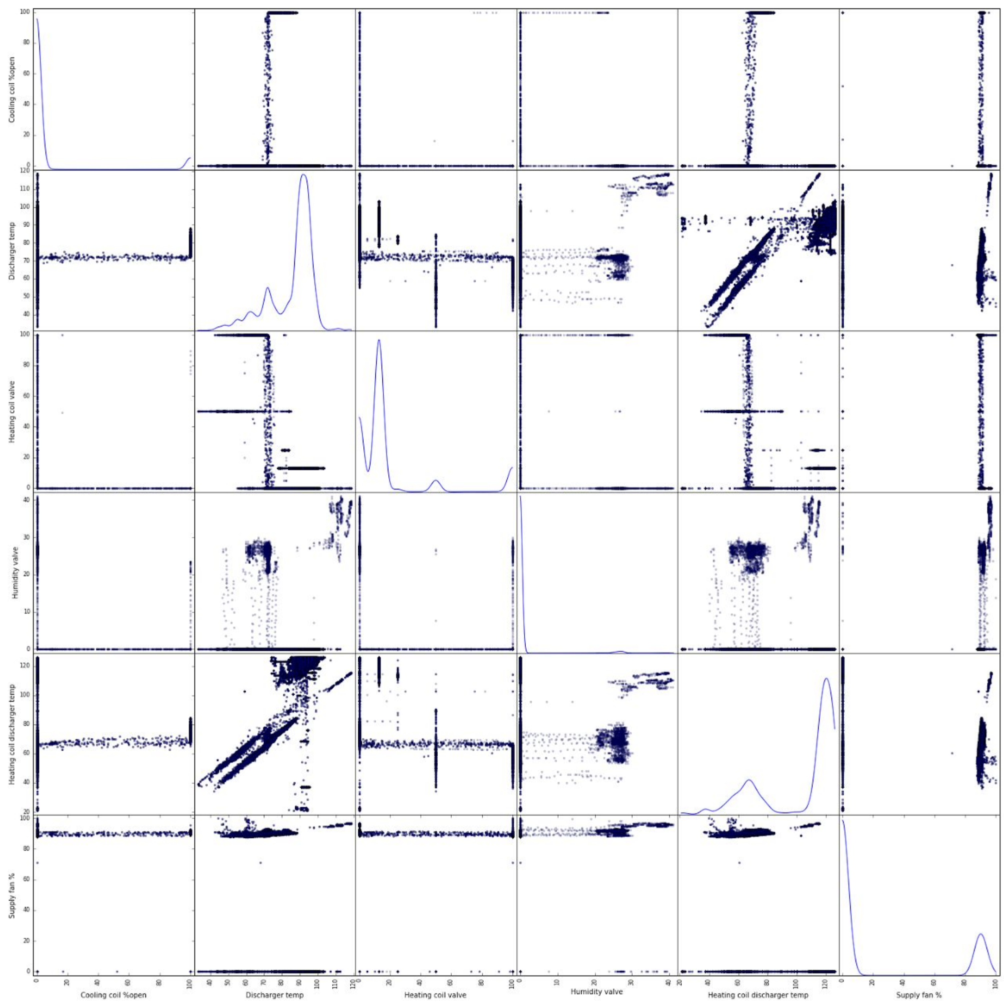


Figure 11 AHU 01 Scatter Plot

AHU 02:

Temperature:

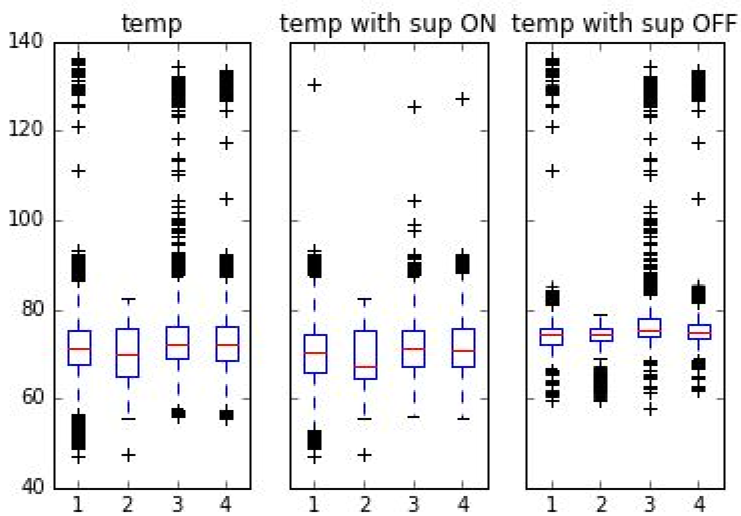


Figure 12 AHU 02 Temperature Box Plot

1. Discharge temp
2. Mixed air temp
3. re-heat coil 1 Discharger temp,
4. re-heat coil 2 Discharger temp

Result: The Discharge temperature has outliers way higher than the temperatures of 120 deg. Mixed Air temperature have below 70 deg. The other temperature values have outliers varied over a period.

Values:

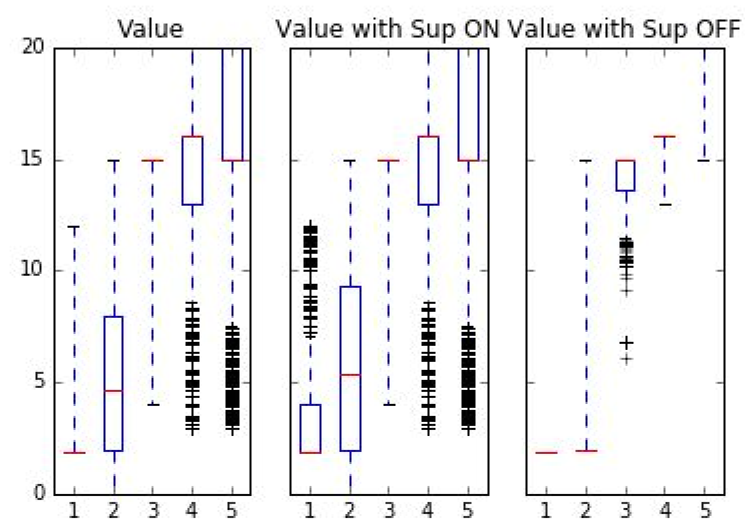


Figure 13 AHU 02 Value Box Plot

1. Cooling coil valve %
2. Mixed air damper %
3. Pre-heat valve %
4. Re-heat coil valve 1 %
5. Re-heat coil valve 2 %

Result: The outliers lie at the Re-heat valves. There is more usage outliers. The cooling coil valve has more outliers when the supply is ON.

Bivariate

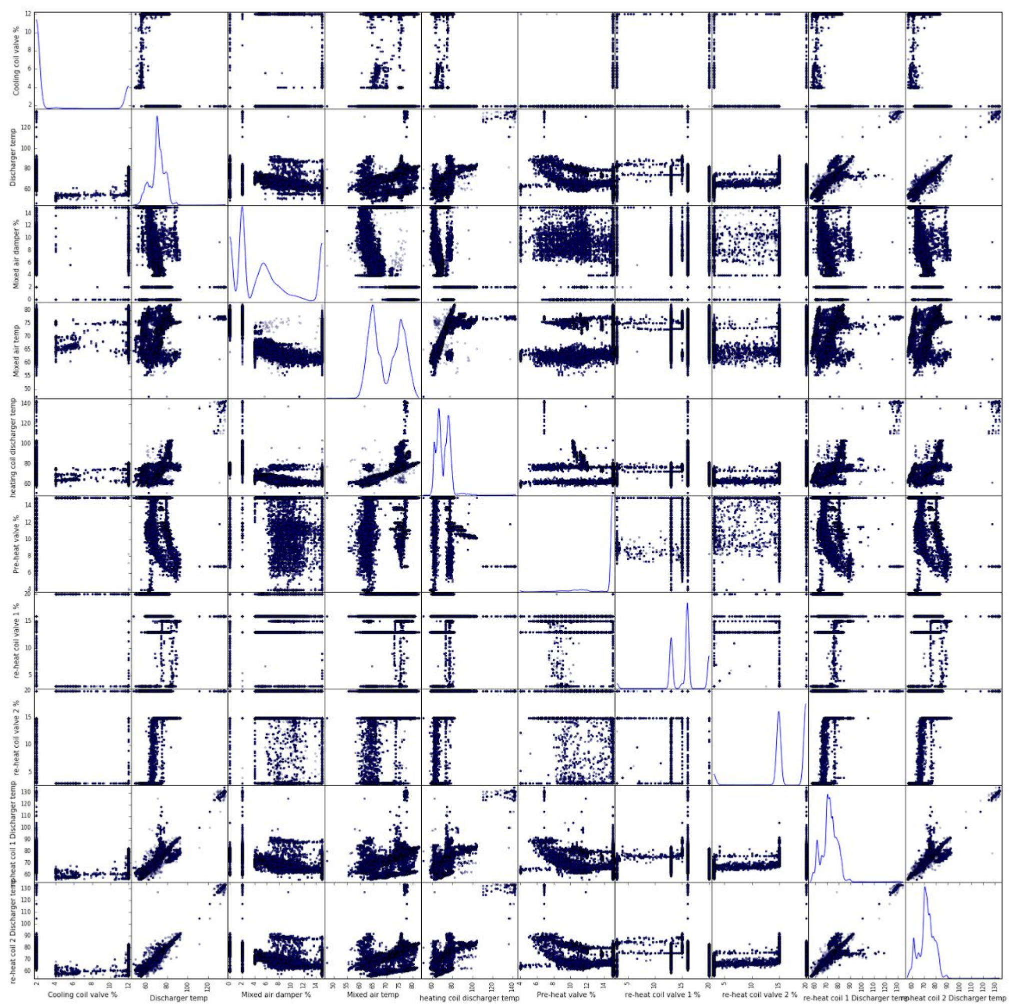


Figure 14 AHU 02 Scatter Plot

AHU 03:

Temperature:

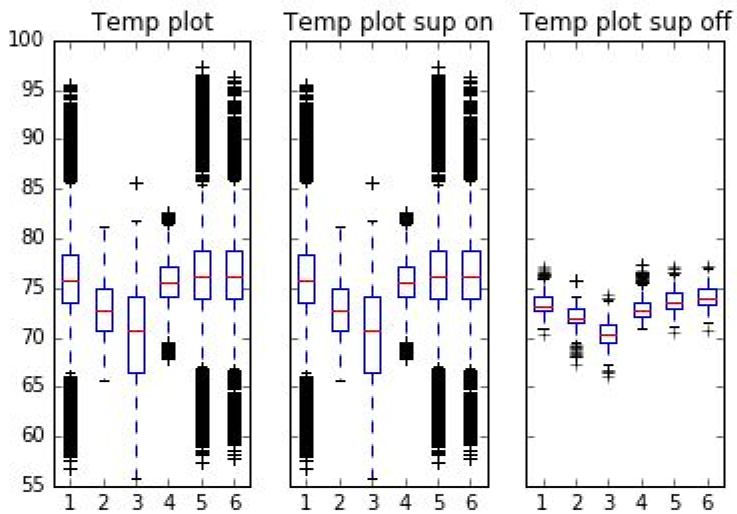


Figure 15 AHU 03 Temperature Box Plot

1. Discharge temperature
2. Mixed air temperature
3. Pre-heat coil discharge temperature
4. Return air temperature
5. re-heat coil 1 Discharge temperature
6. re-heat coil 2 Discharge temperature

Result: The temperature when the supply is ON is similar to the outliers over the whole dataset. There are lesser outliers when the supply fan is off.

Values:

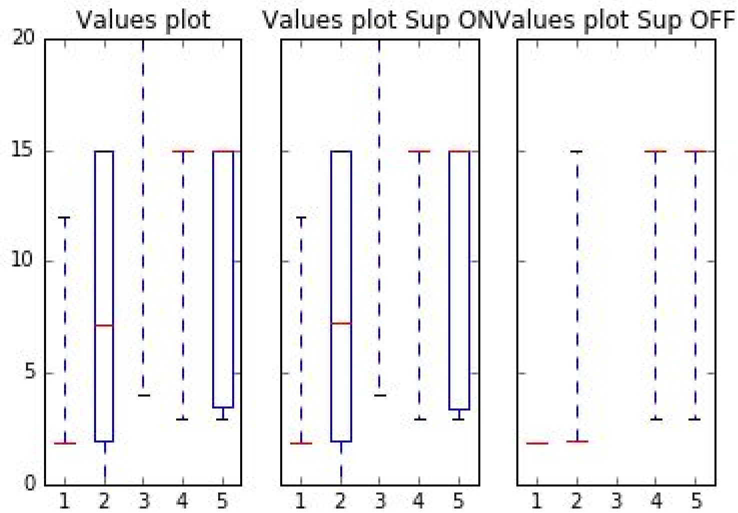


Figure 16 AHU 03 Values Box Plot

1. Cooling coil valve %
2. Mixed air damper %
3. Pre-heat valve %
4. re-heat coil 1 Discharger valve
5. re-heat coil 2 Discharger valve

Result: the value plot shows that the data is spread over all the values.

Bivariate:

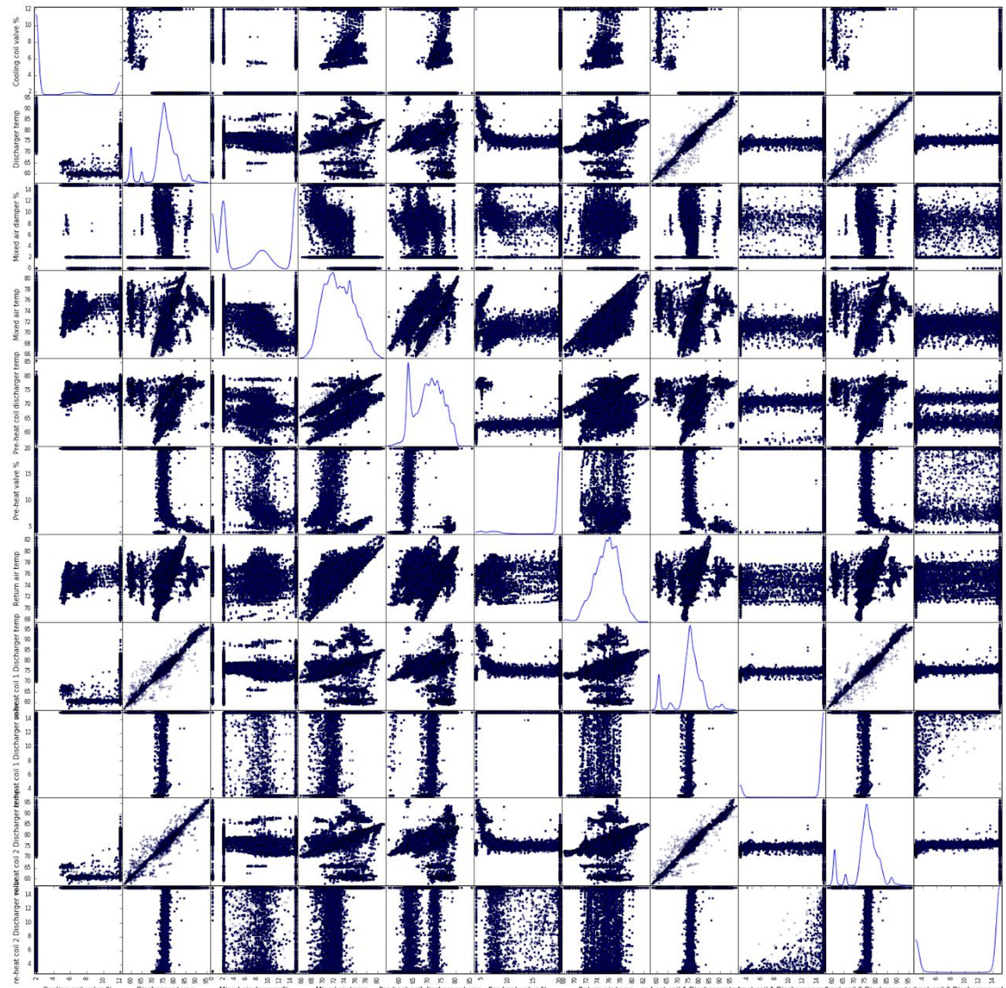


Figure 17 AHU 03 Scatter Plot

AHU 05:
Temperature:

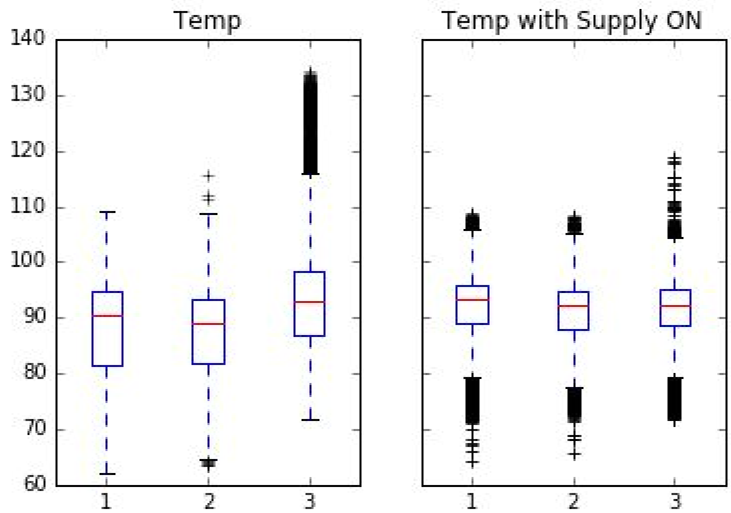


Figure 18 AHU 05 Temperature Box Plot

1. cooling coil discharge temperature
2. Discharge temperature
3. Pre-heat cold discharge temperature

Result: the outliers of are below the temperature of 80 deg

Values:

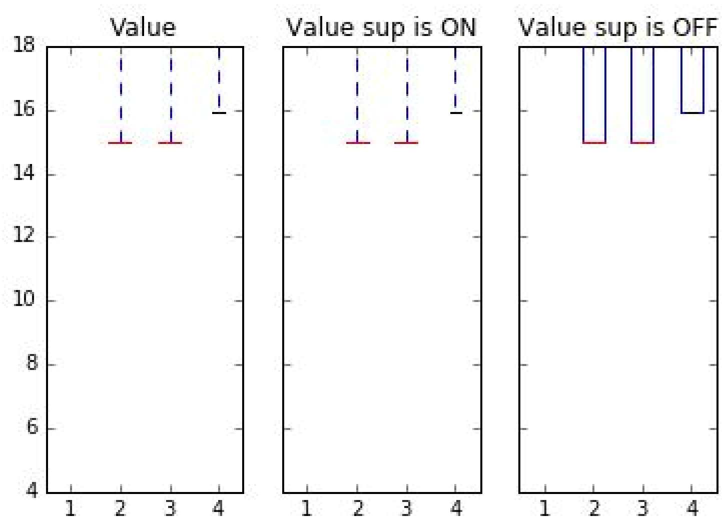


Figure 19 AHU 05 Value Box Plot

1. Cooling coil valve %

2. Pre-heat coil 1 Discharger valve
3. Pre-heat coil 2 Discharger valve
4. Re-heat coil Discharger value

Result: No conclusive result through this method.

Bivariate:

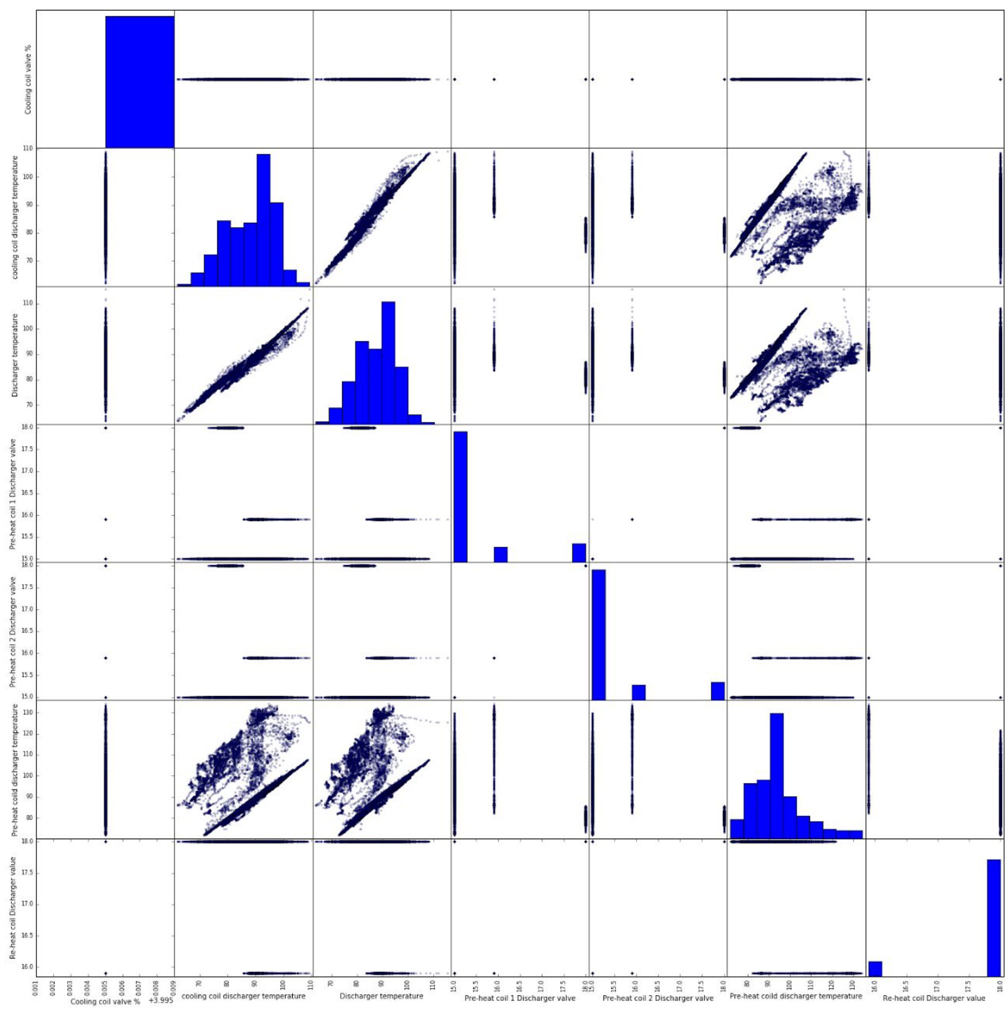


Figure 20 AHU 05 Scatter Plot

AHU 06:

Temperature:

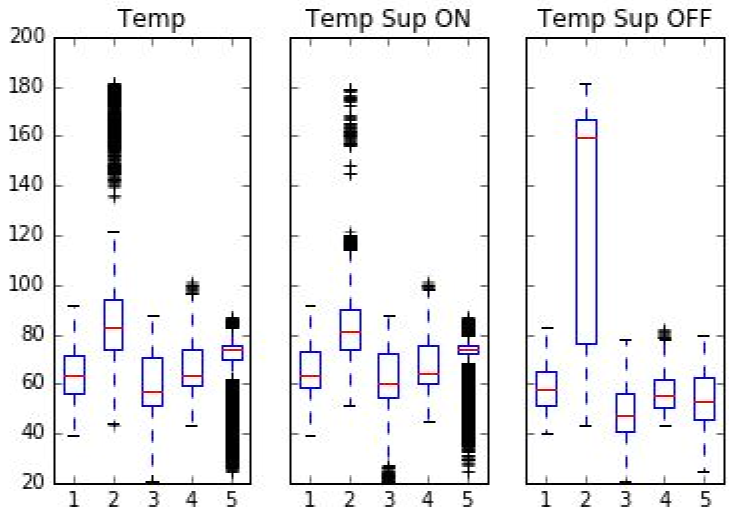


Figure 21 AHU 06 Temperature Box Plot

1. Cold Deck temperature
2. Hot Deck temperature
3. Mixed air temperature
4. Pre-heat coil discharge temperature
5. Return air temperature

Values:

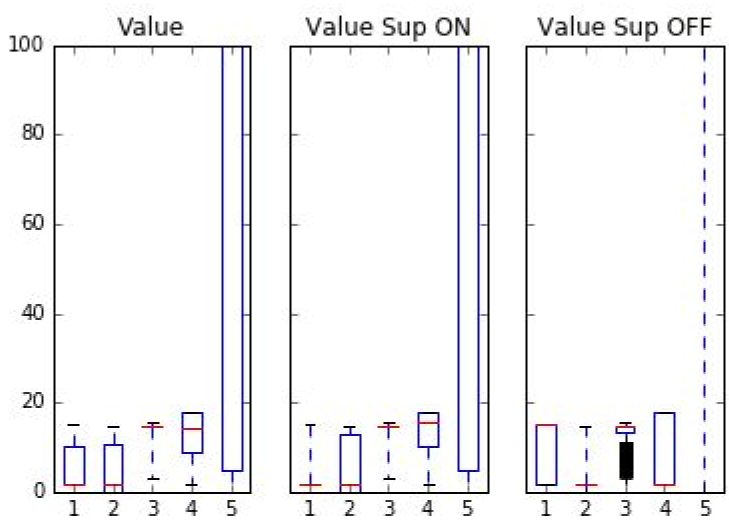


Figure 22 AHU 06 Value Box Plot

1. Cooling coil valve %
2. Mixed air damper %
3. Pre-heat coil Discharge valve
4. Re-heat coil Discharge value
5. Supply fan VFD speed %

Bivariate:

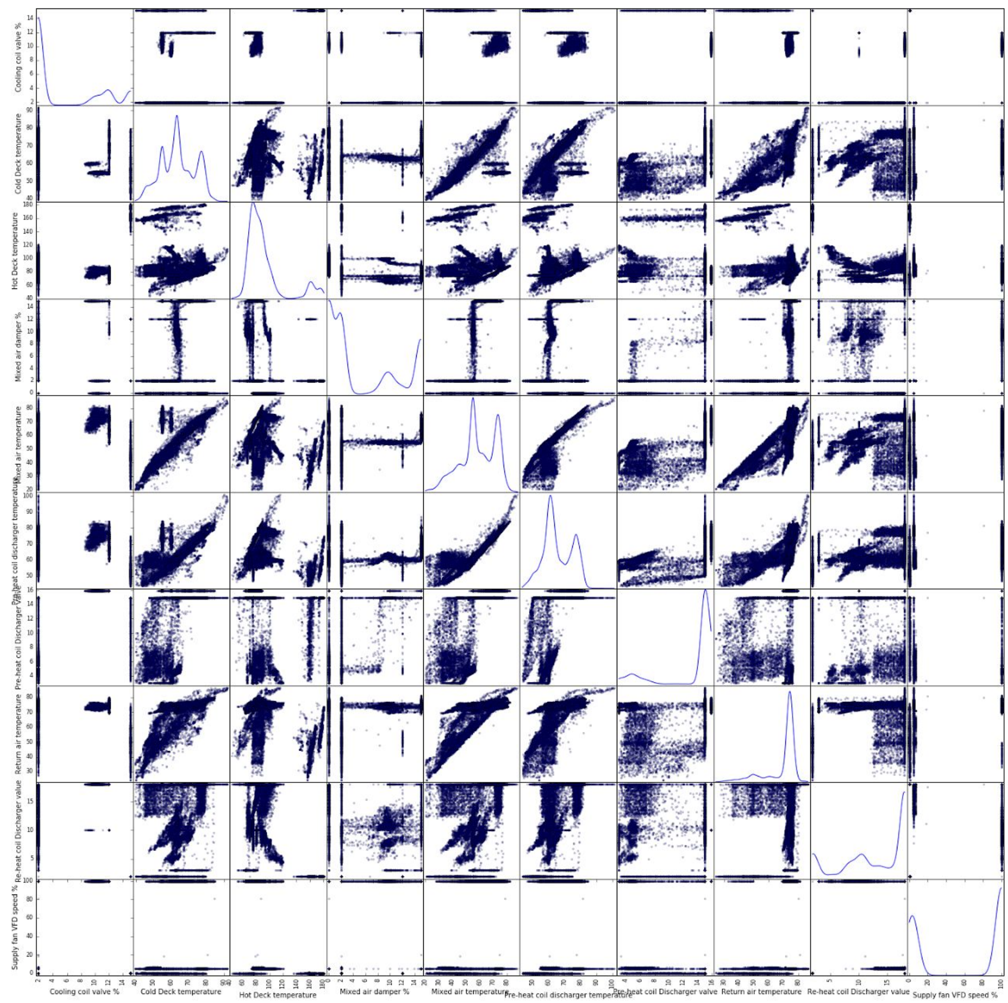


Figure 23 AHU 06 Scatter Plot

AHU 07:
Temperature:

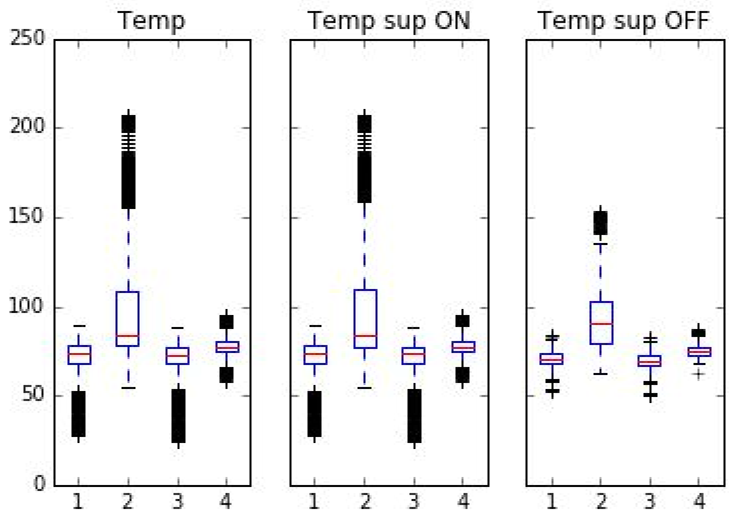


Figure 24 AHU 07 Temperature Box Plot

1. Cold Deck temperature
2. Hot Deck temperature
3. Mixed air temperature
4. Return air temperature

Result: There is an overlap of the outliers in the dataset when the supply fan is ON with that of the overall dataset.

Values:

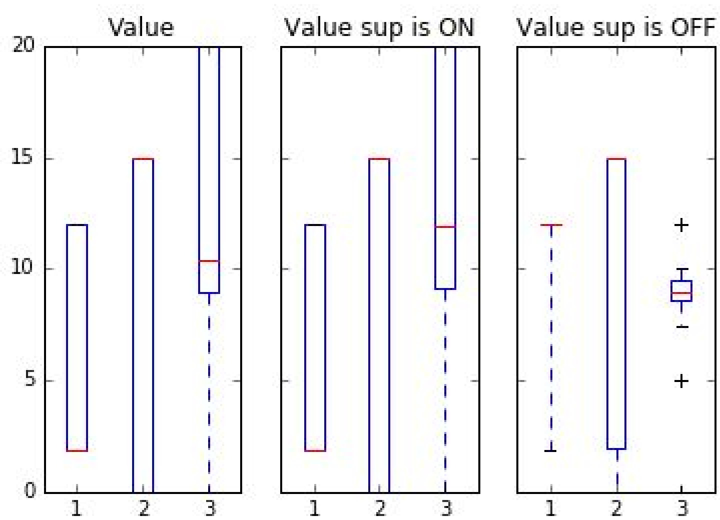


Figure 25 AHU 07 Value Box Plot

1. Cooling coil valve %open
2. Mixed air damper % open
3. Re-heat coil Discharger value

Result: No conclusive results on the dataset.

Bivariate:

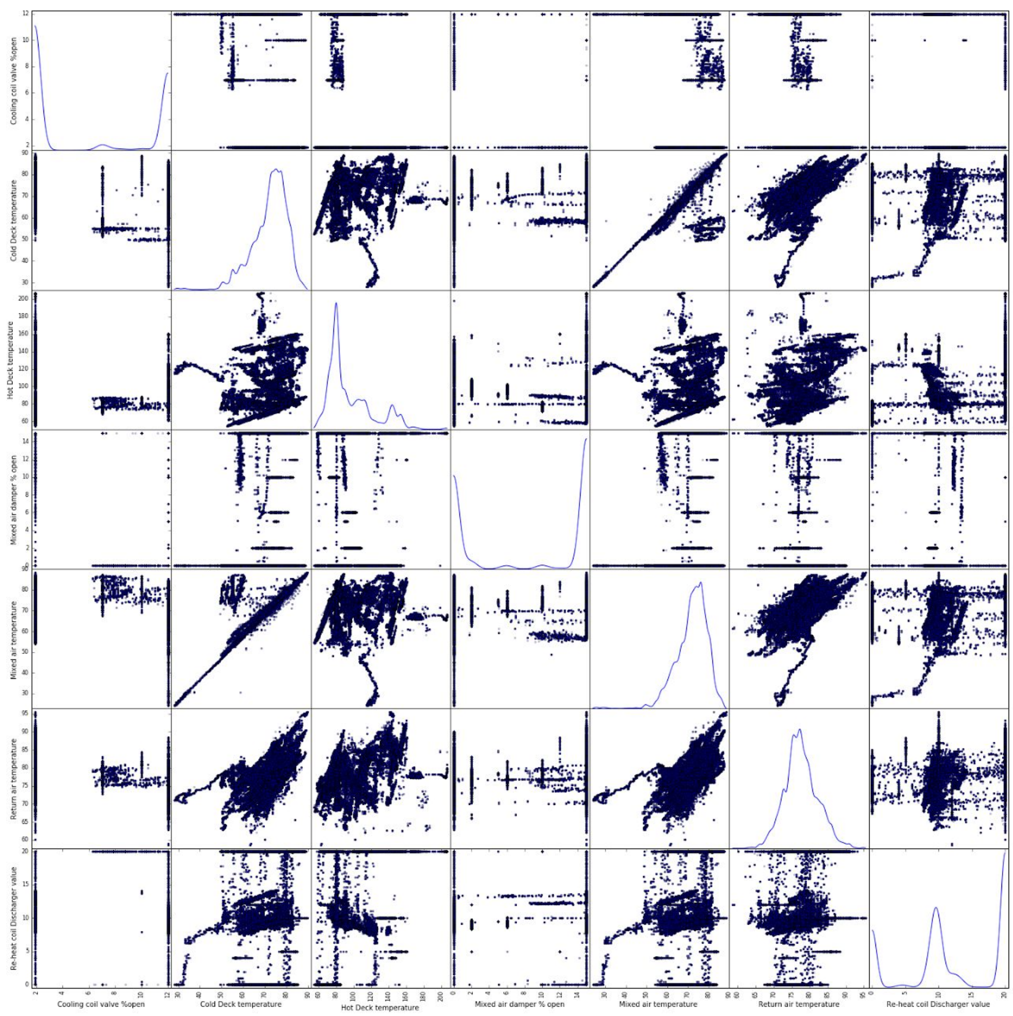


Figure 26 AHU 07 Scatter Plot

AHU 08:

Temperature:

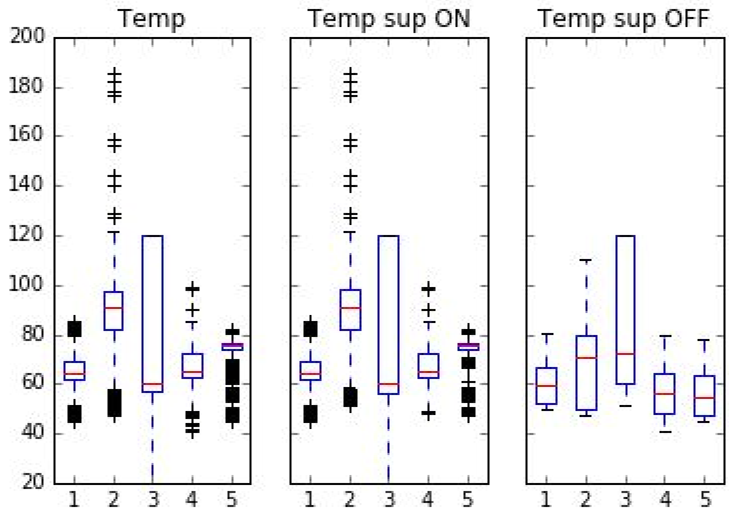


Figure 27 AHU 08 Temperature Box Plot

1. Cold Deck temperature
2. Hot Deck temperature
3. Mixed air temperature
4. Pre-heat coil discharge temperature
5. Return air temperature

Result: the dataset shows that when the supply is OFF, there are almost no outliers but there is an overlap of when the supply is ON over the overall plot.

Values:

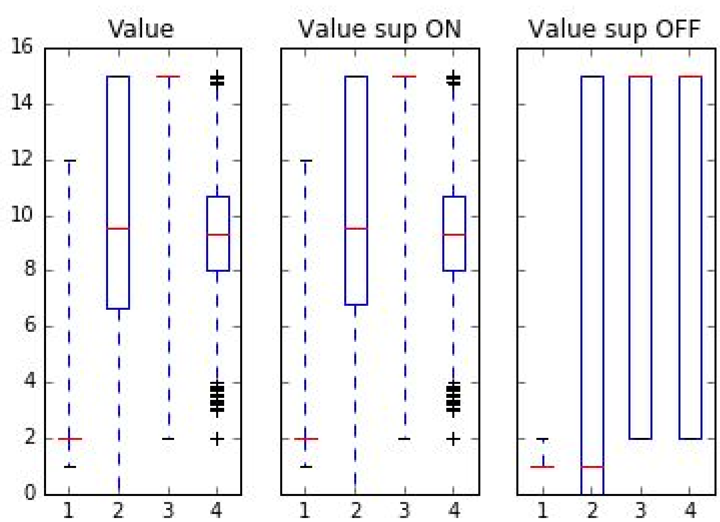


Figure 28 AHU 08 Value Box Plot

1. Cooling coil valve %open
2. Mixed air damper % open
3. Pre-heat coil Discharge value
4. Re-heat coil Discharge value

Result: The result here is inconclusive.
Bivariate:

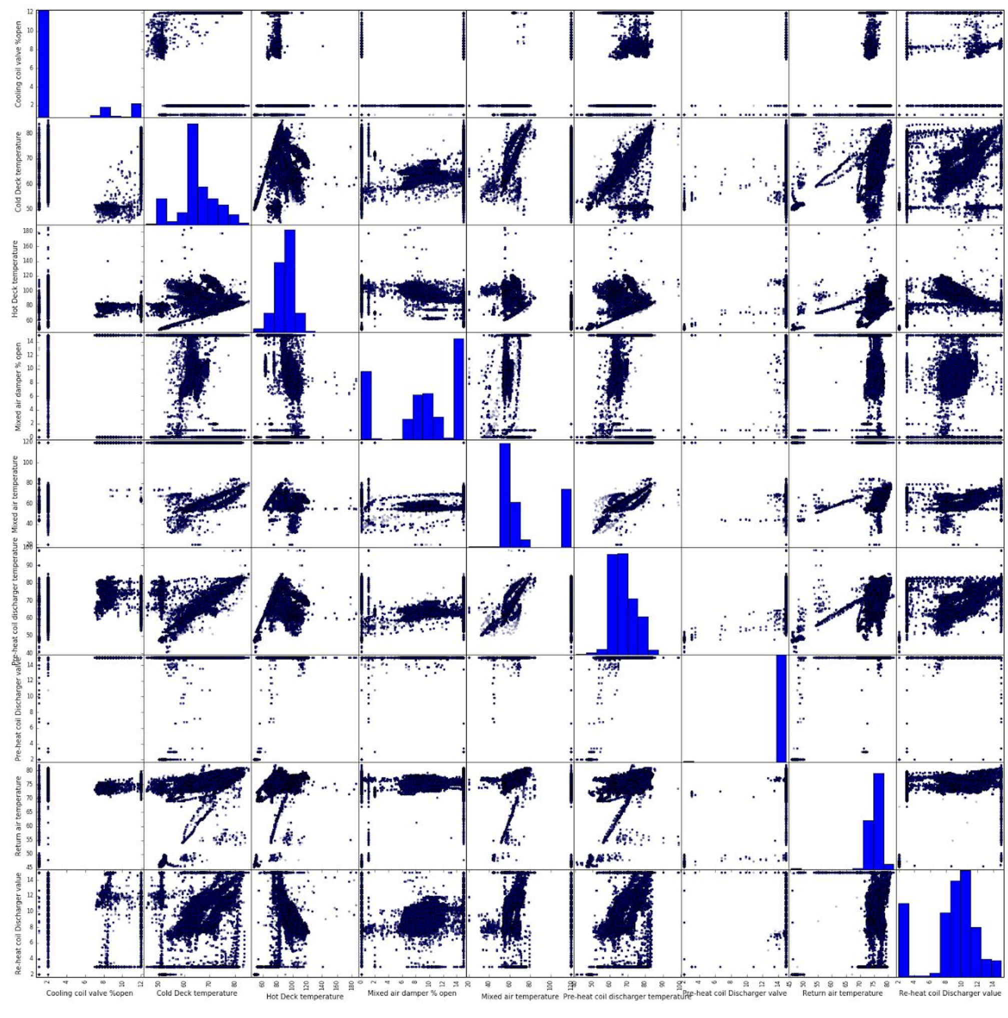


Figure 29 AHU8 Scatter Plot

AHU 10:

Temperature:

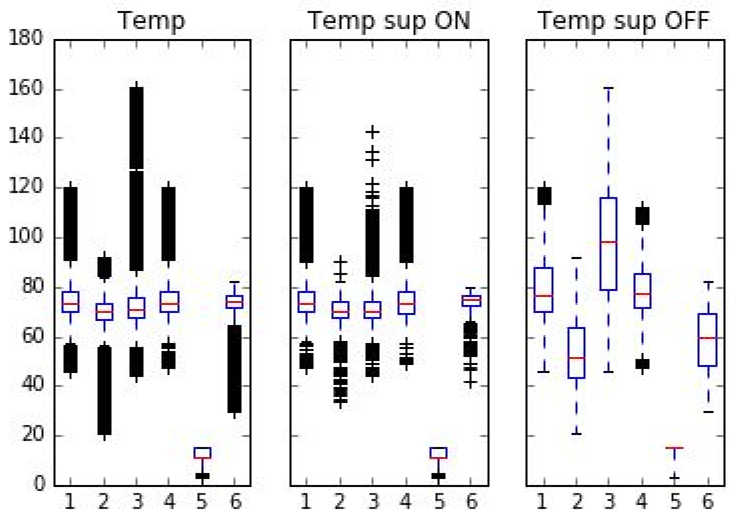


Figure 30 AHU 10 Temperature Box Plot

1. Discharger temperature
2. Mixed air temperature
3. Pre-heat coil discharge temperature
4. Re-heat coil 1 discharger temperature
5. 'Re-heat coil 2 Discharger value
6. Return air temperature

Result: The overall plot shows more outliers cause of the balance of the datasets. Most of the outliers are in the area of the 90 deg and above.

Values:

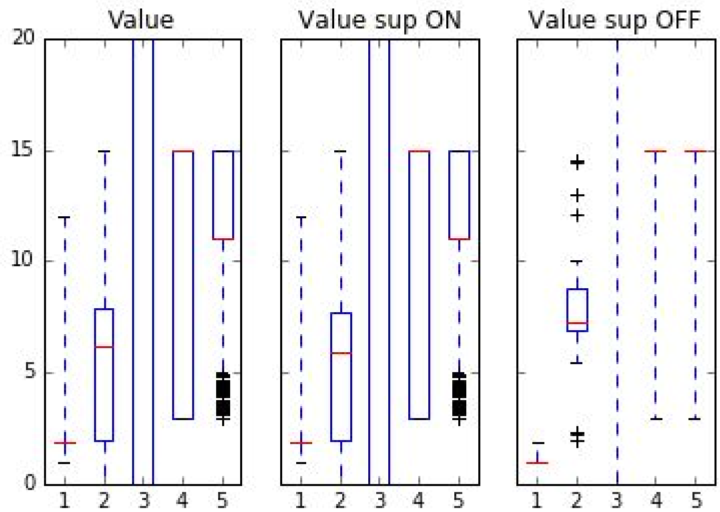


Figure 31 AHU 10 Value Box Plot

1. Cooling coil valve %open
2. Mixed air damper % open
3. Pre-heat coil Discharger value
4. Re-heat coil 1 Discharger value
5. Re-heat coil 2 Discharger value

Bivariate:

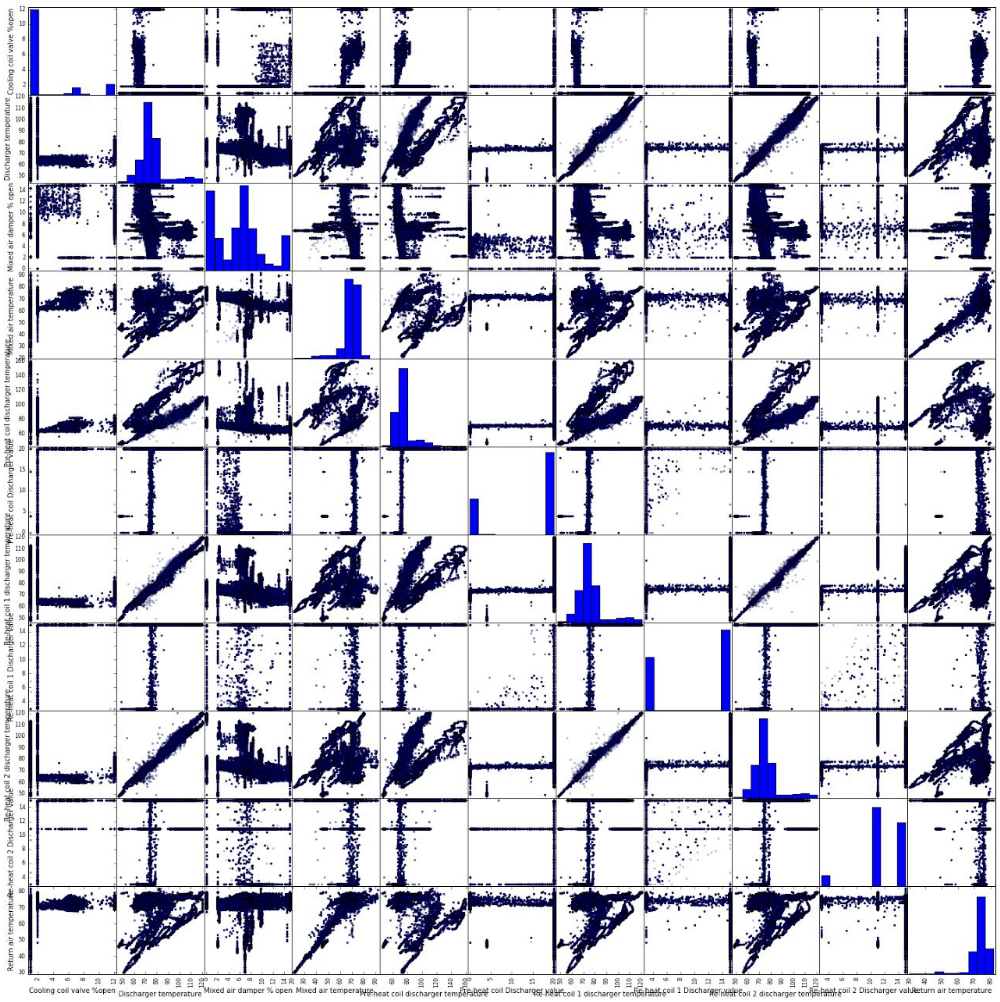


Figure 32 AHU 10 Scatter Plot

AHU 11:

Temperature:

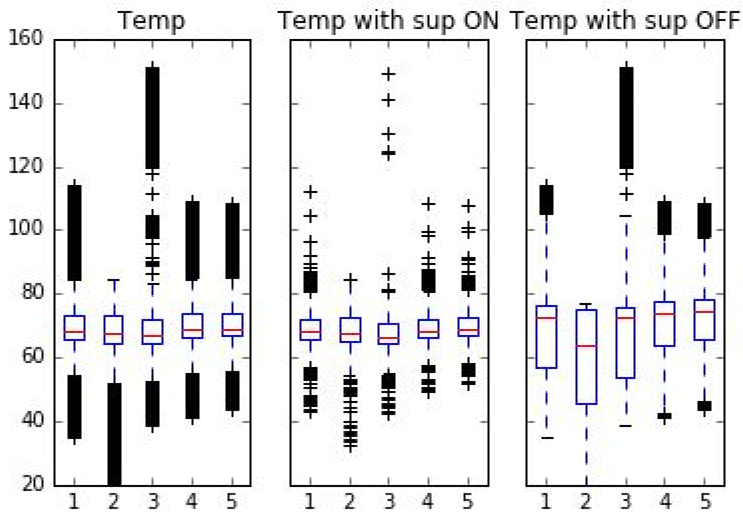


Figure 33 AHU 11 Temperature Box Plot

1. Discharger temperature
2. Mixed air temperature
3. Pre-heat coil discharge temperature
4. Re-heat coil 1 discharge temperature
5. Re-heat coil 2 discharge temperature

Result: The outliers are more when the supply fan is OFF. There are lesser outliers when the supply fan is ON. this could be because there is not much usage of this particular AHU.

Values:

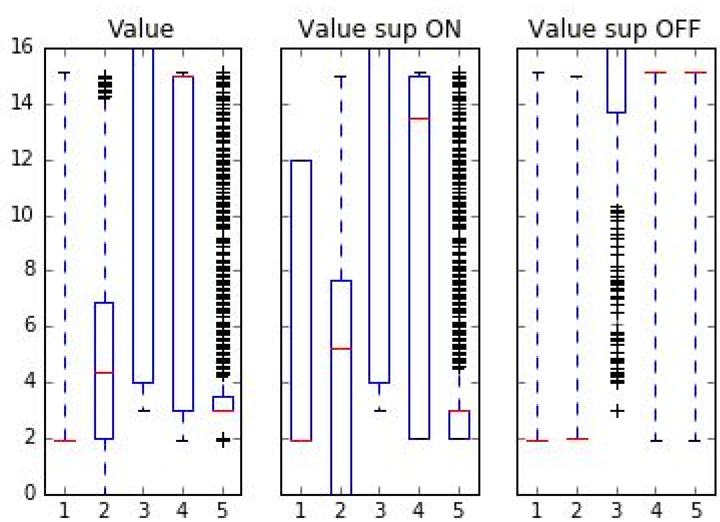


Figure 34 AHU 11 Value Box Plot

1. Cooling coil valve %open
2. Mixed air damper % open
3. Pre-heat coil Discharge valve
4. Re-heat coil 1 Discharge value
5. Re-heat coil 2 Discharge value

Bivariate:

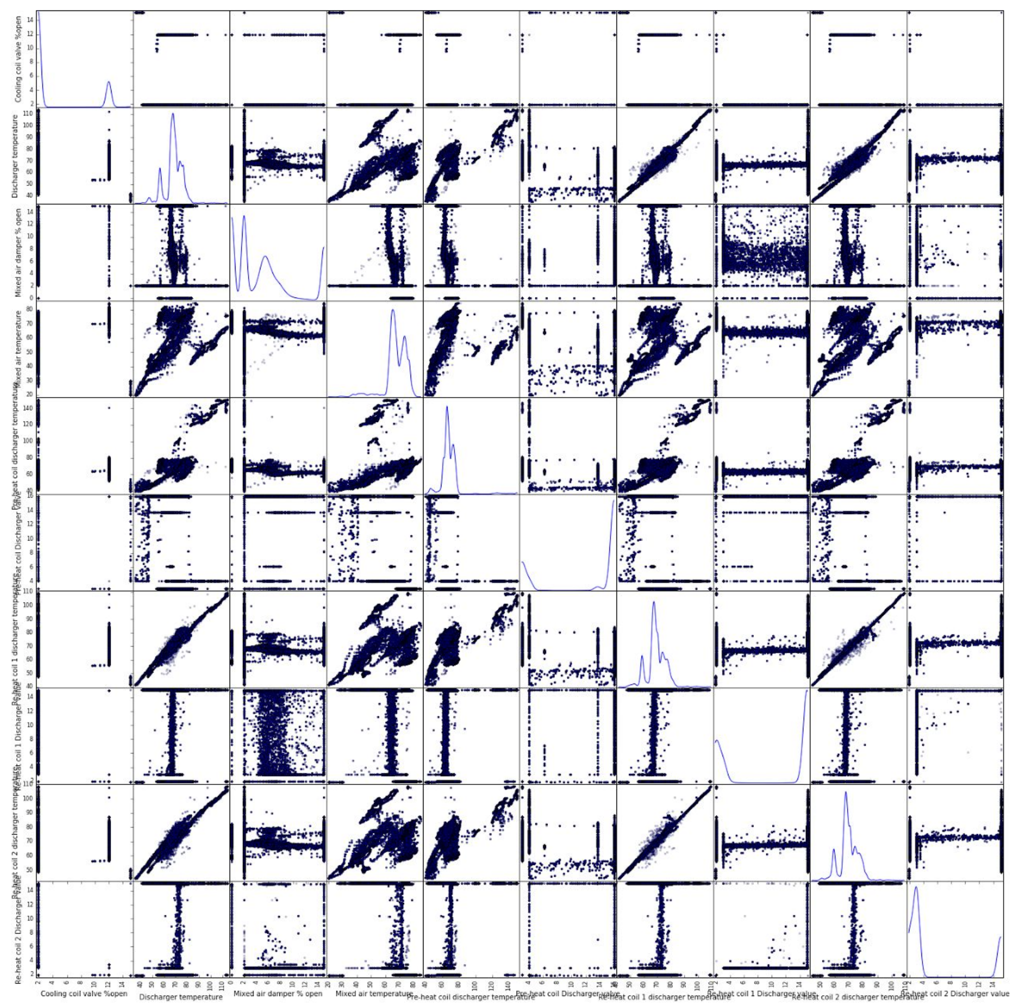


Figure 35 AHU 11 Scatter Plot

AHU 12:

Temperature:

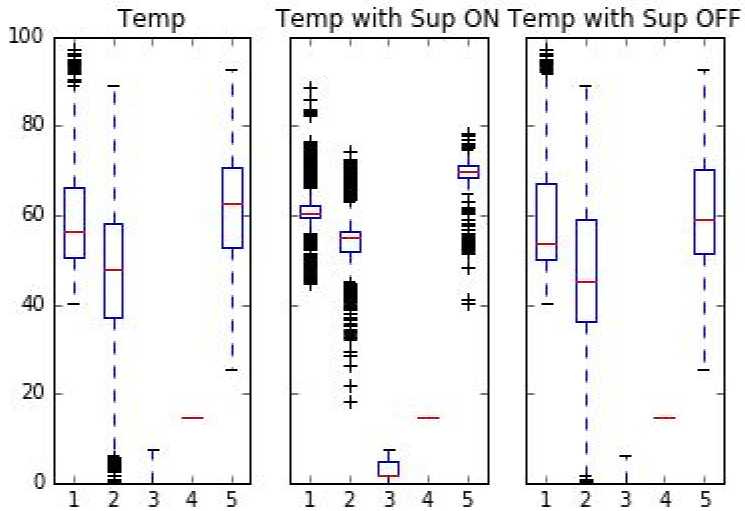


Figure 36 AHU 12 Temperature Box Plot

1. Discharger temperature
2. Mixed air temperature
3. Humidity value
4. Pre-heat value minimum
5. Return air temperature

Result: the outliers are a lot more when the supply fan is ON.over when the supply fan is OFF. There is inconclusive results from the Humidity value and the Pre-heat value Minimum (further analysis shows that this is a constant value.)

Values:

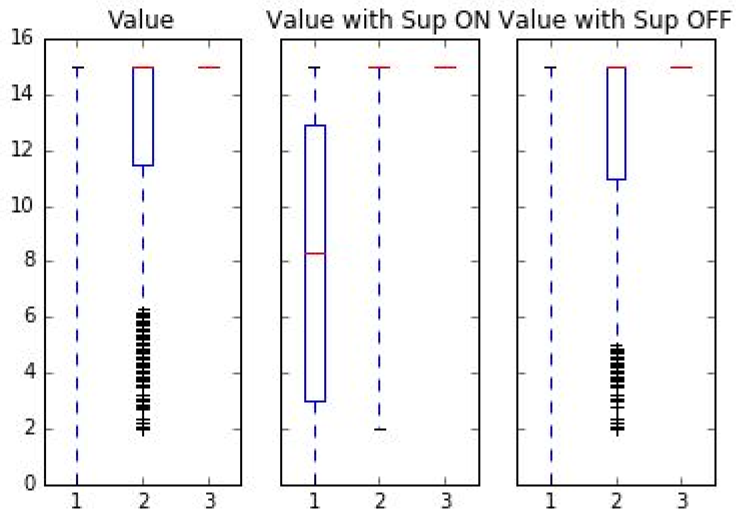


Figure 37 AHU 12 Value Box Plot

1. Mixed air damper % open
2. Pre-heat valve % open
3. Cooling coil valve %open

Result: the Cooling coil value shows an inconclusive value(shows a constant value - its value is not changed over the whole of the time-period). The other values are inconclusive.

Bivariate:

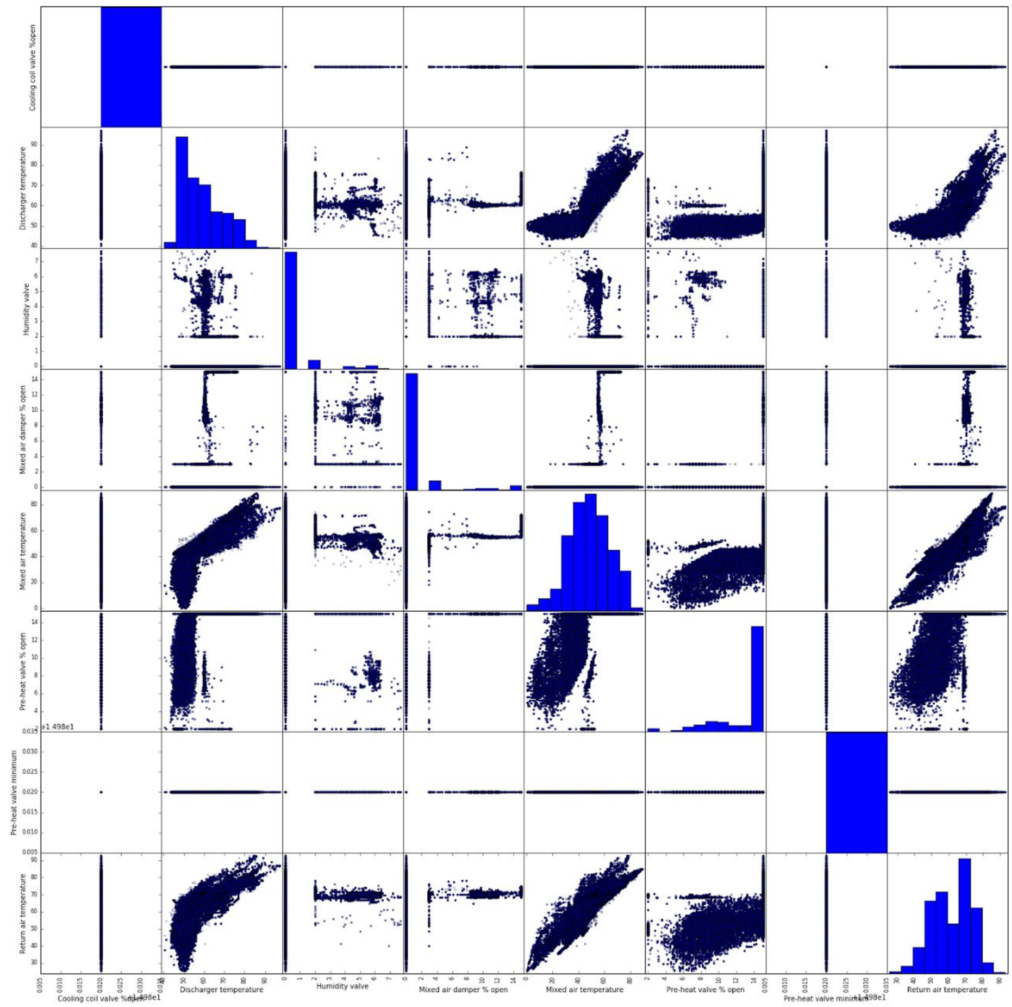


Figure 38 AHU 12 Scatter Plot

AHU 13:

Temperature:

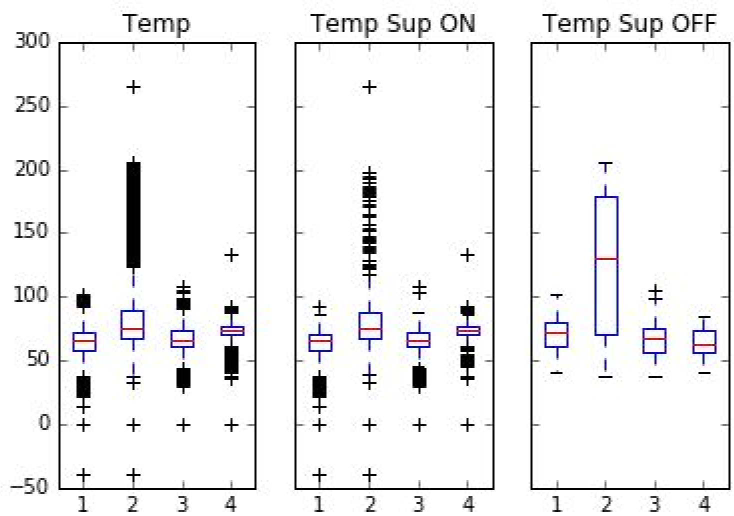


Figure 39 AHU 13 Temperature Box Plot

1. Cold Deck temperature
2. Hot Deck temperature
3. Mixed air temperature
4. Return air temperature

Result: this shows that there are almost no outliers when the supply fan is OFF but there is an overlap in the first two graphs. Hence shows that this AHU runs for long periods of time.

Values:

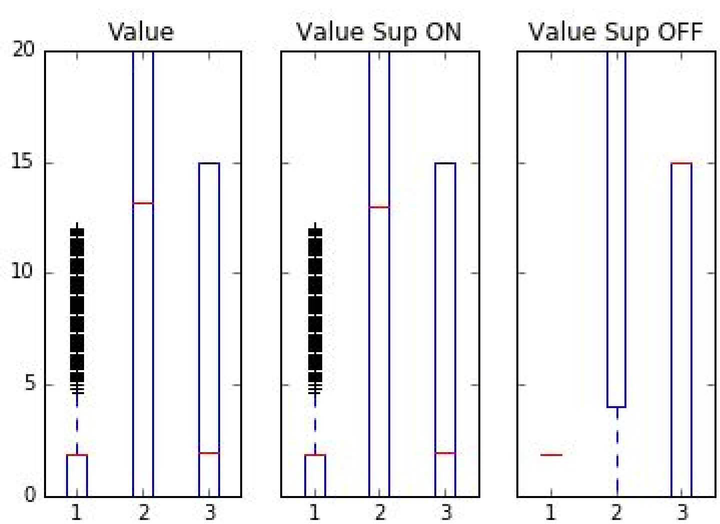


Figure 40 AHU 13 Value Box Plot

1. Cooling coil valve %open
2. Hot Deck Valve
3. Mixed air damper % open

Result: The graphs shows inconclusive results on the dataset.

Bivariate:

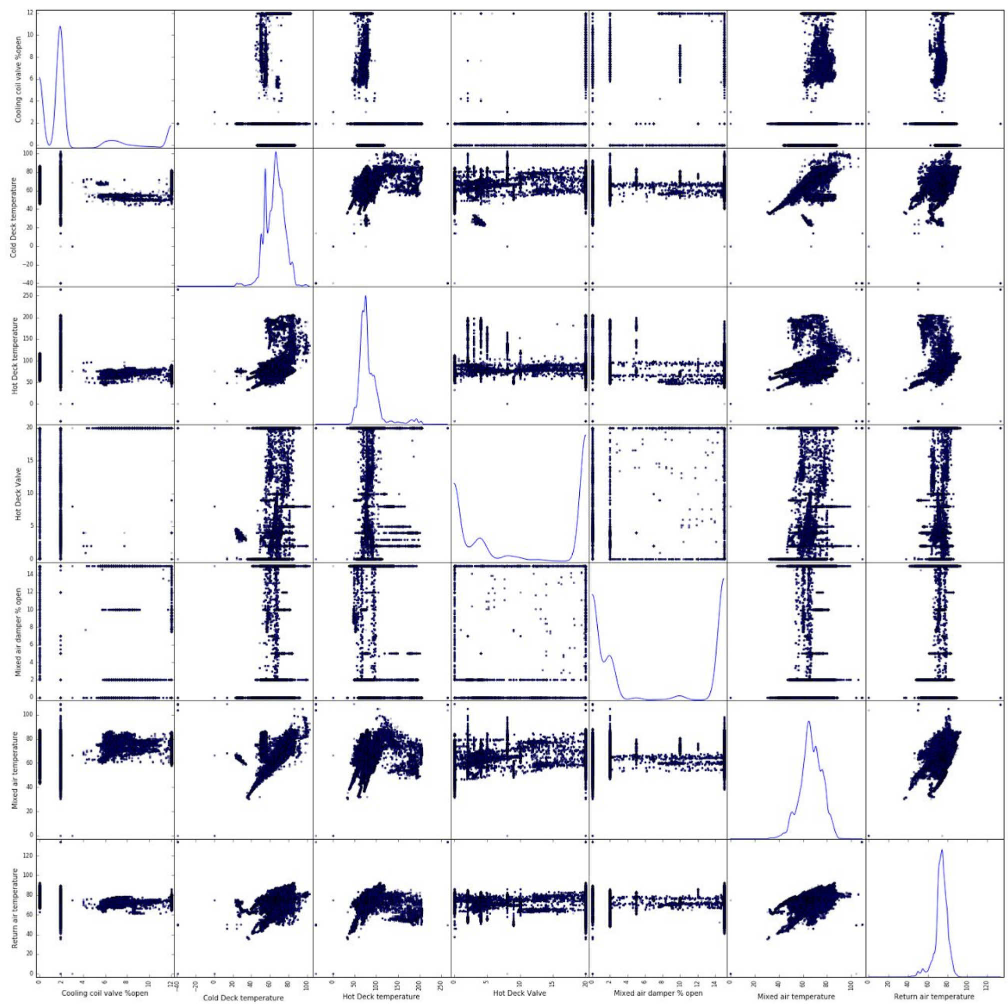


Figure 41 AHU 13 Scatter Plot

AHU 14:

Temperature:

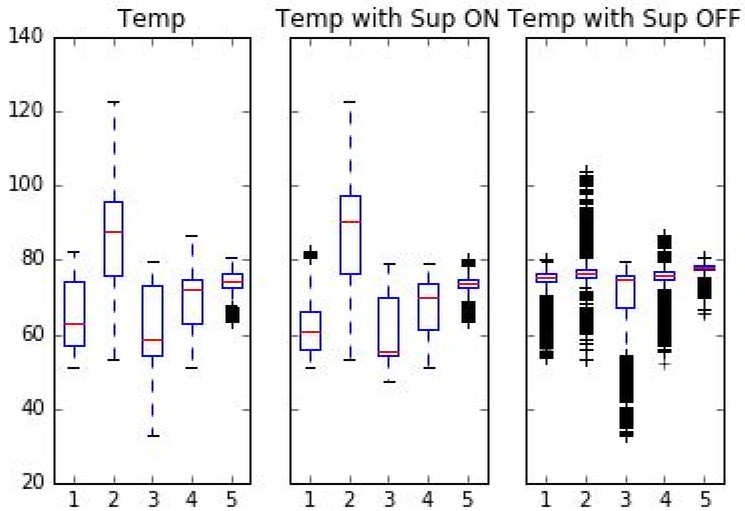


Figure 42 AHU 14 Temperature Box Plot

1. Cold Deck temperature
2. Hot Deck temperature
3. Mixed air temperature
4. Pre-heat coil discharge temperature
5. 'Return air temperature

Result: there are a lot of outliers when the supply fan is OFF. This could mean that the AHU is switch off for most of the times.

Values:

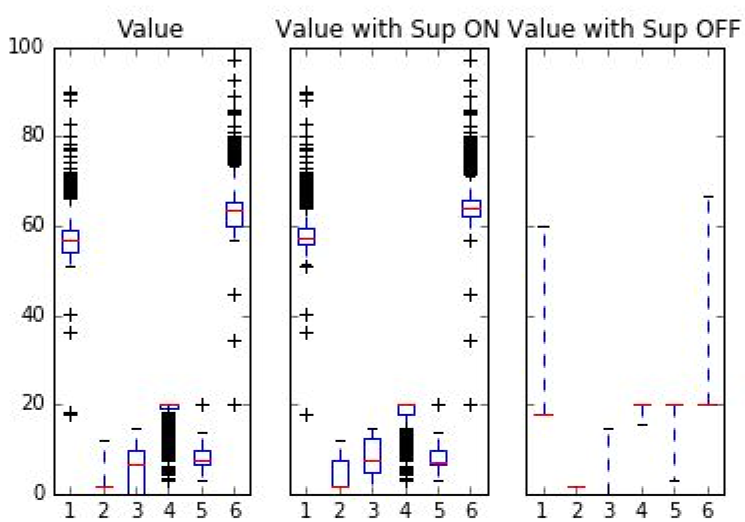
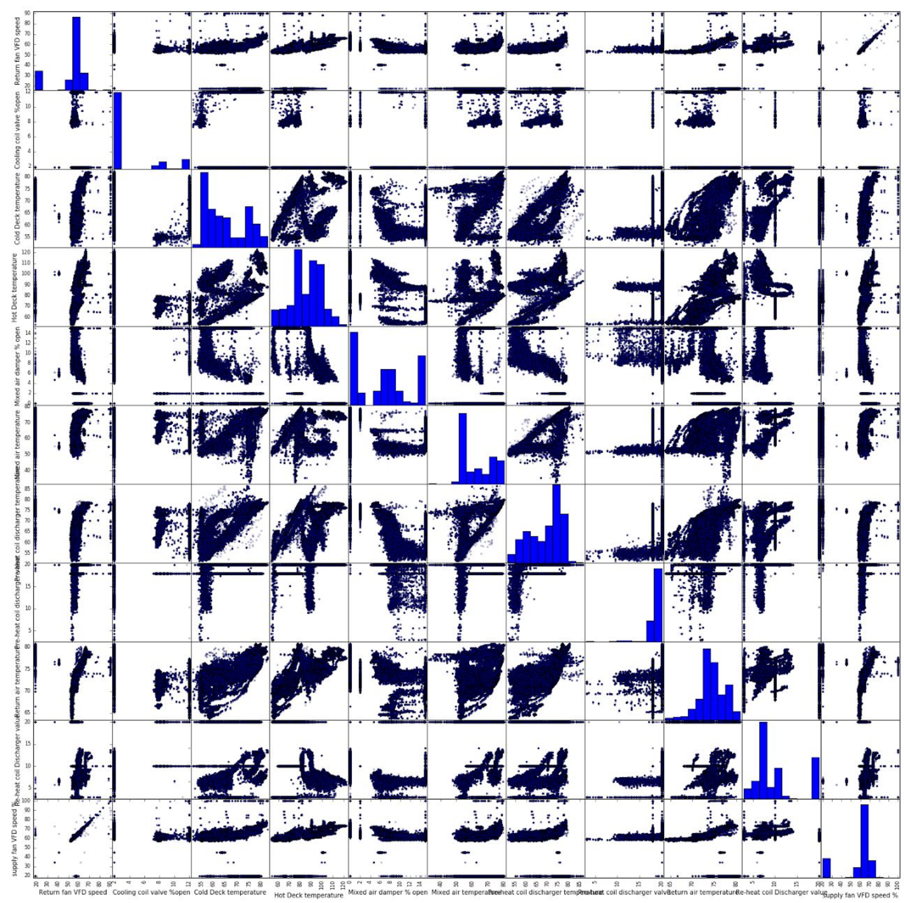


Figure 43 AHU 14 Value Box Plot

1. Return fan VFD speed
2. Cooling coil valve %open
3. Mixed air damper % open
4. Pre-heat coil discharge valve
5. Re-heat coil Discharge value
6. supply fan VFD speed %

Result: the outliers Return fan VFD speed, Pre-heat coil discharge valve and supply fan VFD speed % are having the outliers above, below and above the average data. This AHU runs for most of the time.

Bivariate



:

Figure 44 AHU 14 Scatter Plot

AHU 15:

Temperature:

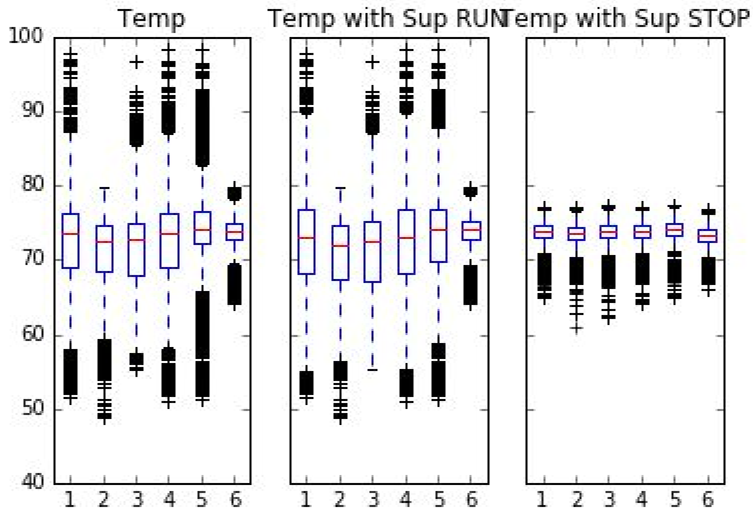


Figure 45 AHU 15 Temperature Box Plot

1. Discharger temperature
2. Mixed air temperature
3. Pre-heat coil discharge temperature
4. re-heat coil 1 Discharge temperature
5. re-heat coil 2 Discharge temperature
6. Return air temperature

Result: there is equal usage of this AHU. The values and the outliers are equally split over the dataset. This shows a lot of outliers at the extremes of the dataset.

Values:

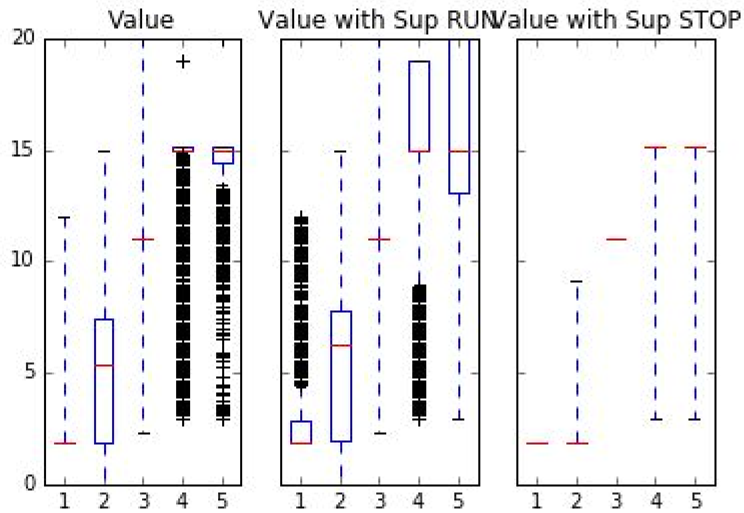


Figure 46 AHU 15 Value Box Plot

1. Cooling coil valve %open
2. Mixed air damper % open
3. Pre-heat coil discharge valve
4. re-heat coil 1 Discharge Valve
5. re-heat coil 2 Discharge Valve

Result: the re-heat coil 1 Discharge Valve,re-heat coil 2 Discharge Valve have outliers in the overall plot but the value when it is stopped, is inconclusive.

Bivariate:

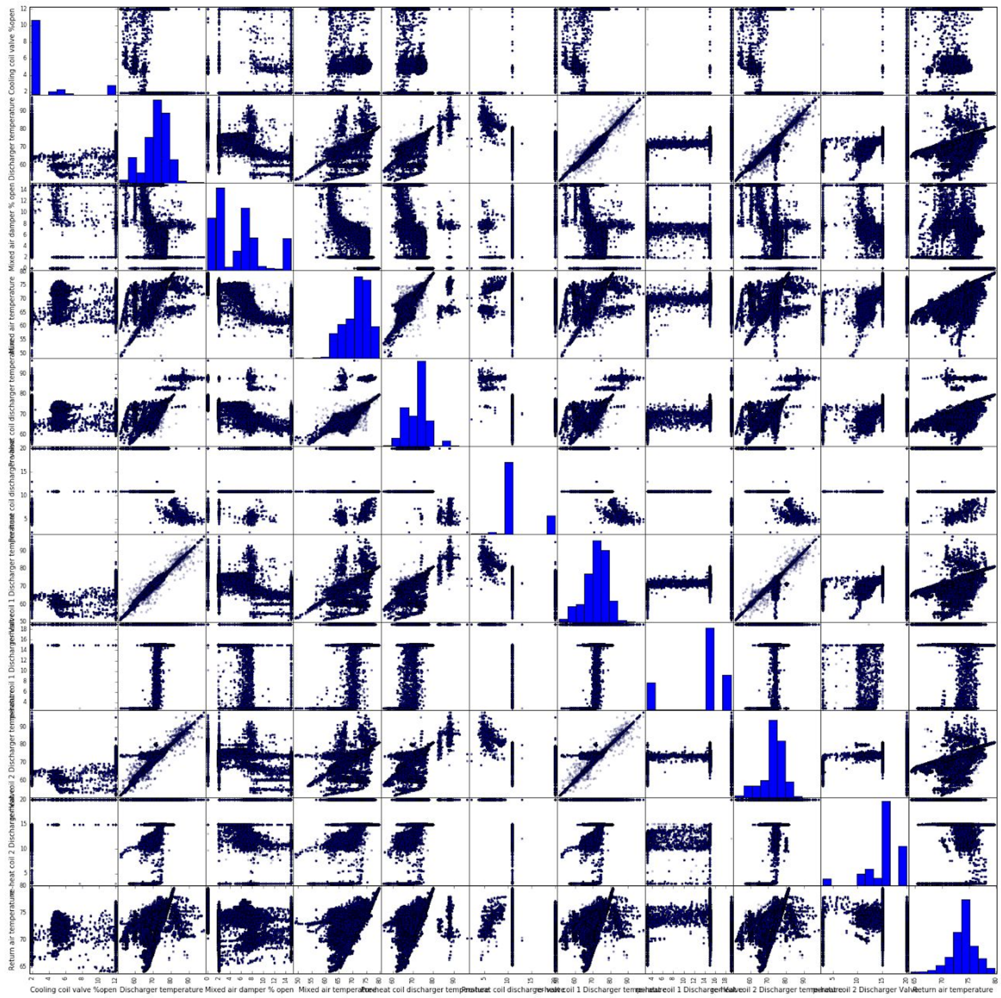


Figure 47 AHU 15 Scatter Plot

AHU 18:

Temperature:

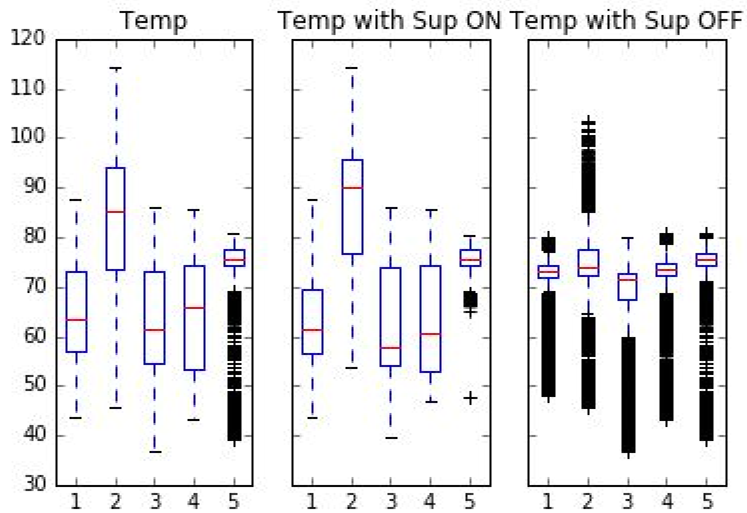


Figure 48 AHU 18 Temperature Box Plot

1. 'Cold Deck temperature
2. Hot Deck temperature
3. 'Mixed air temperature
4. Pre-heat coil discharge temperature
5. Return air temperature

Result: there are a lot of outliers when the supply is OFF. In comparison, the outliers are lesser of the Return Air temperature when the supply is ON. This dataset is chosen for further analysis to find time dependant results.

Values:

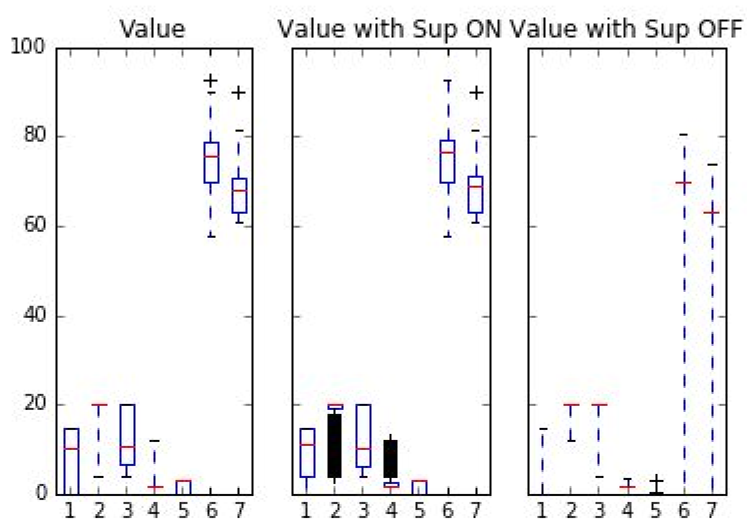


Figure 49 AHU 18 Value Box Plot

1. Mixed Air damper %open
2. 'Pre-heat coil discharge valve
3. 'Return air valve
4. Cooling coil valve %open
5. supply static pressure
6. Supply fan VFD speed
7. 'Return fan VFD speed

Result: the AHU seems to be mostly ON with the values when the supply is switched OFF is inconclusive.

Bivariate:

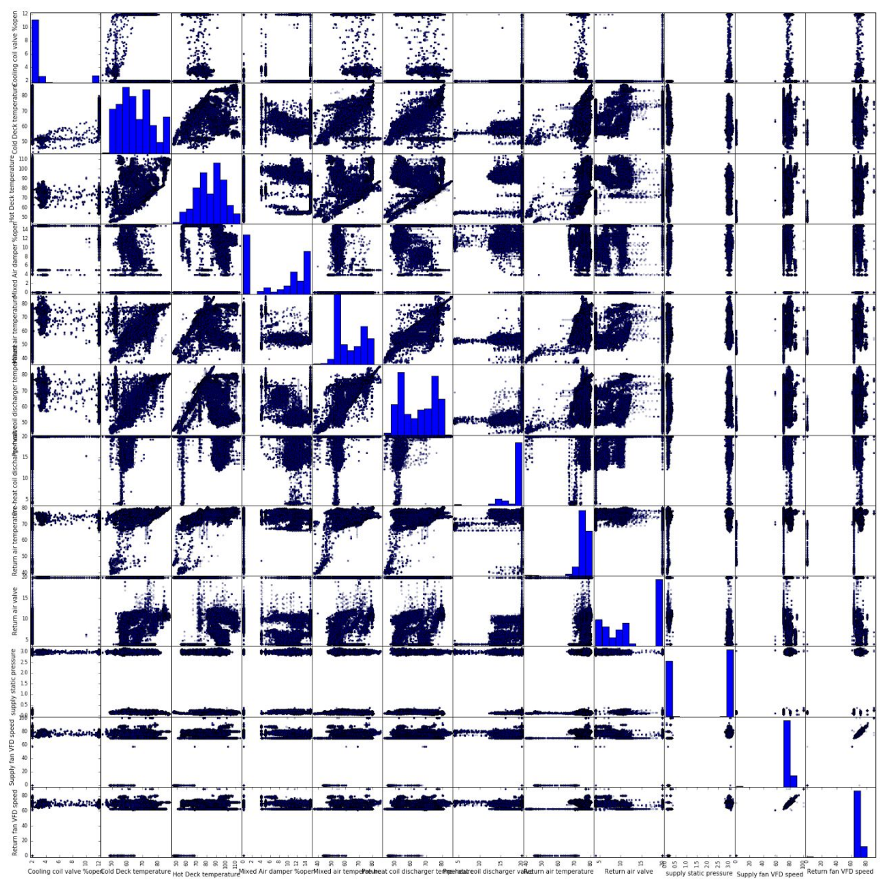


Figure 50 AHU 18 Scatter Plot

CHW:

Temperature:

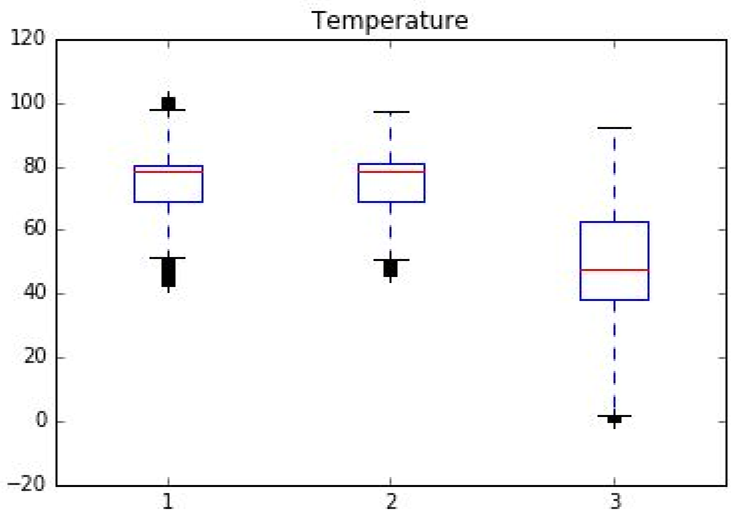


Figure 51 CHW Box Plot

1. Supply Temperature
2. CW return temperature
3. Outside Temperature

Result: The temperatures in the CHW look alright without many outliers. There are however, a lot in the supply temperature.

Values:

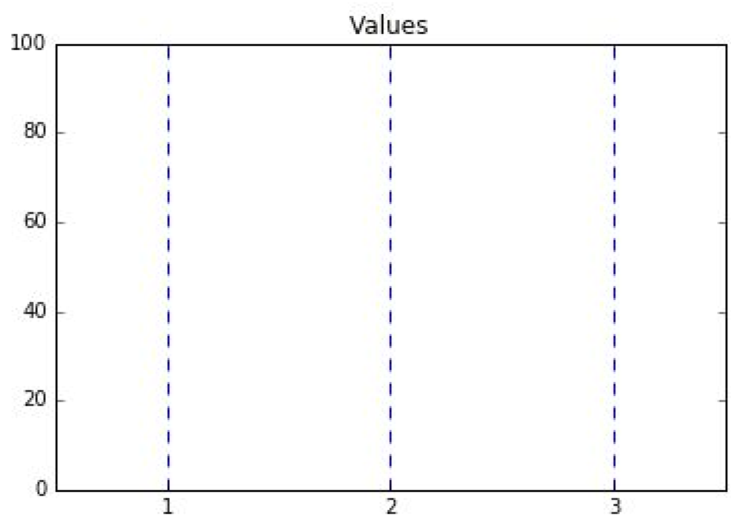


Figure 52 CHW Values Box Plot

1. Chiller 1 Capacity
2. Chiller 2 Capacity

3. Chiller 3 Capacity

Result: The result on the capacity is inconclusive as the ranges are varied too much.

Bivariate:

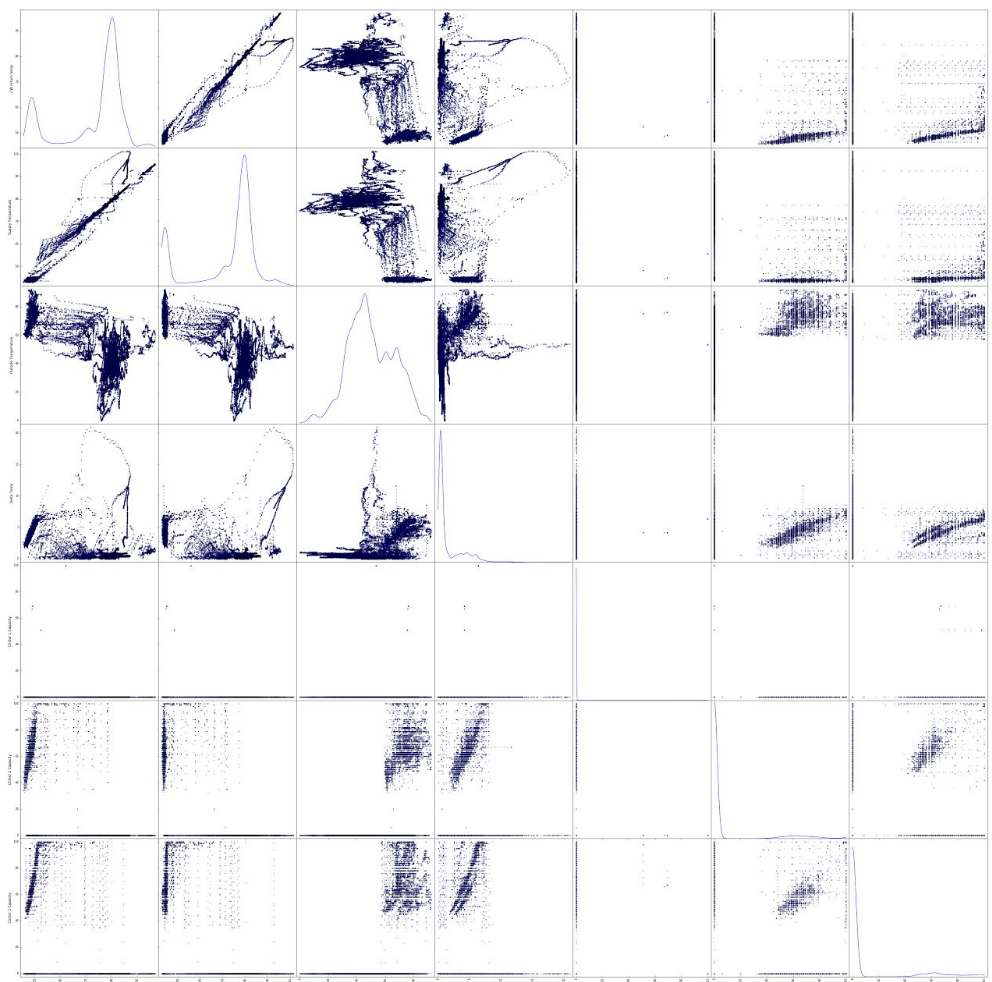


Figure 53 CHW Scatter Plot

3.3 Animation

<u>Purpose:</u>

To figure out the efficiency of the chilled water unit, we create the animation of between delta temperature vs chiller capacity. Check the efficiency by see how fast the temperature changes while the chiller changes the percentage. Also check whether chiller is work when the outside temperature is high, to determine chiller not work properly if chiller not turn on when the outside temperature is high.

From the animation, we can detect whether chiller work properly in general, figure out the when and which chiller has potential errors.

<u>Data Used:</u>

The original data is record from 10/31/2015 to 6/23/2016 of chiller 1, chiller 2 and chiller 3 for IIT Tower building. **But the data for animation** only contain at least one chiller is working (chiller 1, chiller 2 and chiller 3) from 4/20/2016 to 6/23/2016

Chiller 2 and Chiller 3 vs Outside Temperature

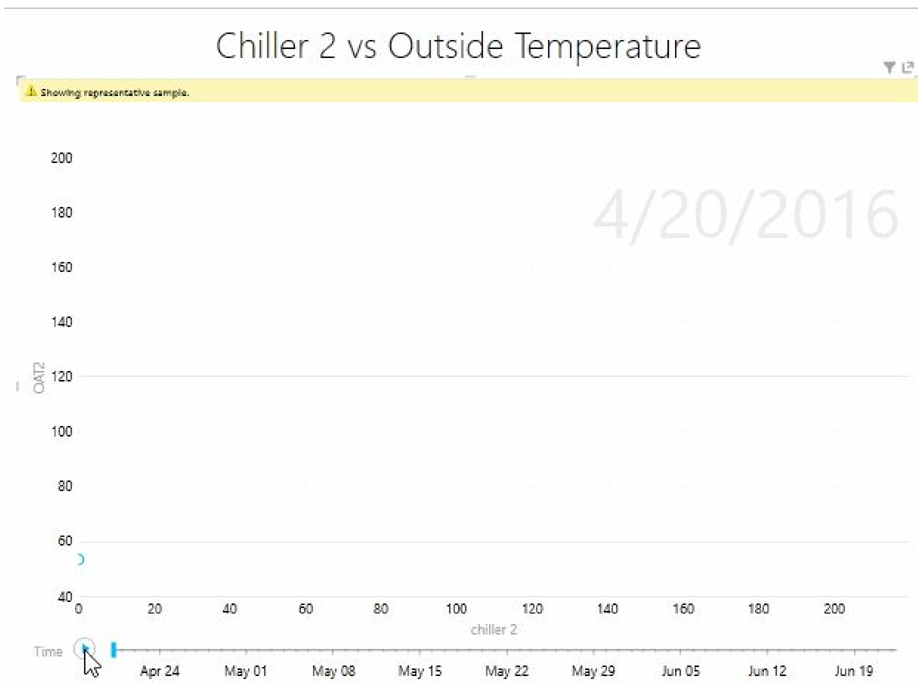


Figure 54 Chiller 2 VS Outside Temperature

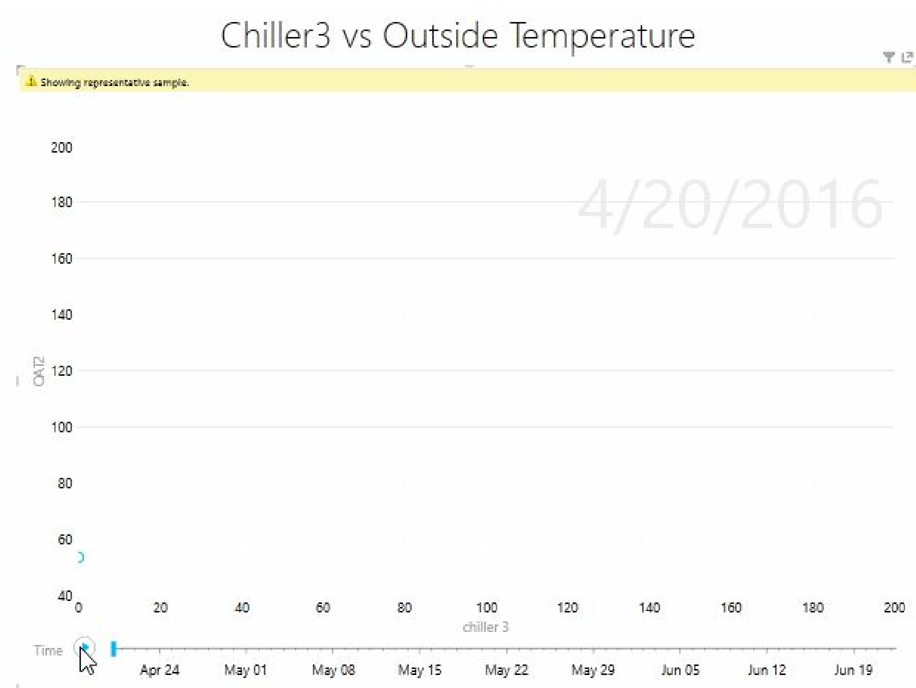


Figure 55 Chiller 3 VS Outside Temperature

Conclusion:
We can see chillers trigger to work when the Outside temperature reach 60 to 70 F, but no other clear pattern between chiller and outside temperature.

Chiller 2 and Chiller 3 vs Delta Temperature between Supply temperature and Return temperature

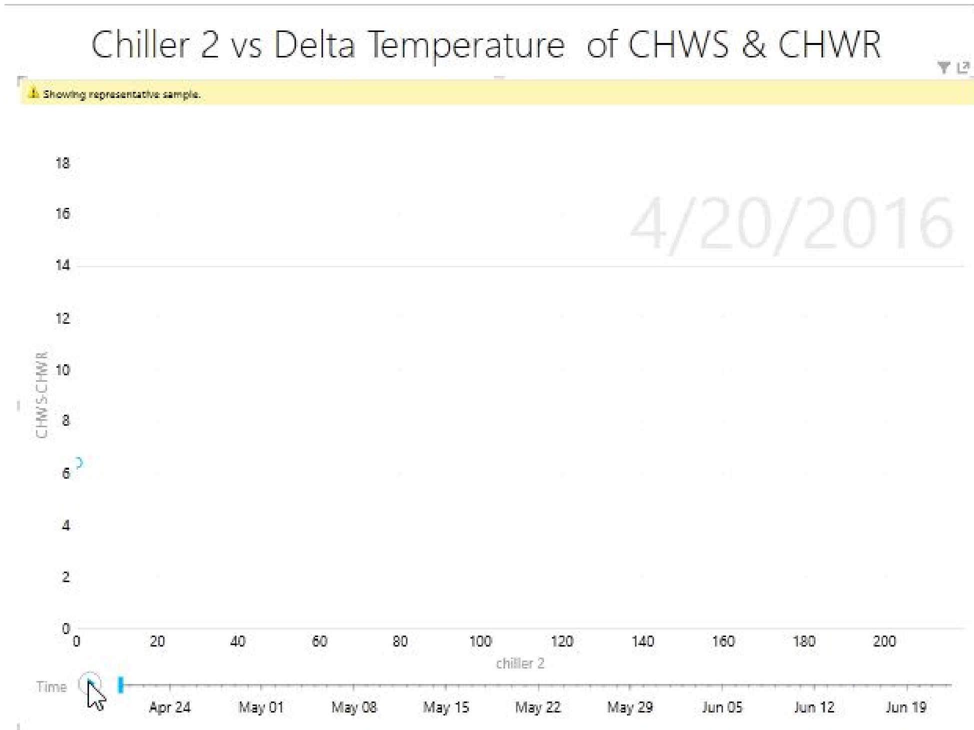


Figure 56 Chiller 2 VS Delta Temperature of CHWs and CHWR

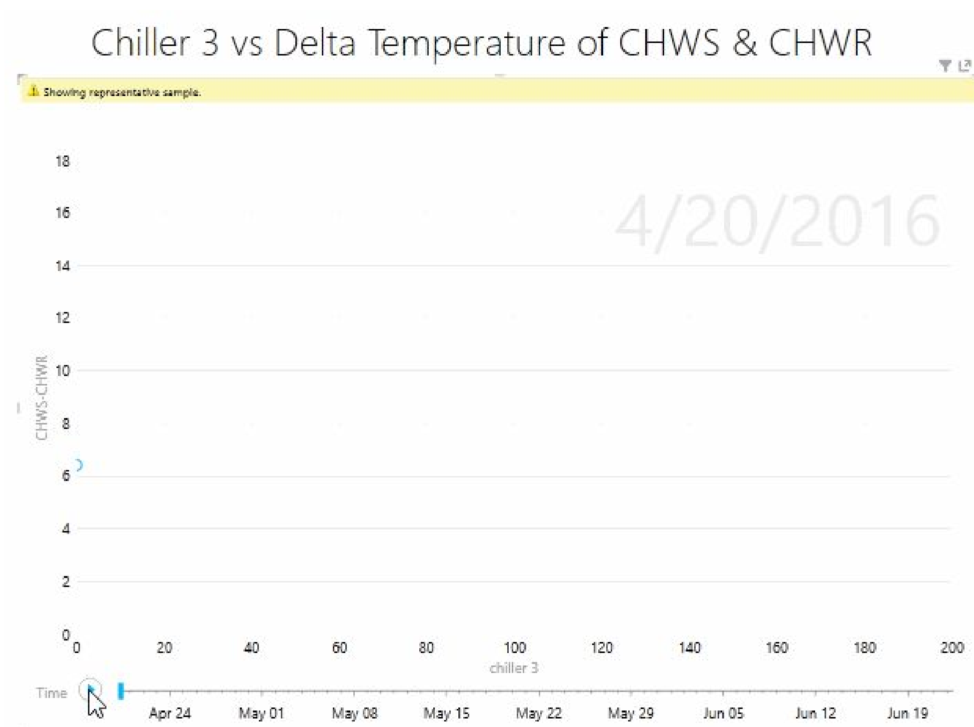


Figure 57 Chiller 3 VS Delta Temperature of CHWs and CHWR

Conclusion:
There are clear linear pattern between delta temperature and the chiller open, which indicates the larger the chiller open, the more difference between supply temperature and the return temperature, It's a sign that chiller works properly to change the temperature. In the figure and animation show that Chiller 3 have a better linear relationship with the delta temperature.

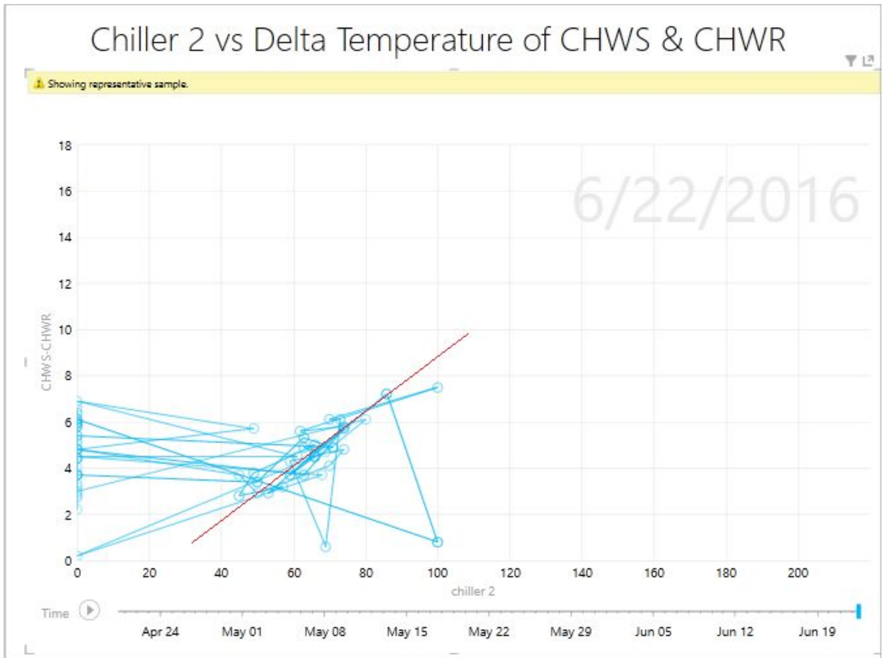


Figure 58 Chiller 2 VS Delta Temperature of CHWs and CHWR

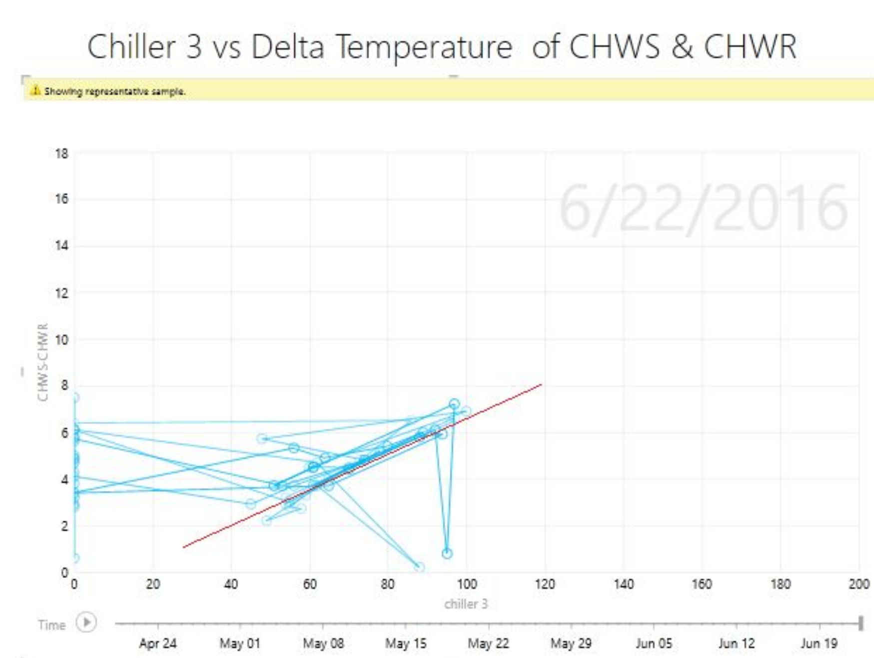


Figure 59 Chiller 3 VS Delta Temperature of CHWs and CHW

3.4 Time series:comparison of the various effects over time for a work order.

	Wo Number	Description	Assigned To	Building	Floor	Room	Craft	Enter Date	Wo Close Date	Hours SUM	Date	Time
750	WO122352	ROOM IS TOO COLD	TJABCZYN	IT	18.0	18C9-1	HVAC	2015-01-13 09:48:35	2015-01-14 14:37:09	1.00	2015-01-14	14:37:09
924	WO122529	ROOM IS TOO COLD	TARIANOU	IT	18.0	18F4-1	HVAC	2015-01-16 10:54:44	2015-01-16 15:32:41	0.75	2015-01-16	15:32:41
4579	WO126183	ROOM IS TOO COLD	TJABCZYN	IT	18.0	18E4-1	HVAC	2015-04-02 11:37:40	2015-04-03 14:30:12	2.00	2015-04-03	14:30:12
4600	WO126205	ROOM IS TOO COLD - 18E4-1	DBROOKS4	IT	18.0	18E4-1	HVAC	2015-04-02 15:00:17	2015-04-03 13:33:07	0.50	2015-04-03	13:33:07
4718	WO126322	ROOM IS TOO COLD	TJABCZYN	IT	18.0	18E4-1	HVAC	2015-04-06 09:49:47	2015-04-08 15:01:49	0.50	2015-04-08	15:01:49
5392	WO126995	ROOM IS TOO COLD	TJABCZYN	IT	18.0	18D3-2	HVAC	2015-04-21 08:51:55	2015-04-22 14:40:14	1.00	2015-04-22	14:40:14
5400	WO127004	ROOM IS TOO COLD	TJABCZYN	IT	18.0	18C7-1	HVAC	2015-04-21 09:26:50	2015-04-22 14:44:26	1.00	2015-04-22	14:44:26
6427	WO128034	ROOM IS TOO COLD	TJABCZYN	IT	18.0	18F8-1	HVAC	2015-05-12 09:57:35	2015-06-01 15:06:57	1.00	2015-06-01	15:06:57
6433	WO128038	ROOM IS TOO COLD - 18E4-1	TJABCZYN	IT	18.0	18E4-1	HVAC	2015-05-12 10:43:16	2015-05-29 14:56:37	0.50	2015-05-29	14:56:37
6647	WO128240	ROOM IS TOO COLD	TJABCZYN	IT	18.0	18E4-1	HVAC	2015-05-18 10:24:11	2015-05-29 14:55:14	0.50	2015-05-29	14:55:14
9372	WO130987	ROOM IS TOO COLD	TJABCZYN	IT	18.0	18E4-1	HVAC	2015-07-13 13:18:27	2015-07-14 14:45:19	1.50	2015-07-14	14:45:19
14011	WO135624	ROOM IS TOO COLD	TJABCZYN	IT	18.0	18E4-1	HVAC EQUIP	2015-10-05 09:50:52	2015-10-08 15:03:05	1.00	2015-10-08	15:03:05
19516	WO141158	ROOM IS TOO COLD	TARIANOU	IT	18.0	18D3-1	HVAC EQUIP	2016-01-21 15:08:25	2016-01-22 16:06:45	1.00	2016-01-22	16:06:45
20548	WO142182	ROOM IS TOO COLD	TJABCZYN	IT	18.0	18D3-2	HVAC EQUIP	2016-02-10 10:48:55	2016-02-16 14:57:32	1.00	2016-02-16	14:57:32
24388	WO146029	ROOM IS TOO COLD	TJABCZYN	IT	18.0	n/a	HVAC EQUIP	2016-04-27 09:12:11	2016-04-27 15:09:15	0.50	2016-04-27	15:09:15

Table 4 Comparison of the Various Effects

The last three overlap with the AHU data. This was taken into consideration for the comparison on the data. The report takes 7 days prior to the work order request. Here is a time comparison between the various variables of the AHU18 over time for this 7 day time period. The last day /block is when there was a call on the service.

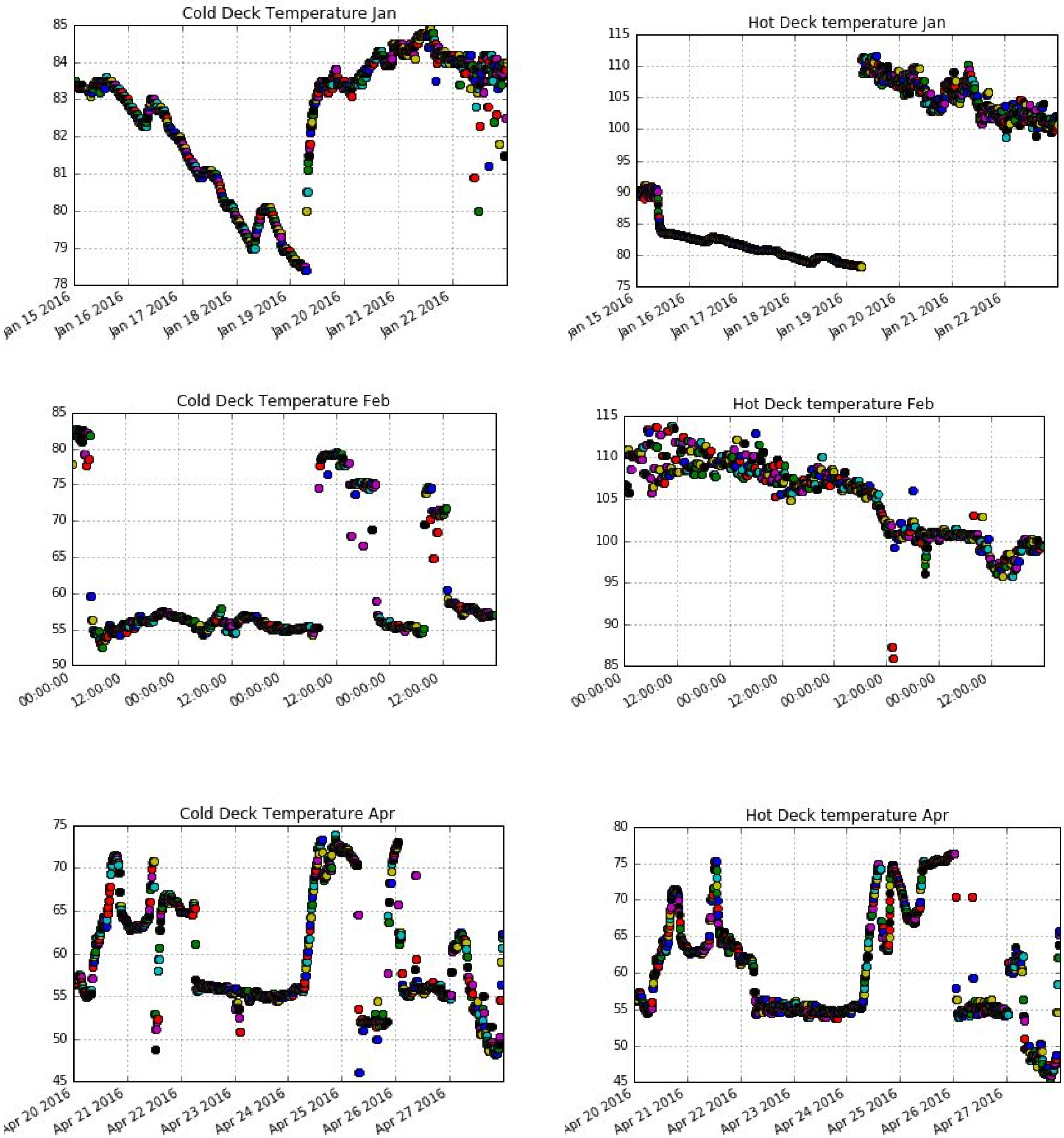


Figure 60 Hot and Cold Deck Temperature Comparison with time

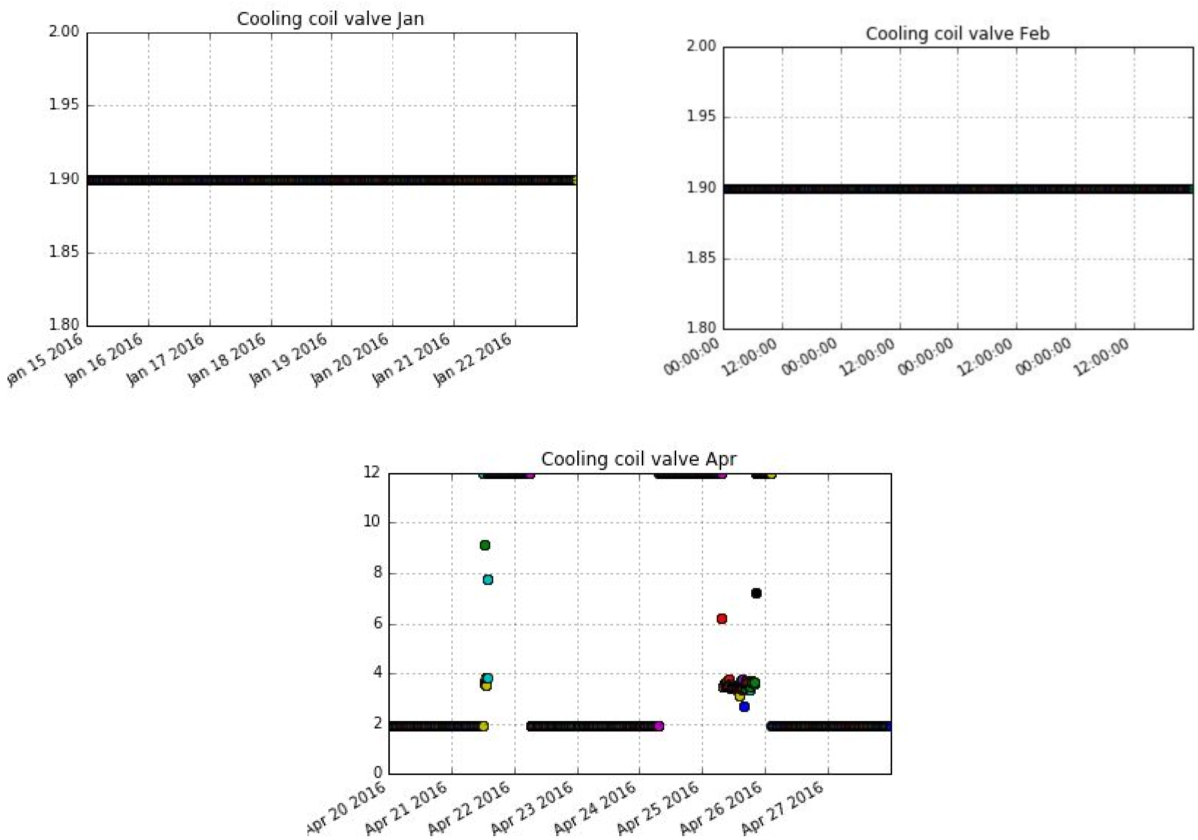
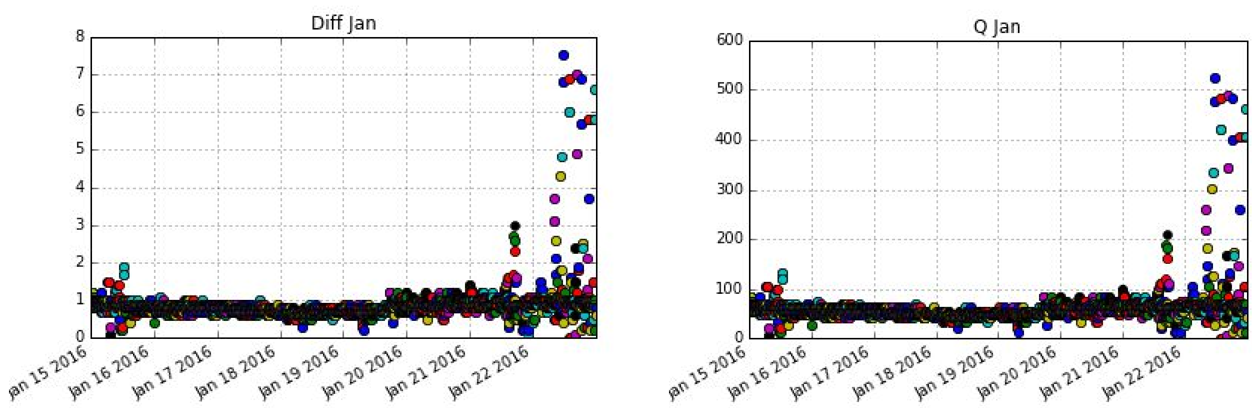


Figure 61 Cooling Coil Status



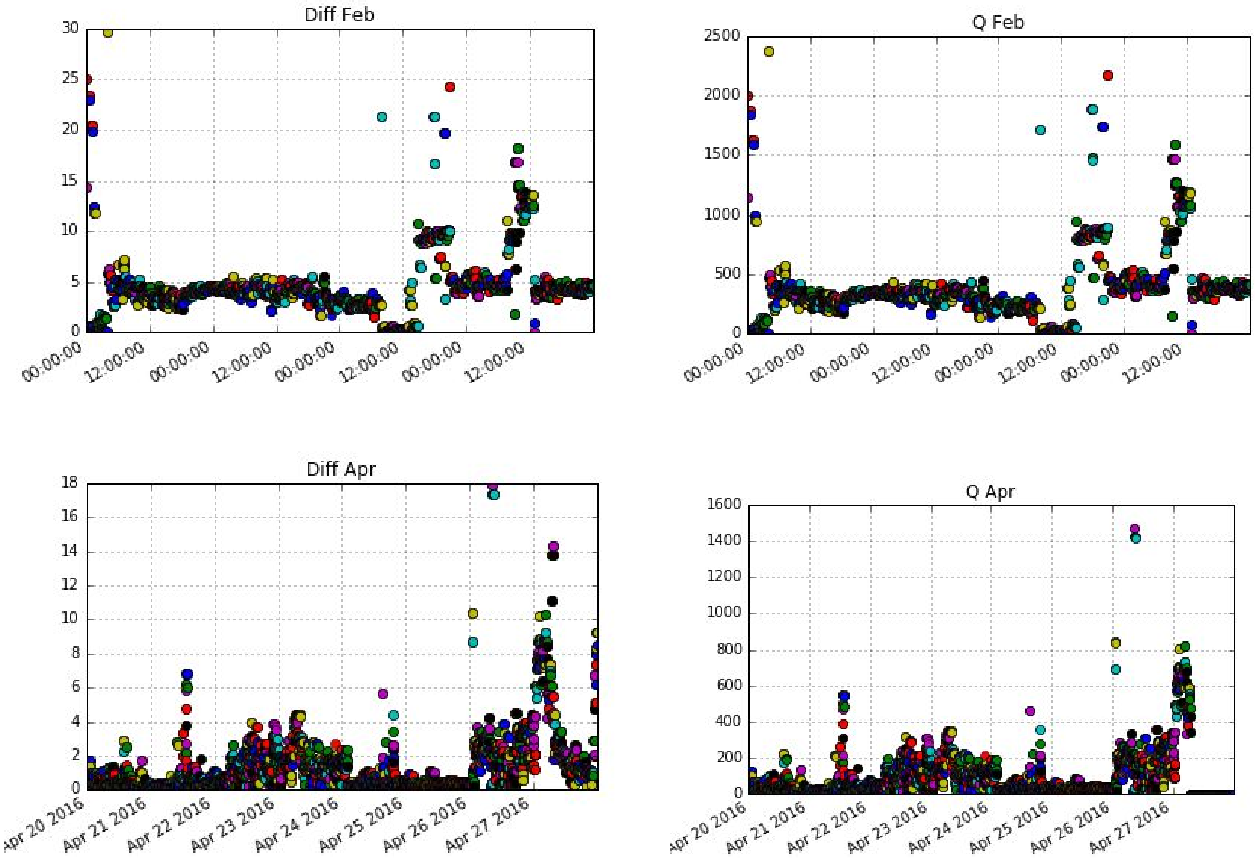
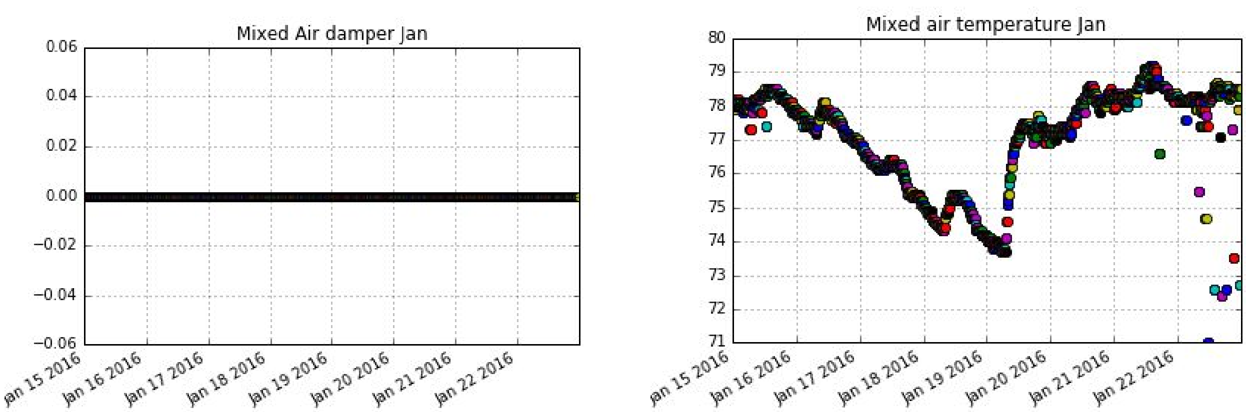


Figure 62 Diff and Q Comparison



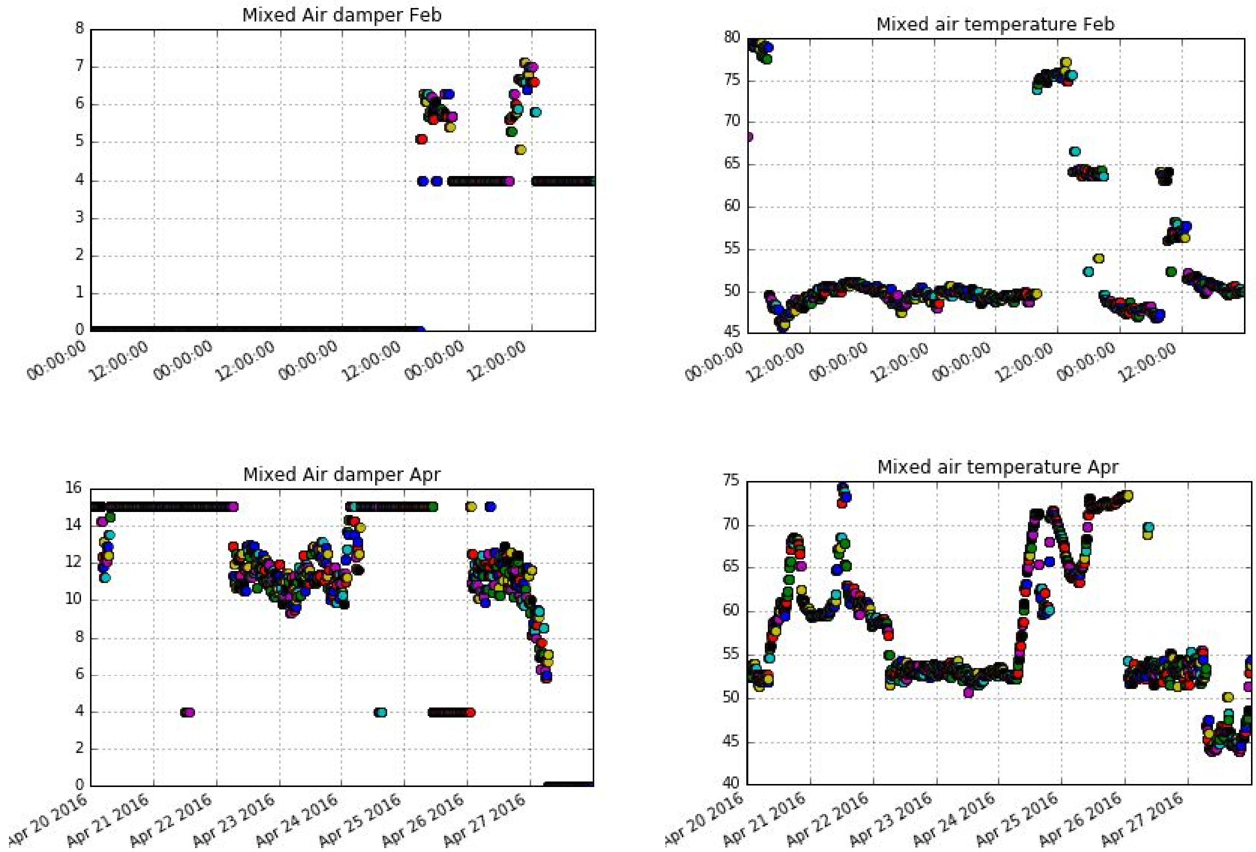
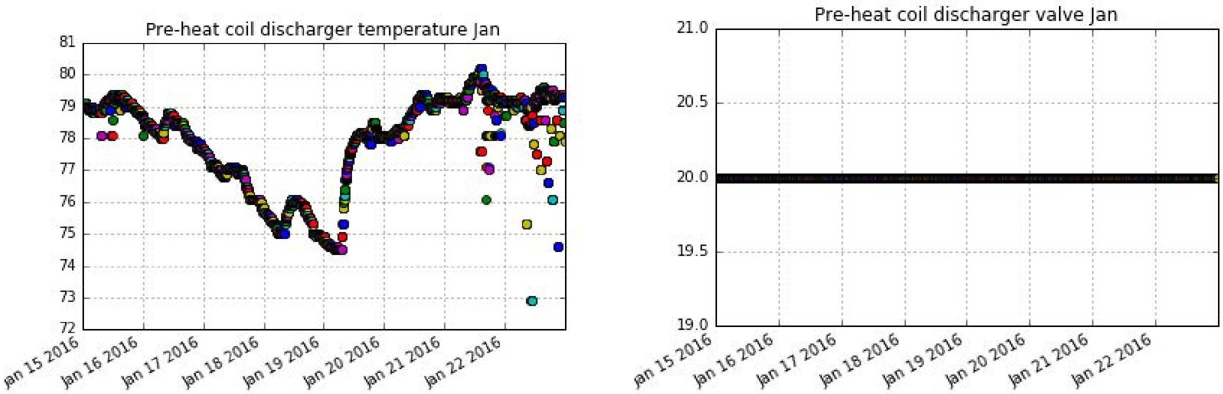


Figure 63 Mixed Air and Mixed Air Temperature Comparison



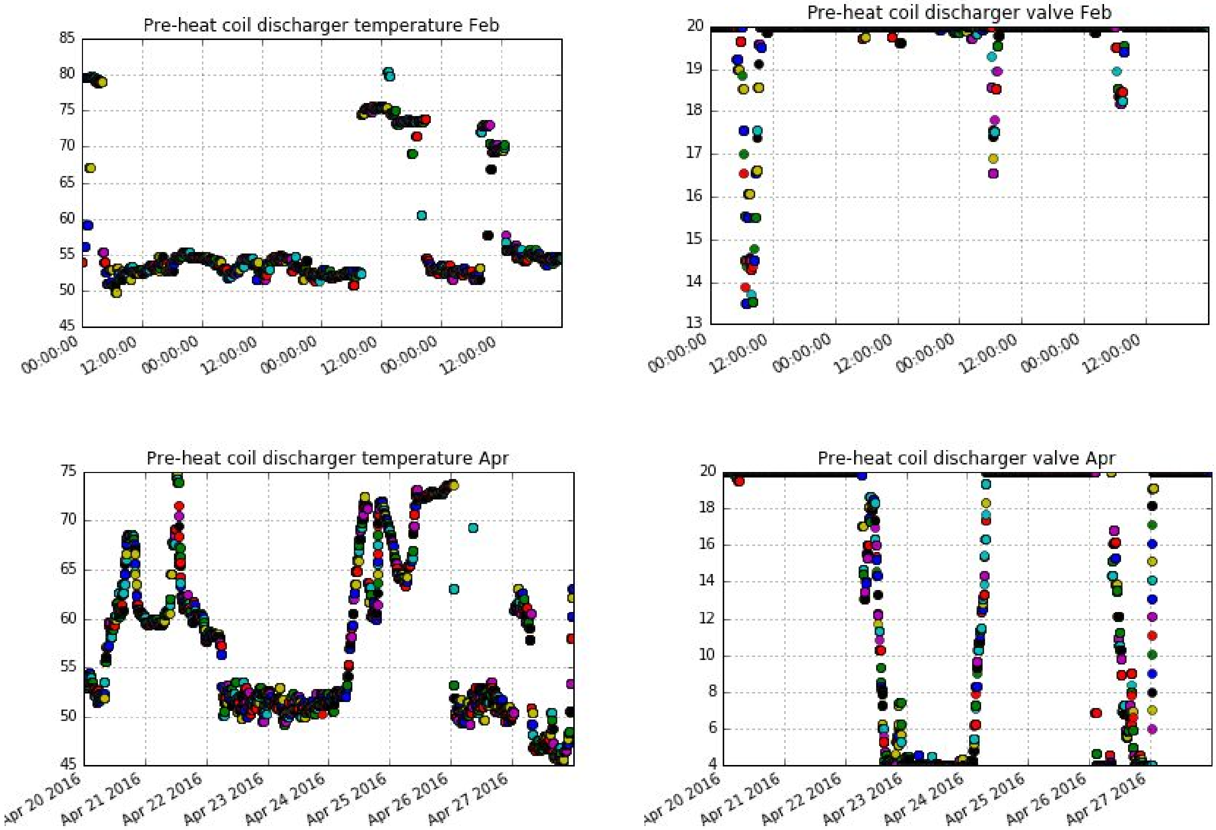
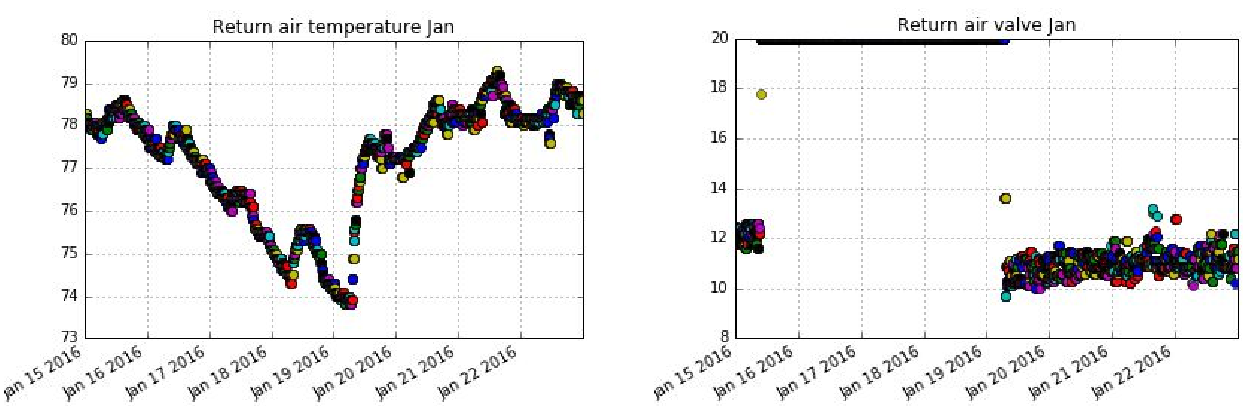


Figure 64 Preheat Coil Discharge Temperature and Valve Comparison



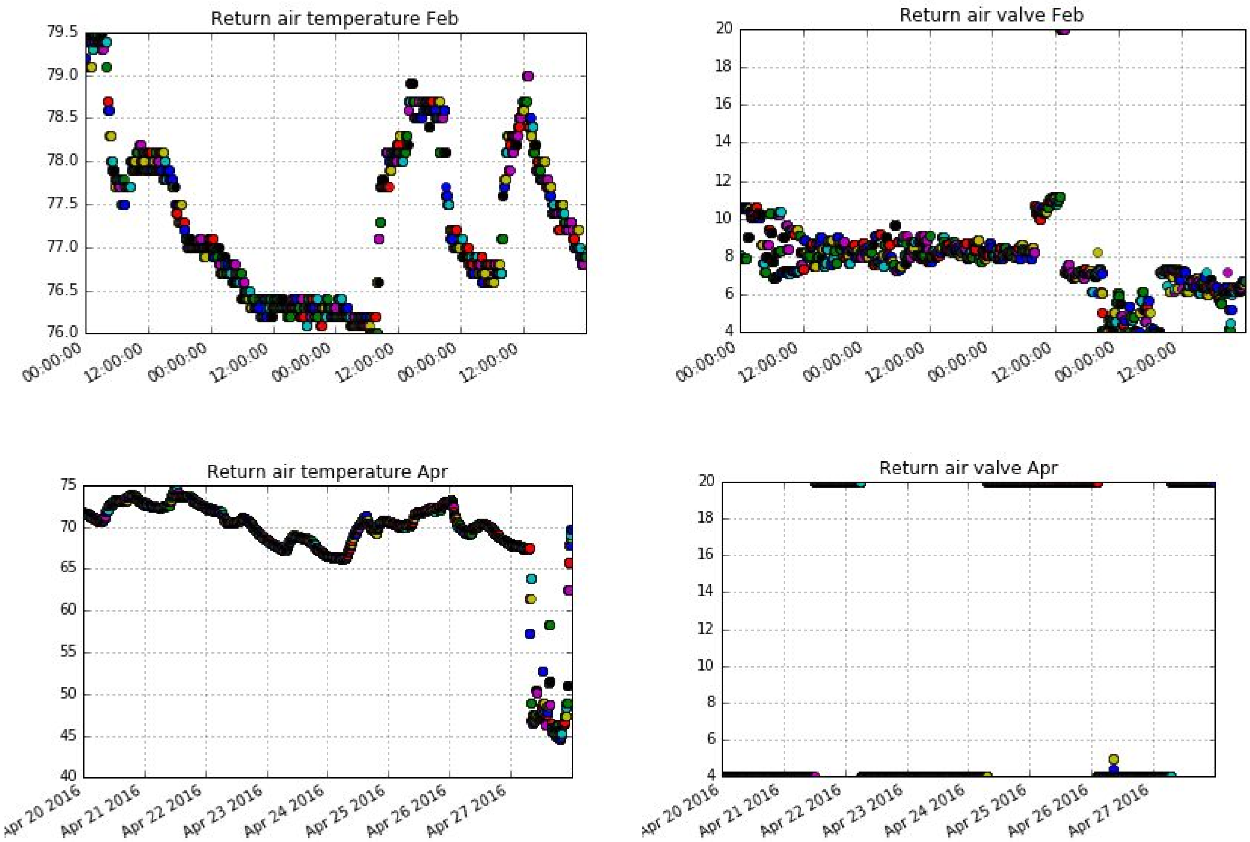


Figure 65 Return Air Temperature and Valve Comparison

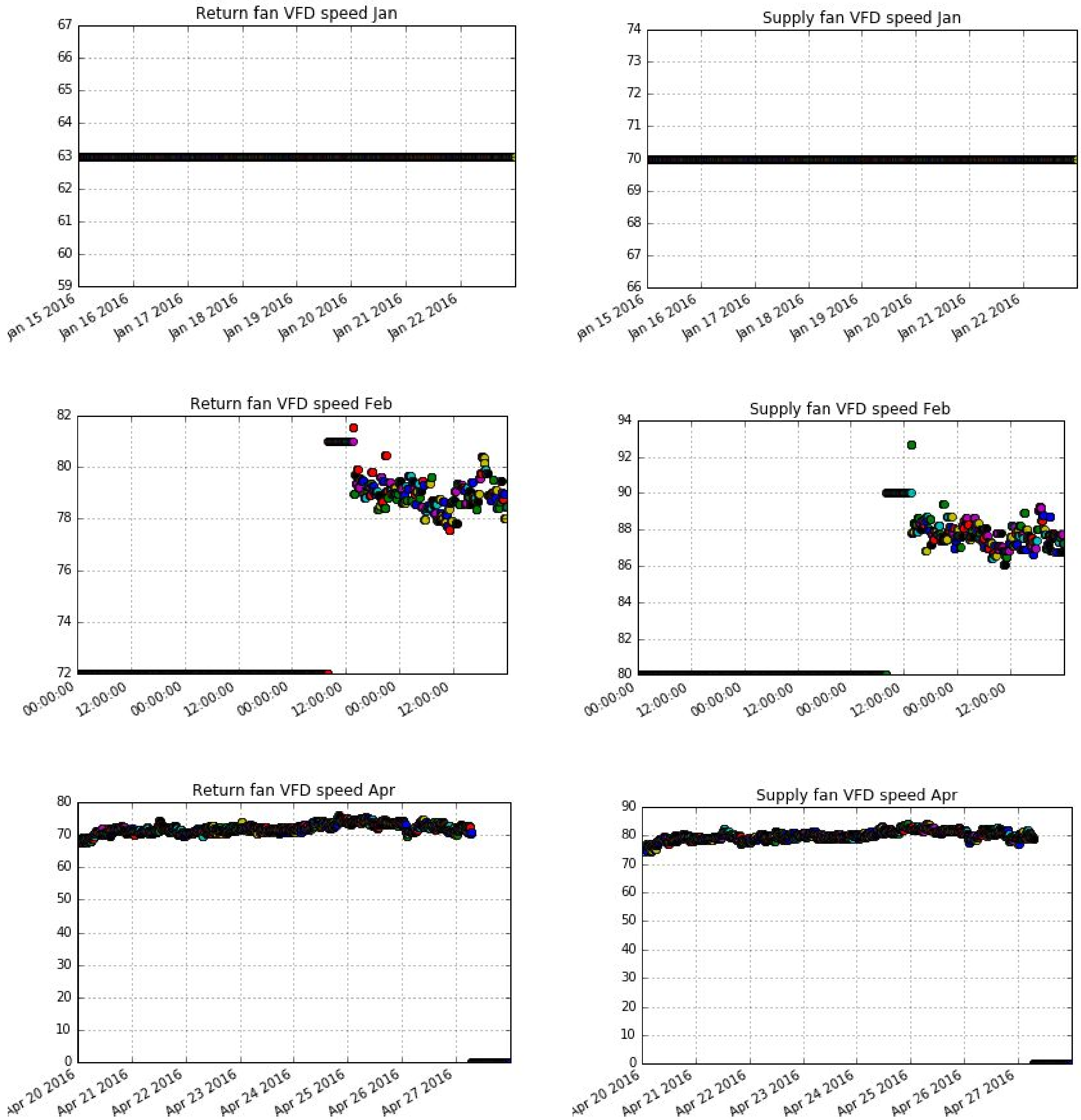


Figure 66 Supply Fan VFD Status

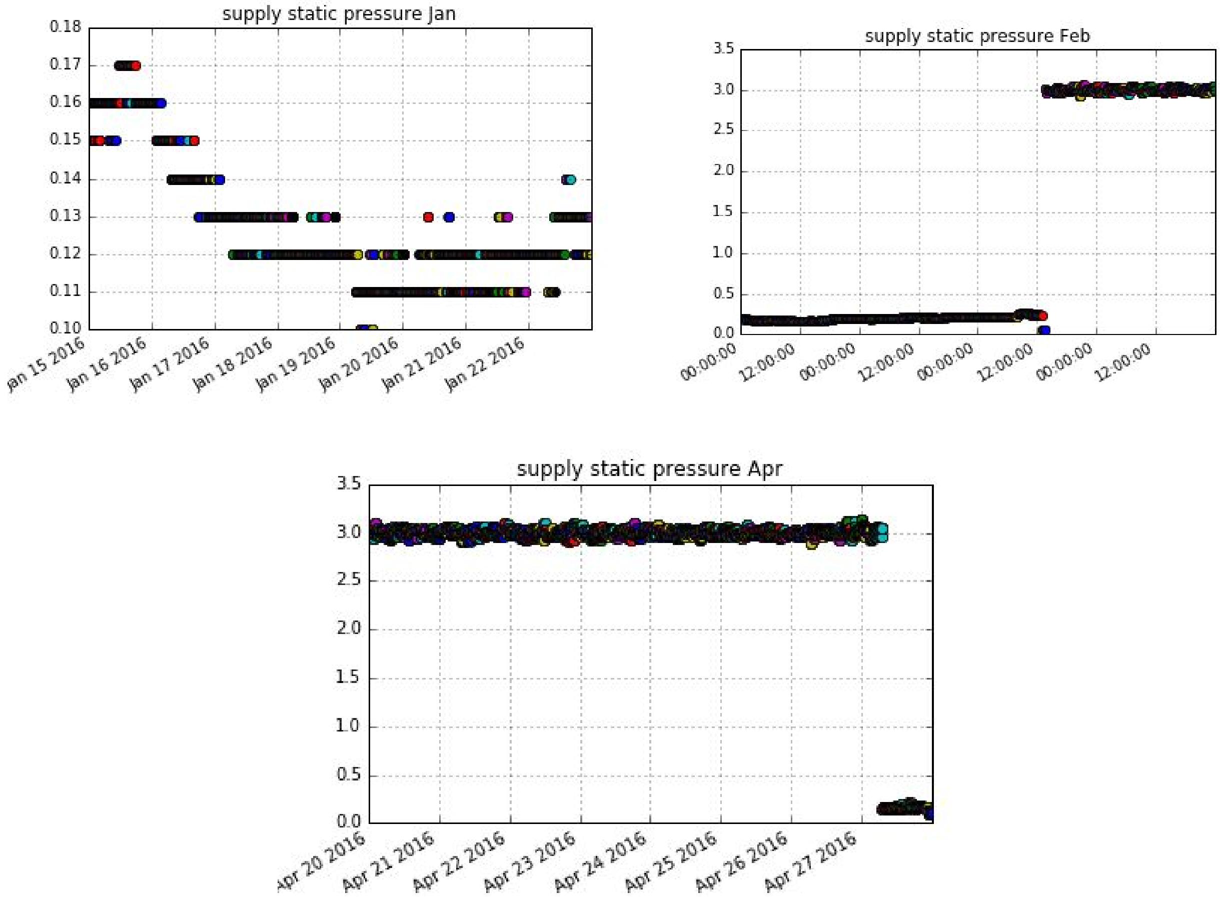


Figure 67 Supply Static Pressure Status

Result: we do not find any conclusive pattern to analyse for this particular work order.

4. Postmortem

1. Collect same feature data for the same type devices(ex. Each AHU collect different feature data, makes it hard to compare them in general)
2. Figure out better way (numerical way) to assess the building condition rather than work order description
3. Database - Suggested Column Headings
    a. **AHU:**

Abbr	Available
R	1
SAF/S	14
CCO/CCV	18
CDAT	1
CDT	6
DAT	10
DSP_S17	1
DX1_S17	1
DX2_S17	1
EFCAWA/E	12
EFGEN	2
EFNEC	1
FBD_S17	1
FBO	1
HCO	1
HDT	6
HDV	1
HVO/HV	2
p	1
KH_S17	1
KL_S17	1
MAD	16

MAM_S17	1
MAT	16
MND	4
PHT	13
PHV/PH1V	15
PH2V	1
PVM	1
RAT	12
RH1T/R1T_S10	8
RH1V_S3/RHV_S6/R1V_S10	14
RH2T/R2T_S10	8
RH2V_S3/R2V_S10	8
RVD	1
SAF	3
SDT	1
SFVFD:RUN.STOP_S15	1
SFVFD:SPEED/SVD/SFSPD	9
SSP_18	1
TMD	1
E13RVD_S14	1
S15E15	1
S15RAT	1
E23RVD	1
Q = VFD*(PHT-MAT)	0
Diff = PHT - MAT	0

Table 5 AHU Postmortem

b. **WorkOrder:**
   i. Work order number
   ii. Description
   iii. Building

iv. Floor
v. Room
vi. Craft
vii. Enter Date
viii. Close Date
ix. Hours SUM
x. Requestor

Softwares Used:

1. R (ggplot2, RODBC,)
2. PowerBI - for graphs
3. Python (matplotlib, pandas, numpy, collections, nltk, string)
4. Microsoft (Excel (VB-Macros, graphs),Microsoft SQL Server Management System, Azure, Access DB)
5. Google (Drive, Documents)
6. Basecamp2
7. SQL

Appendix

Tables:

Figures: