Predicting energy consumption pattern of HVAC using ML techniques

Ayan Nath
Ravikiran Anapagaddi
Rajat Nandan NS

Initial proposed title: Predicting optimum cooling air supply of room indoor environment using ML techniques

Introduction

Energy used by buildings worldwide is over 40%[1] of global energy consumption

Recently multiple building data are available from the diff. sensor data connected to building management systems

ML techniques can be analyzed to evaluate various sensor data and relate with its energy consumption pattern
These data/techniques can help to develop strategies for reducing energy consumption of buildings

Objective:
We want to analyze the energy consumption pattern of building using ML techniques
How to operate the HVACs to provide effective cooling by minimizing the energy consumption

Explanation of raw dataset

3 years data set* for building is available which includes:

1. Whole-building and end-use energy consumption
2. HVAC system operating conditions
3. Indoor and outdoor environmental parameters
4. Occupant counts

The office building in Berkeley, California.

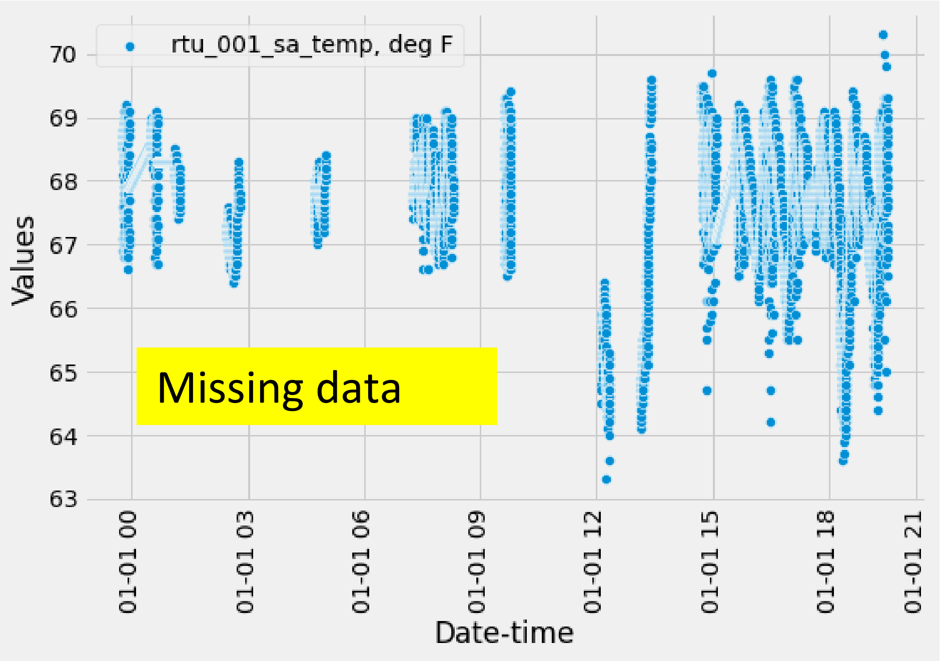
Data	Description	No. of dataset	Sampling frequency
Energy use data	Miscellaneous electric load for North & South Wing	2	15 min
	Lighting load for the South Wing	1	15 min
	Heating Ventilation and Air Conditioning load for North & South Wing	2	15 min
Outdoor environmental data	air temperature, Outdoor air dew temperature, air relative humidity, solar radiation	5	15 min
Indoor environmental data	Cooling, Heating temperature setpoint of Zone *	41+41	5 min
	Zone temperature of interior, exterior zone	16+51	10 min
	CO2 concentration of each zone	13	1 min
Occupant data	Occupant counts in south half of each floor	2	10min
Wifi data	Wifi connection counts in the different building zones	4	10min
HVAC operational data	Roof Top Unit operational characteristics (fan speed, Chilled water temp. etc)	159	1 min

*Hong, Tianzhen; Luo, Na; Blum, David; Wang, Zhe (2022), A three-year building operational performance dataset for informing energy efficiency , Dryad, Dataset

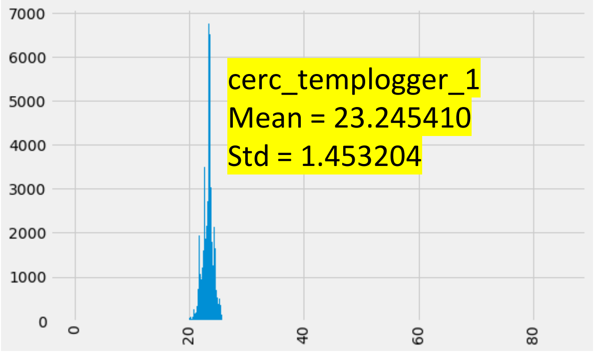
Challenges present in the dataset

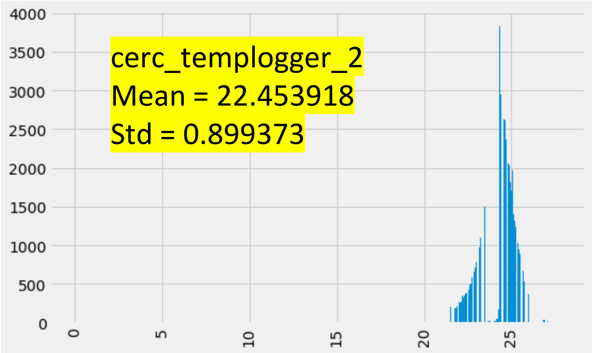
Challenges

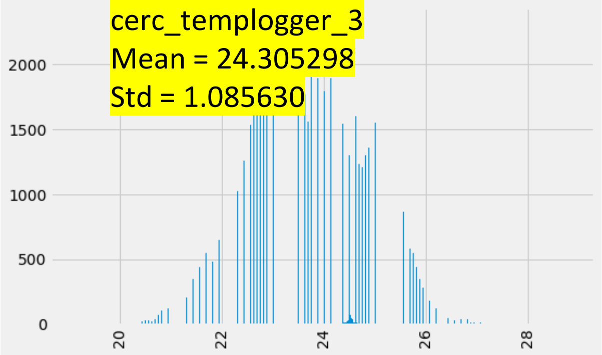
- Data is missing for some ranges
- Some data points were wrong
- Temperature set points are not varying
- All the HVAC systems are not operating continuously making complexity in the predictions
- Data reaching max values

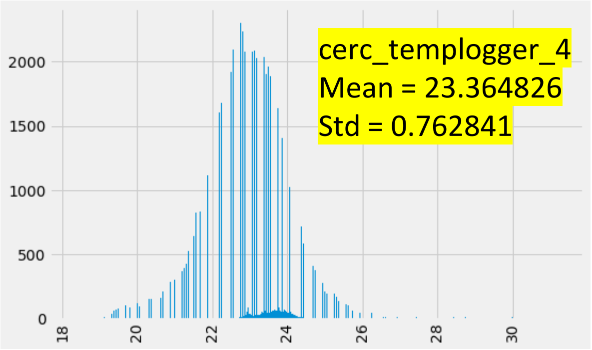


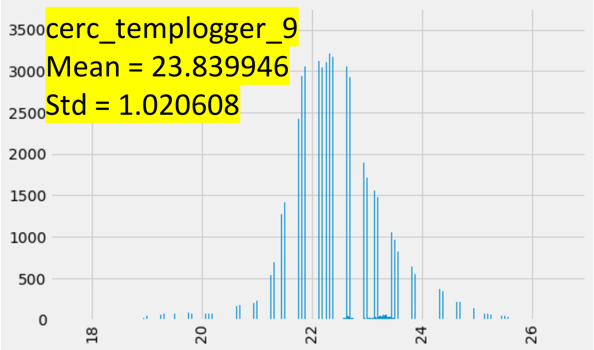
Visualizing the dataset: temp. pattern inside the building

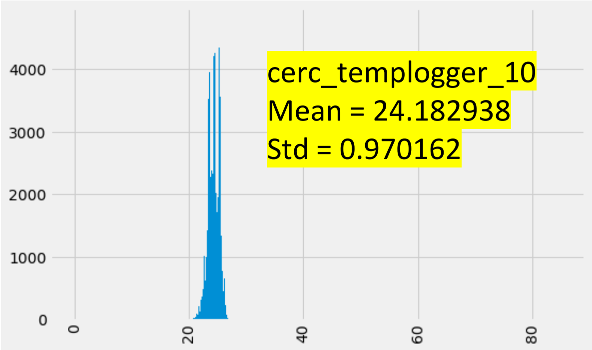


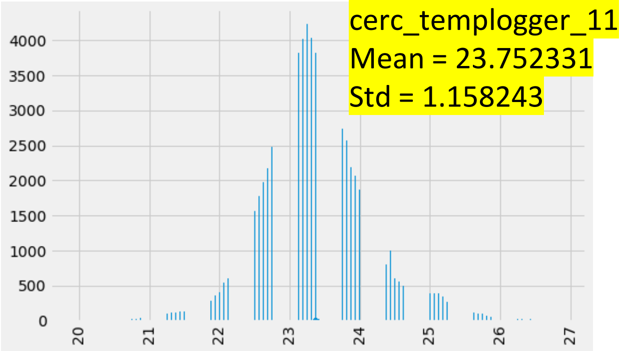


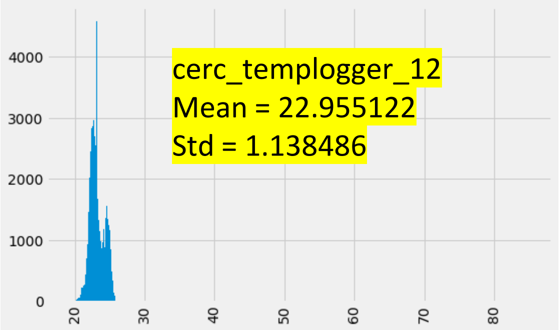






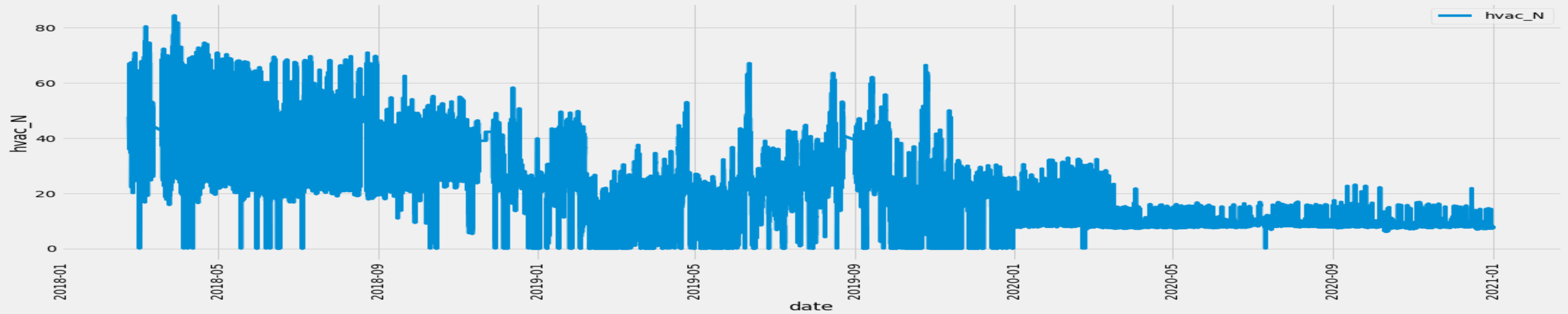


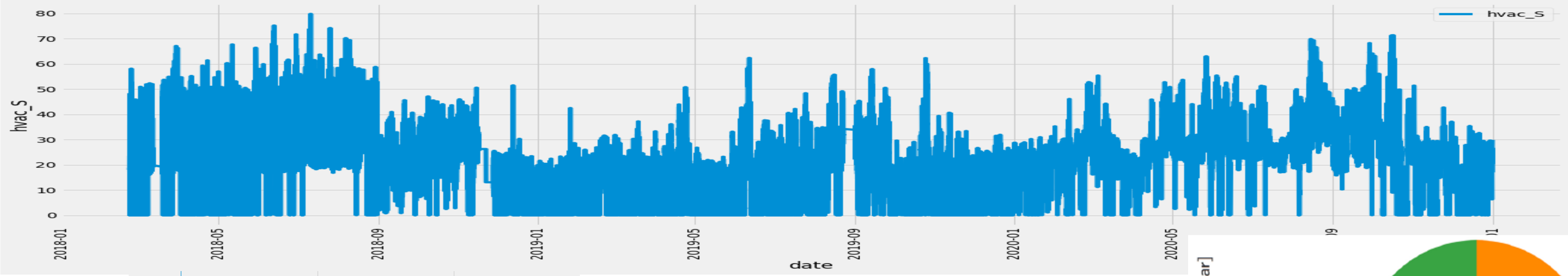


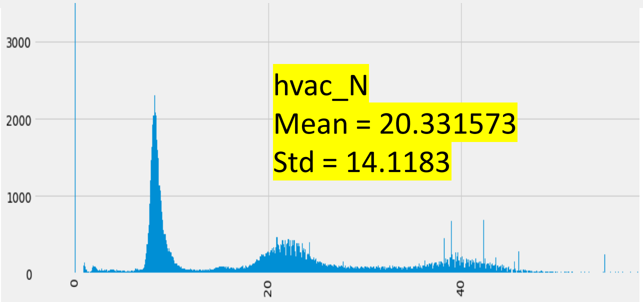


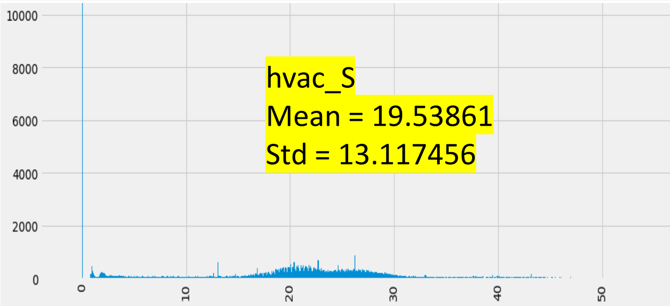
Energy consumption pattern of HVAC system

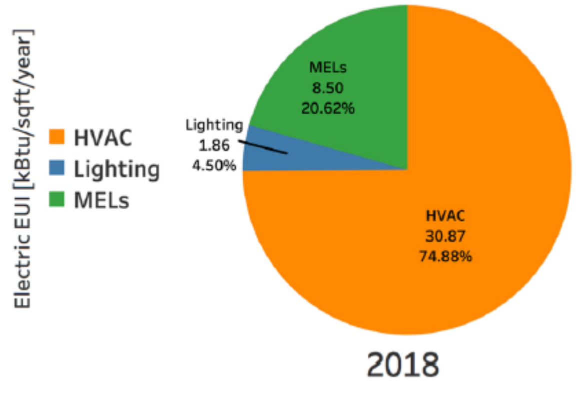
Most of the energy consumption in the building is contributed by HVAC system



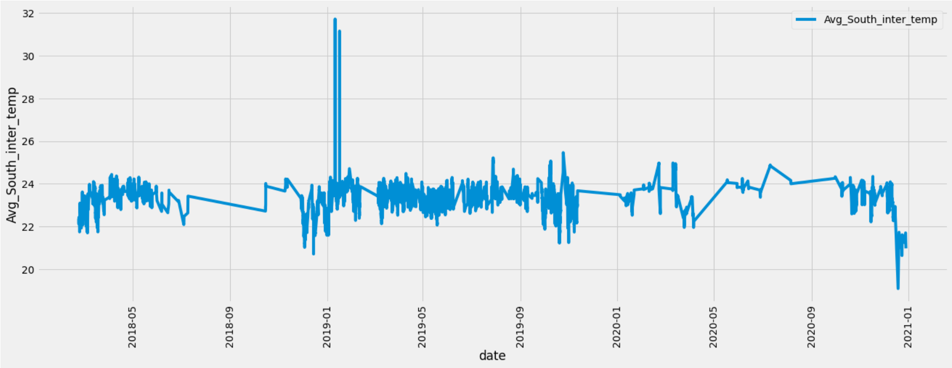


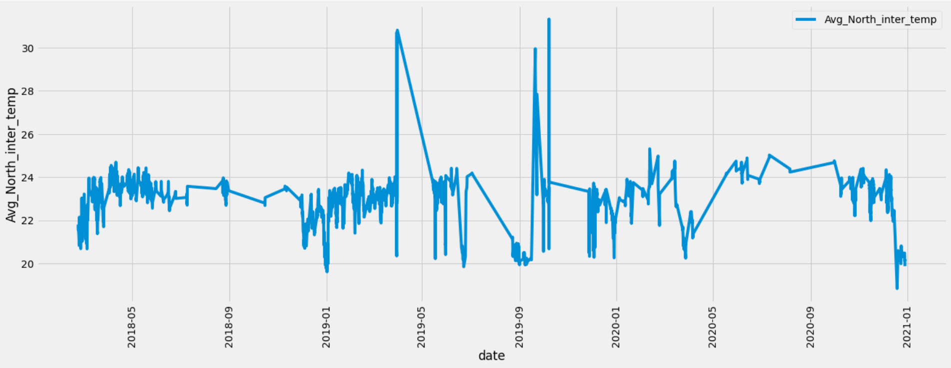


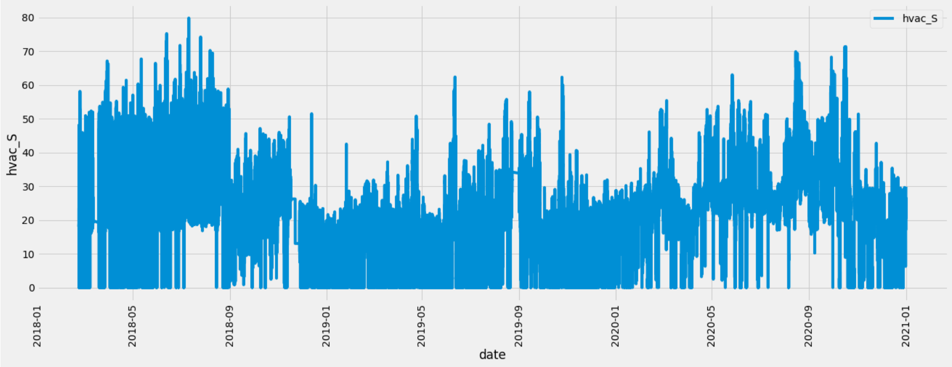


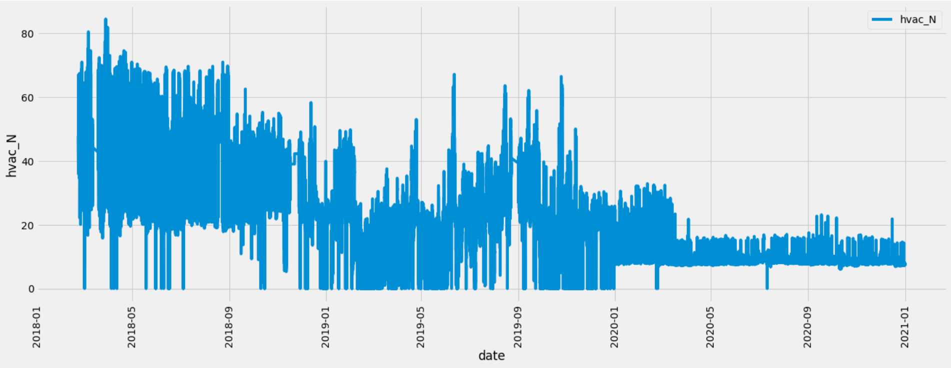


Relate the avg temp & HVAC energy consumtiotion:









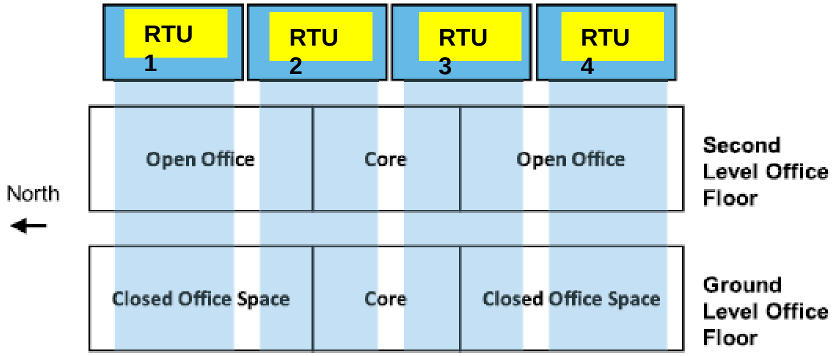
Lighting zone	RTU	Thermal zones
North Wing	1	36, 37, 38, 39, 40, 41, 42, 64, 65, 66, 67, 68, 69, 70
North Wing	2	19, 20, 27, 28, 29, 30, 31, 32, 33, 34, 35, 43, 44, 49, 50, 57, 58, 59, 60, 62, 63, 71, 72
South Wing	3	18, 25, 26, 45, 48, 55, 56, 61
South Wing	4	16, 17, 21, 22, 23, 24, 46, 47, 51, 52, 53, 54

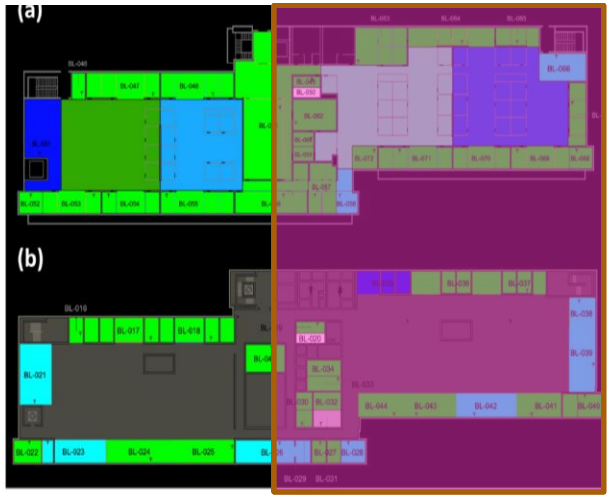
Table 1. Key Electrical Panels.

Correlate fan with multiple zone temp’s

Lighting zone	RTU	Thermal zones
North Wing	1	36, 37, 38, 39, 40, 41, 42, 64, 65, 66, 67, 68, 69, 70
North Wing	2	19, 20, 27, 28, 29, 30, 31, 32, 33, 34, 35, 43, 44, 49, 50, 57, 58, 59, 60, 62, 63, 71, 72
South Wing	3	18, 25, 26, 45, 48, 55, 56, 61
South Wing	4	16, 17, 21, 22, 23, 24, 46, 47, 51, 52, 53, 54

Individual zone temp.

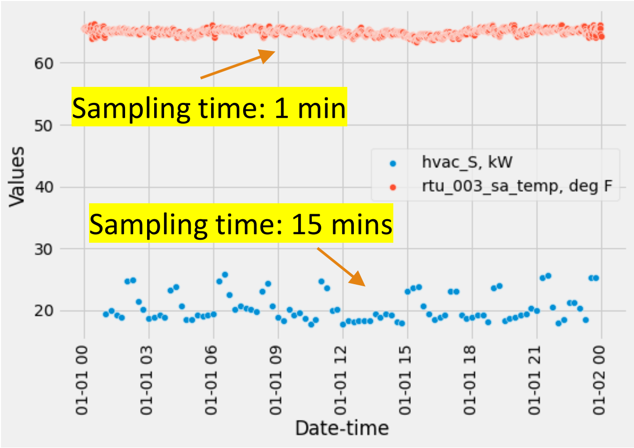




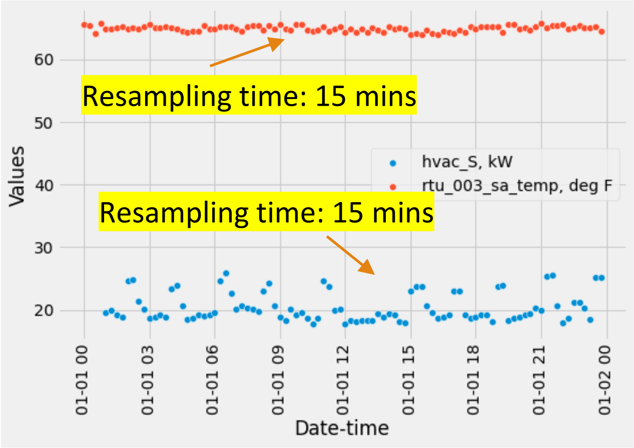
uft_fan_spd.csv	**Supply air fan speed of Zone : 1,2,3,4**
rtu_fan_spd.csv	**supply fan speed : 1,2,3,4** **return fan speed : 1, 2, 3, 4**
rtu_sa_fr.csv	filtered supply air flow rate
rtu_oa_fr.csv	outdoor air flow rate

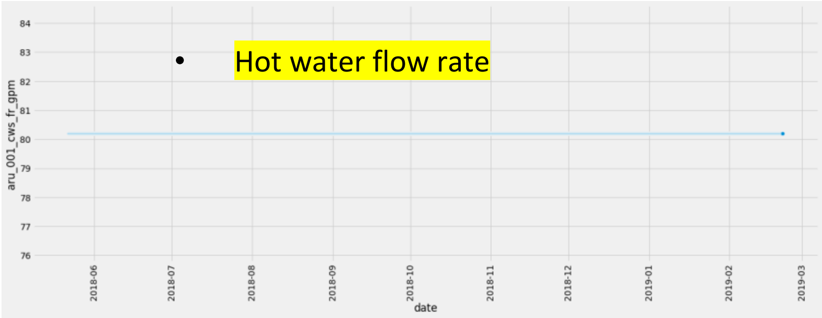
zone_temp_interior.csv
zone_temp_exterior.csv

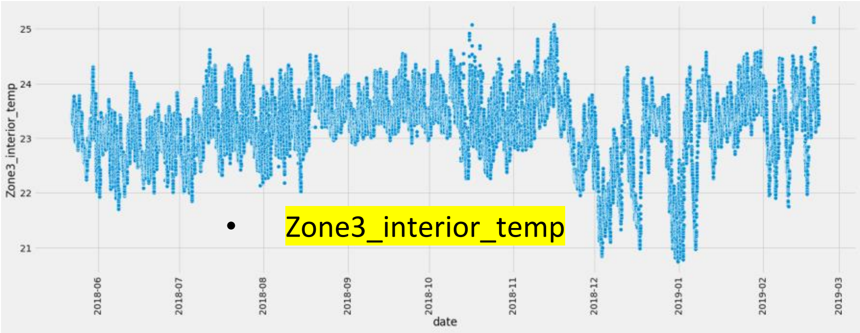
Data filtering

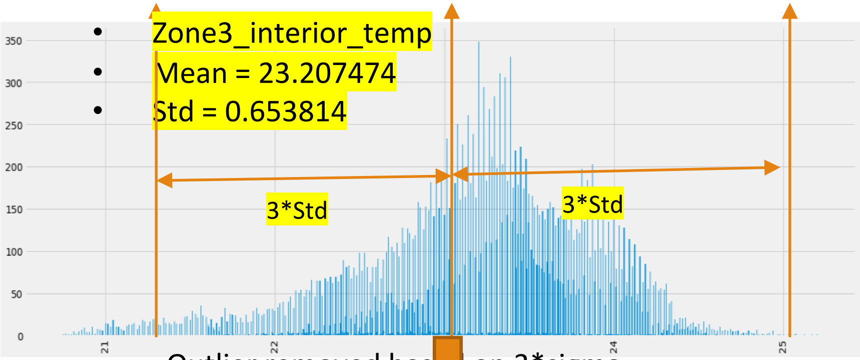


Resampling data
15 mins interval





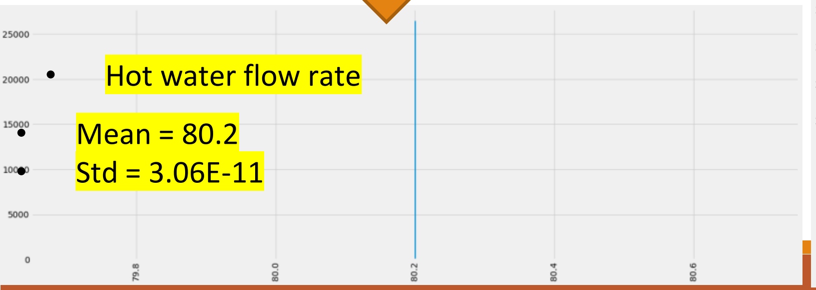


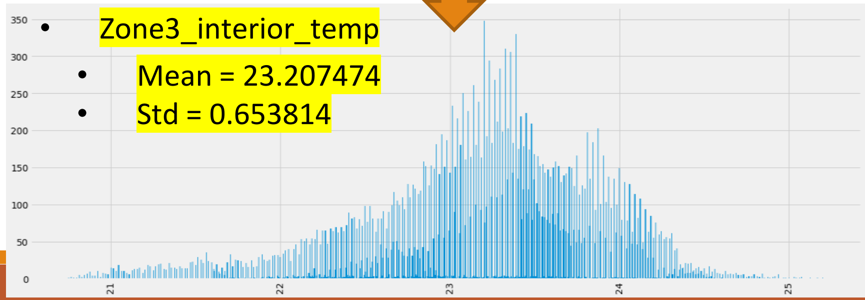


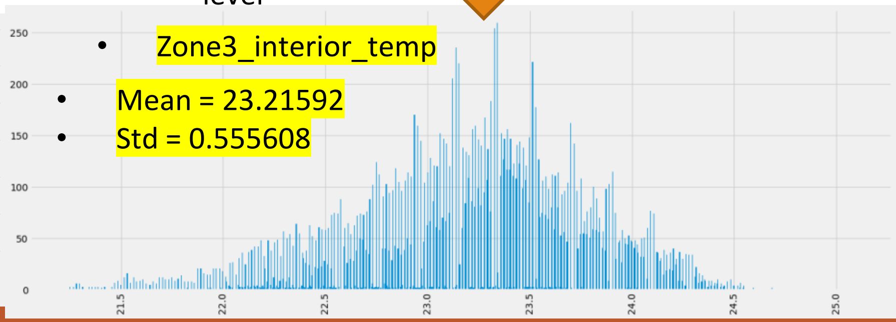
Calculate the stats and if the std is not changed much remove from the list

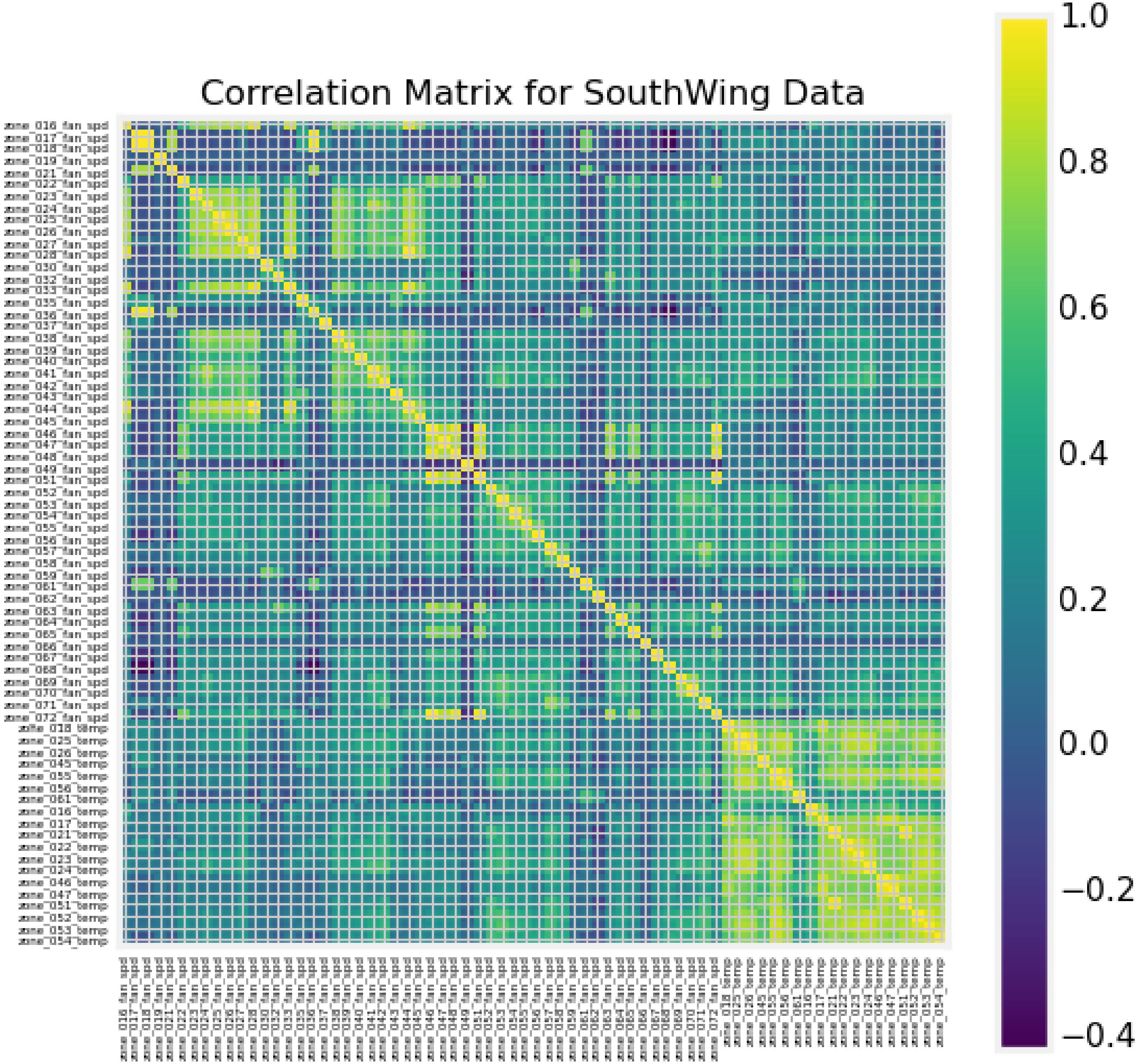
This signal has variability. And can not be removed

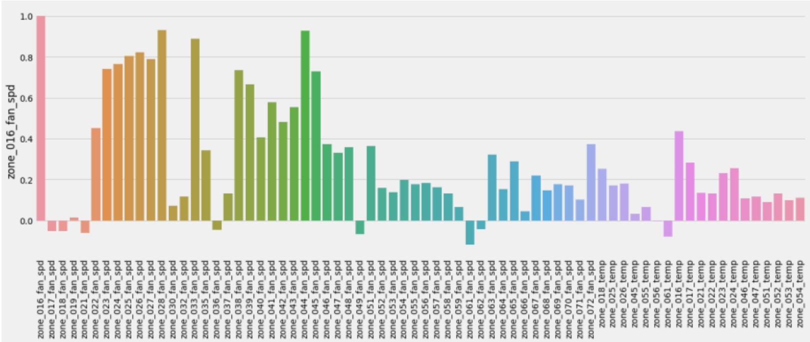
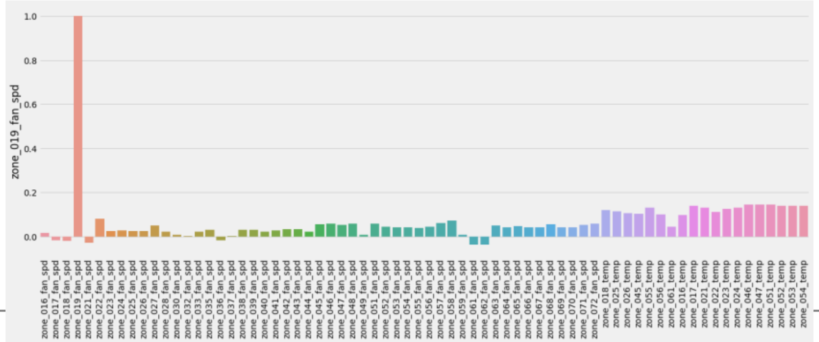
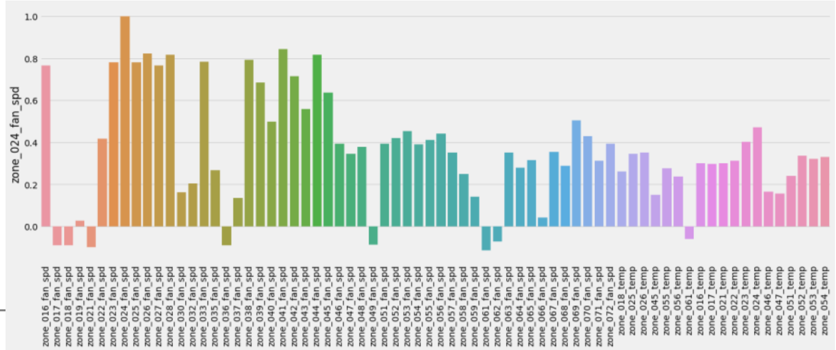
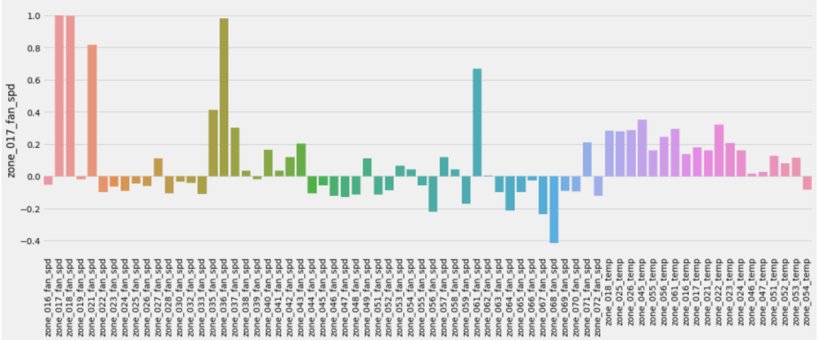
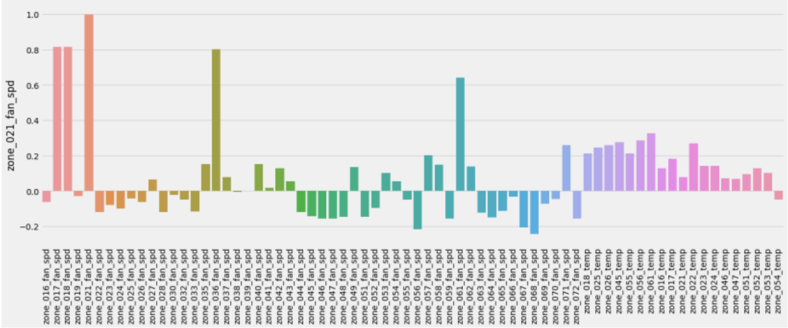
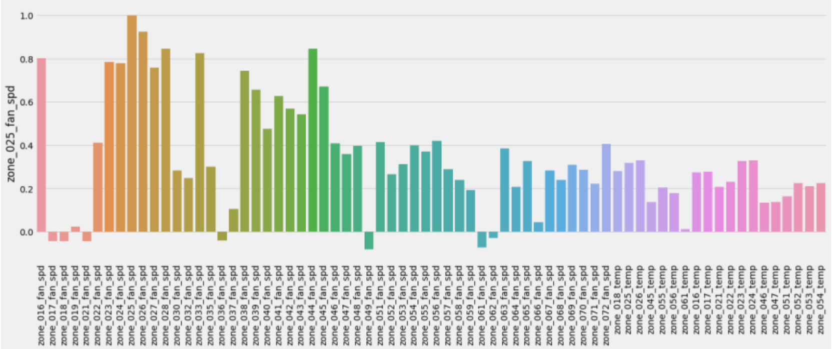
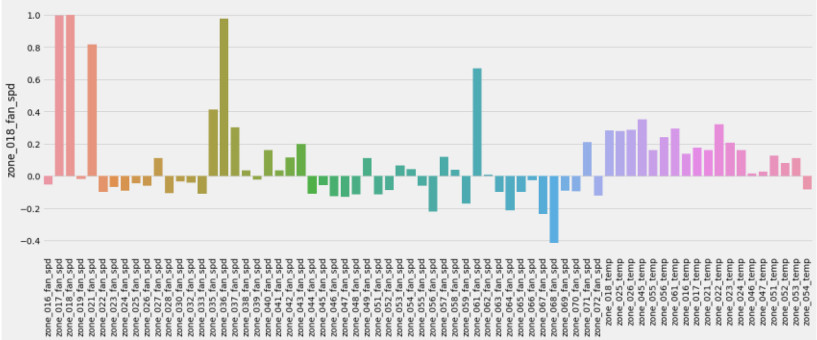
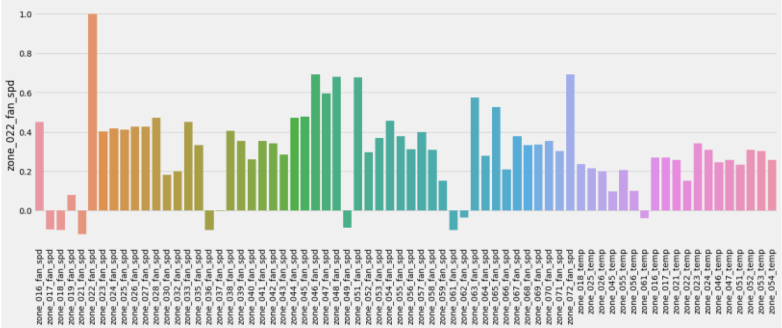
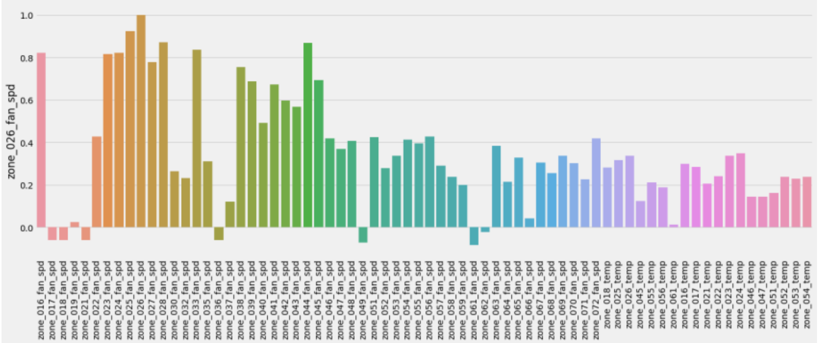
Outlier removed based on 3*sigma level

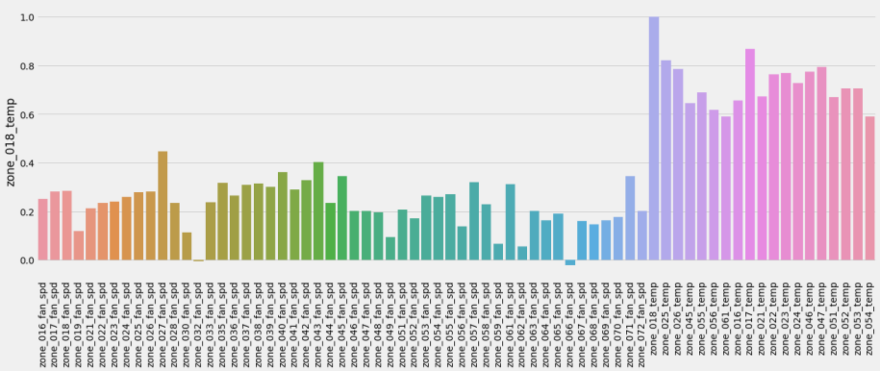


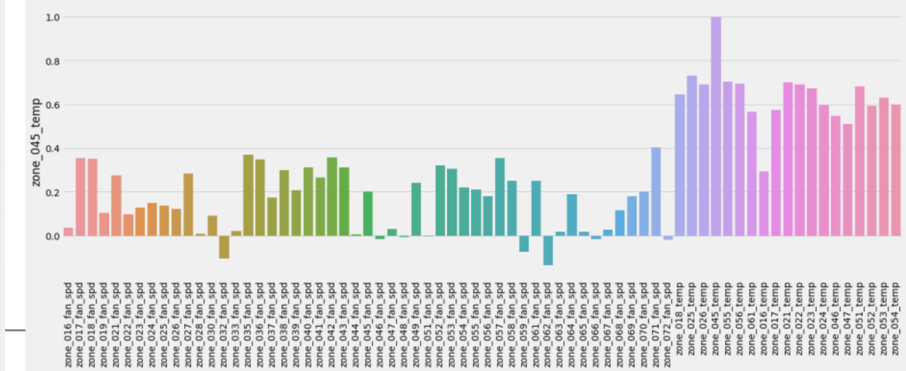


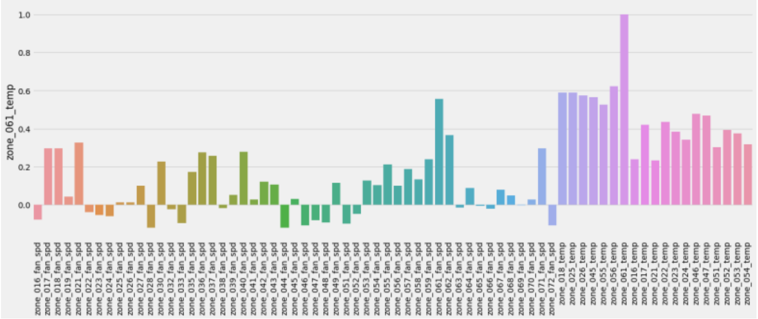


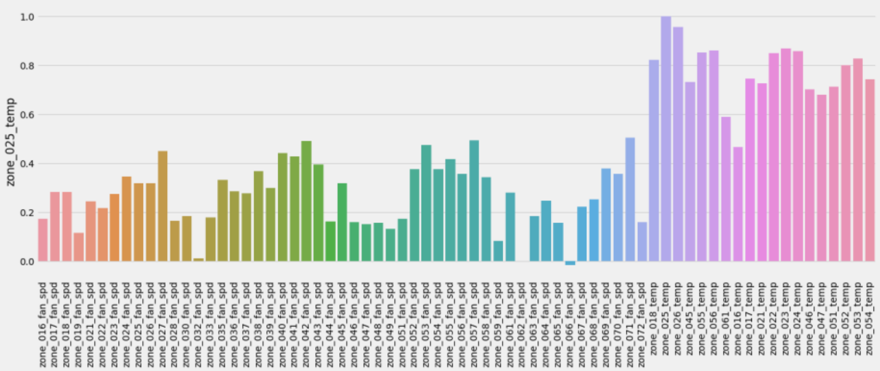

Correlation Matrix for SouthWing Data
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4

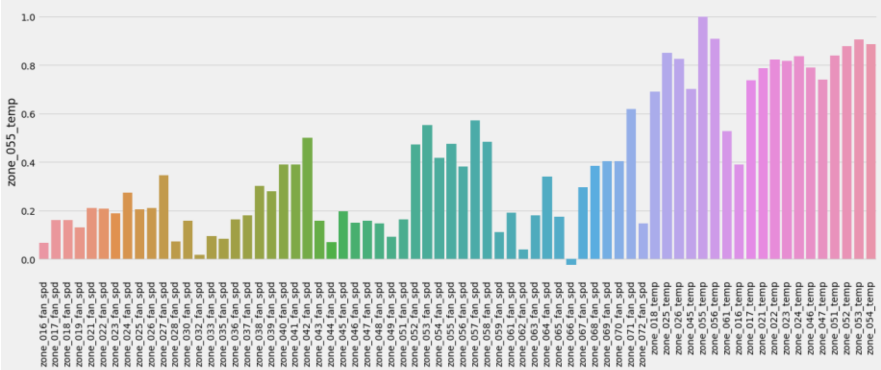

zone_016_fan_spd

zone_019_fan_spd

zone_024_fan_spd

zone_017_fan_spd

zone_021_fan_spd

zone_025_fan_spd

zone_018_fan_spd

zone_022_fan_spd

zone_026_fan_spd

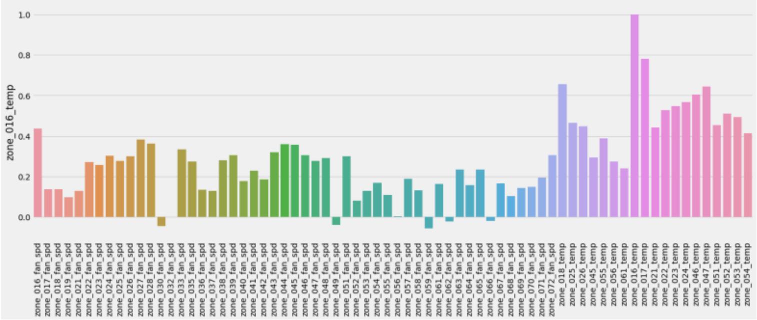

zone_018_temp

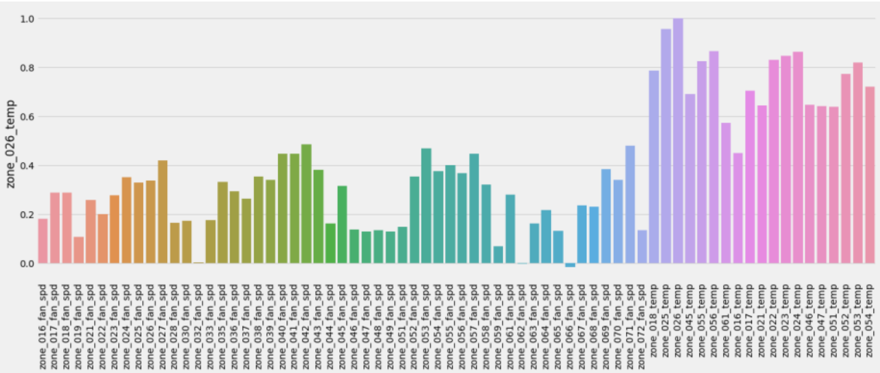

zone_045_temp

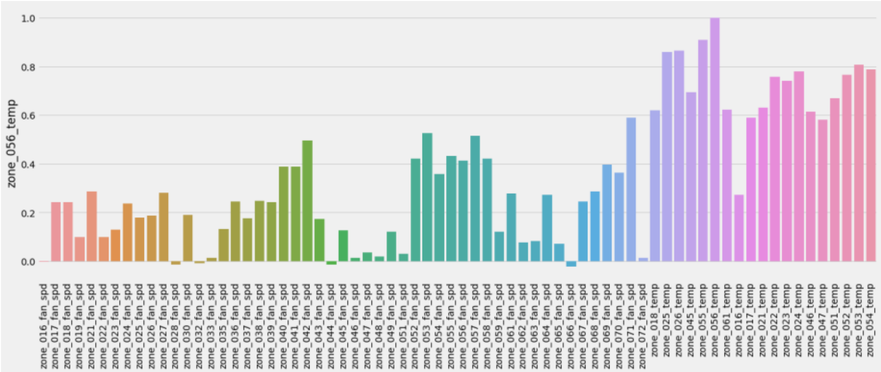

zone_061_temp

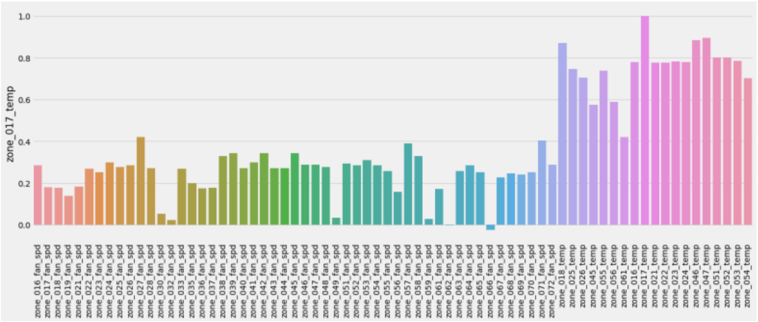

zone_025_temp


zone_055_temp

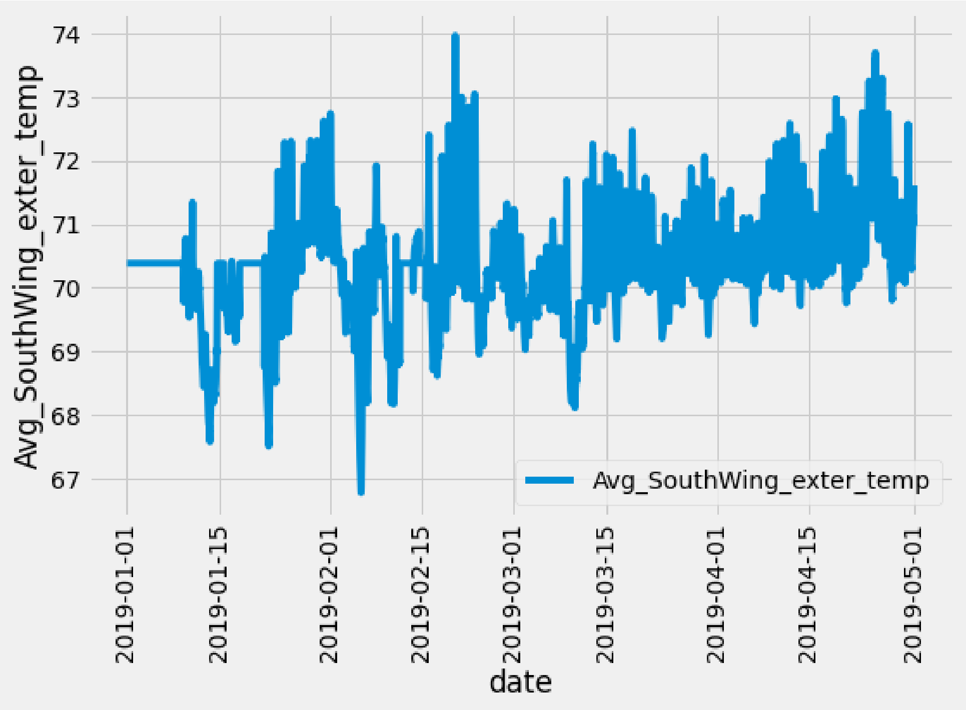

zone_016_temp

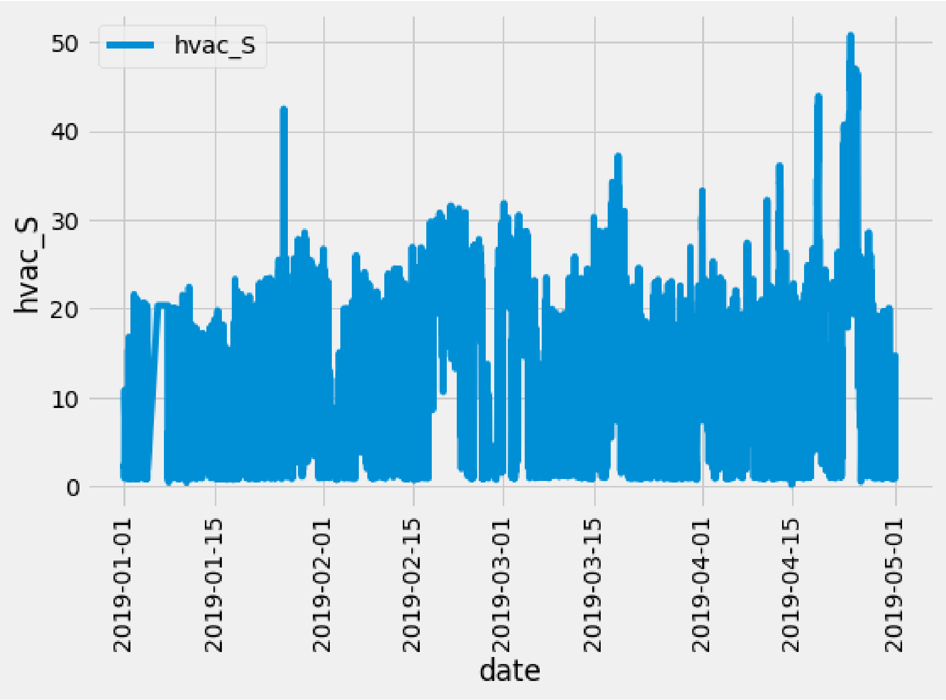

zone_026_temp


zone_056_temp


zone_017_temp

ML models: relation between Exterior temp and HVAC





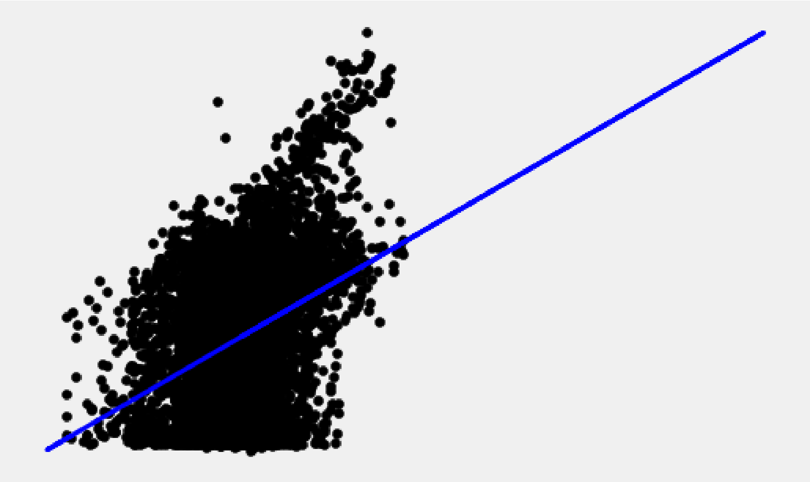
ML models: relation between Exterior temp and HVAC

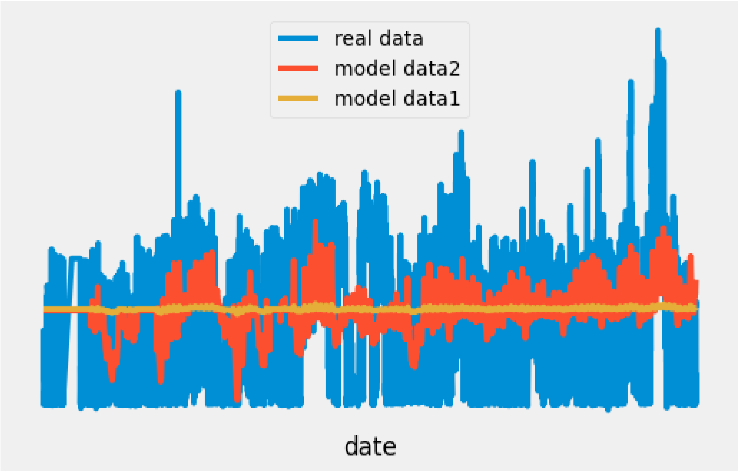
Model 1

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.110
Model:                            OLS   Adj. R-squared:                  0.110
Method:                 Least Squares   F-statistic:                     1009.
Date:                Tue, 25 Apr 2023   Prob (F-statistic):          6.09e-209
Time:                        16:01:46   Log-Likelihood:                -28880.
No. Observations:                8190   AIC:                         5.776e+04
Df Residuals:                    8188   BIC:                         5.778e+04
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const       -219.7805      7.348    -29.909      0.000    -234.185    -205.376
x1             3.3137      0.104     31.767      0.000       3.109       3.518
==============================================================================
Omnibus:                      356.748   Durbin-Watson:                   0.477
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              170.449
Skew:                           0.146   Prob(JB):                     9.72e-38
Kurtosis:                       2.356   Cond. No.                     5.70e+03
==============================================================================
```

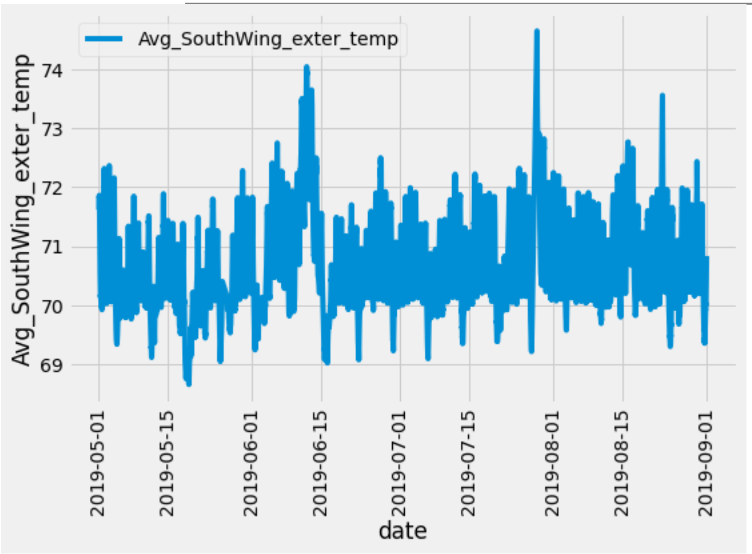
Model 2

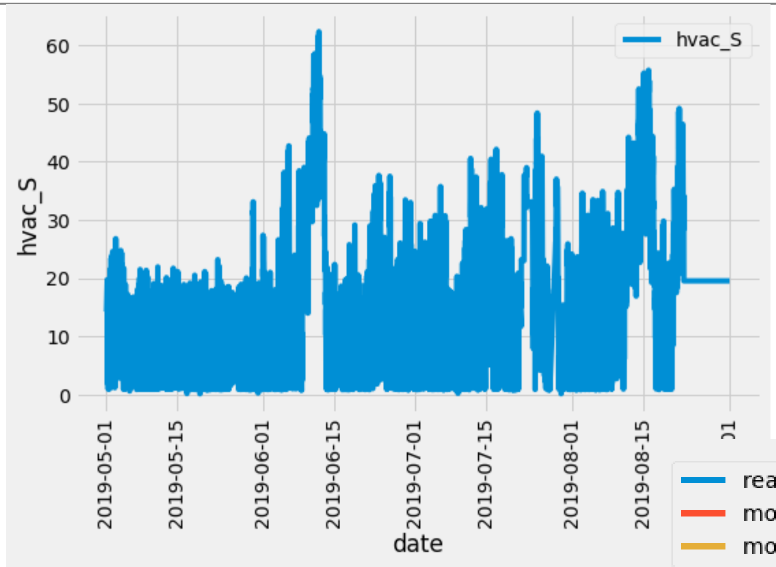
```
                                 OLS Regression Results
=======================================================================================
Dep. Variable:                      y   R-squared (uncentered):                   0.713
Model:                            OLS   Adj. R-squared (uncentered):              0.713
Method:                 Least Squares   F-statistic:                          2.039e+04
Date:                Tue, 25 Apr 2023   Prob (F-statistic):                        0.00
Time:                        16:01:46   Log-Likelihood:                         -29305.
No. Observations:                8190   AIC:                                  5.861e+04
Df Residuals:                    8189   BIC:                                  5.862e+04
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
x1             0.1941      0.001    142.788      0.000       0.191       0.197
==============================================================================
Omnibus:                      201.156   Durbin-Watson:                   0.430
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              215.530
Skew:                           0.395   Prob(JB):                     1.58e-47
Kurtosis:                       2.919   Cond. No.                         1.00
==============================================================================

Notes:
[1] R² is computed without centering (uncentered) since the model does not contain a constant.
[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

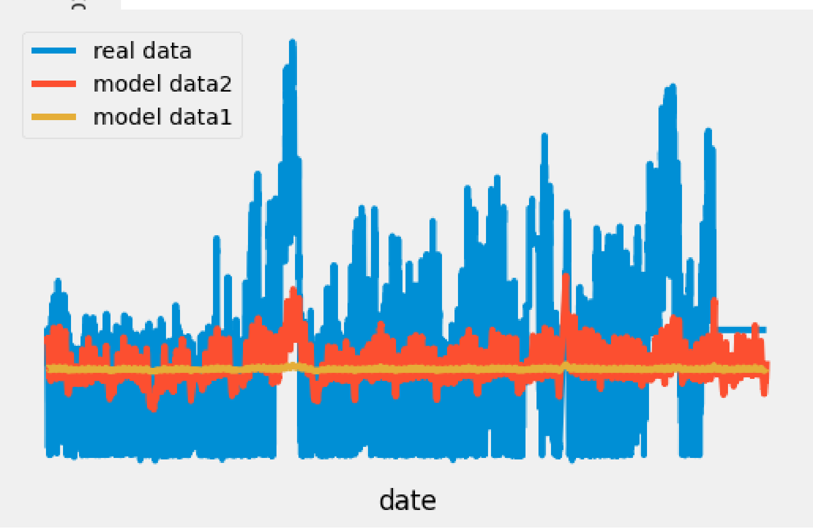




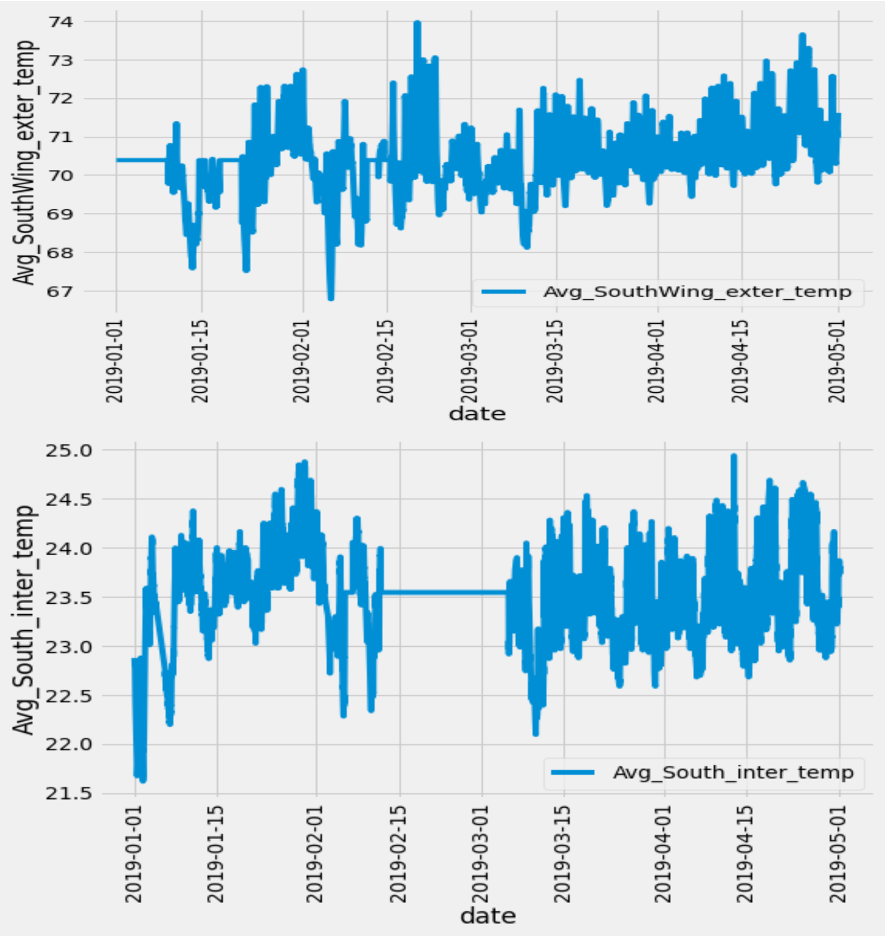
ML models: relation between Exterior temp and HVAC

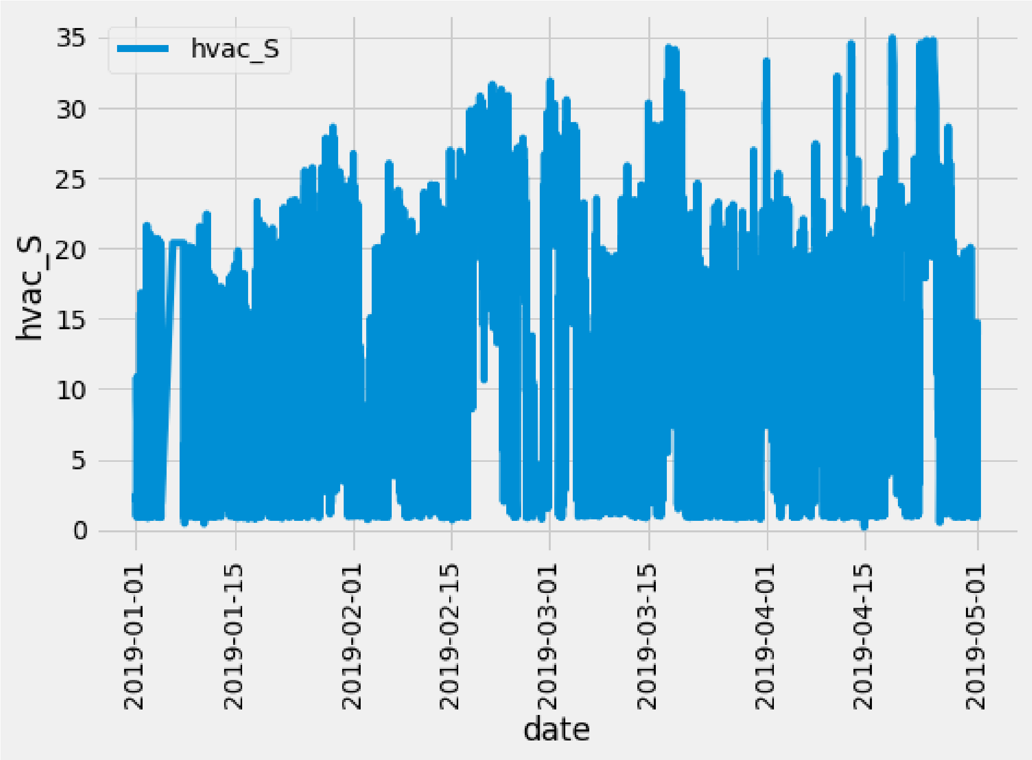






ML models: relation between (Exterior temp, interior temp) and HVAC

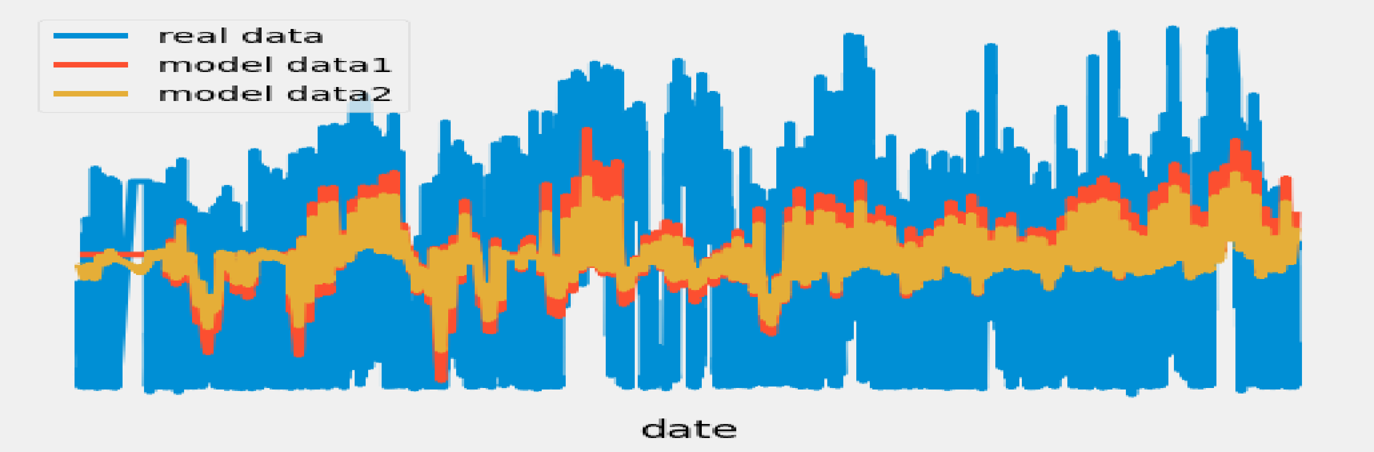




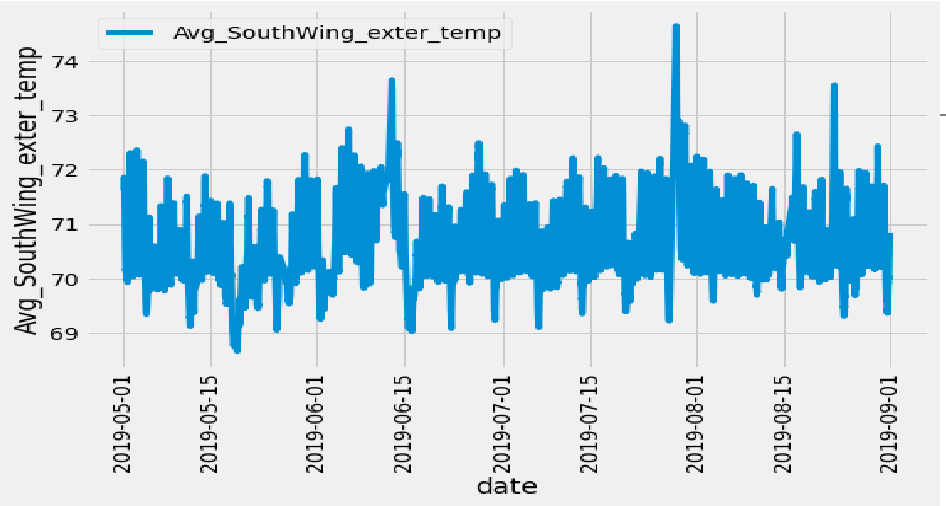
ML models: relation between (Exterior temp, interior temp) and HVAC

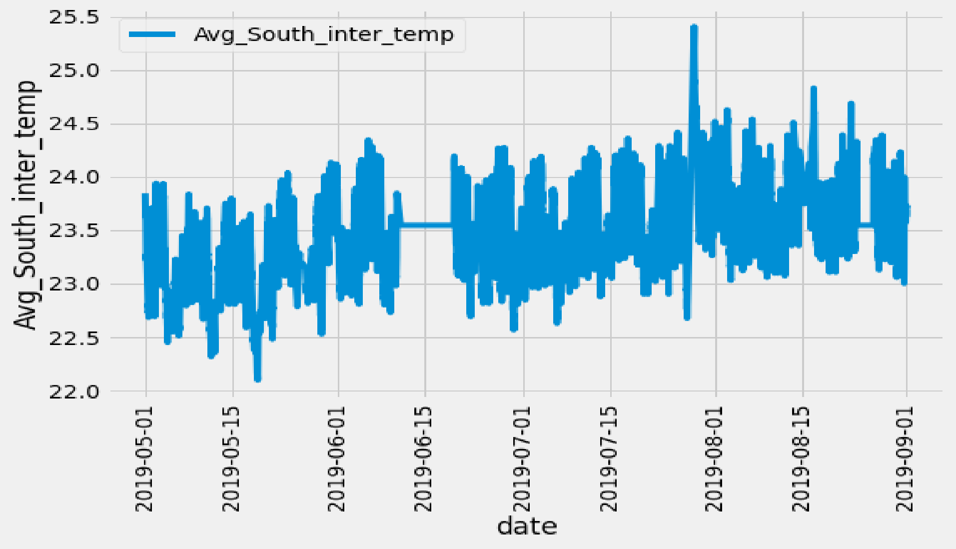
```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.065
Model:                            OLS   Adj. R-squared:                  0.064
Method:                 Least Squares   F-statistic:                     278.5
Date:                Tue, 25 Apr 2023   Prob (F-statistic):          1.09e-117
Time:                        17:30:03   Log-Likelihood:                -28057.
No. Observations:                8066   AIC:                         5.612e+04
Df Residuals:                    8063   BIC:                         5.614e+04
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const       -160.0721      7.372    -21.713      0.000    -174.524    -145.621
x1             2.1107      0.126     16.808      0.000       1.865       2.357
x2             1.0487      0.240      4.369      0.000       0.578       1.519
==============================================================================
Omnibus:                     2112.553   Durbin-Watson:                   0.529
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              353.473
Skew:                           0.002   Prob(JB):                     1.76e-77
Kurtosis:                       1.974   Cond. No.                     6.27e+03
==============================================================================
```

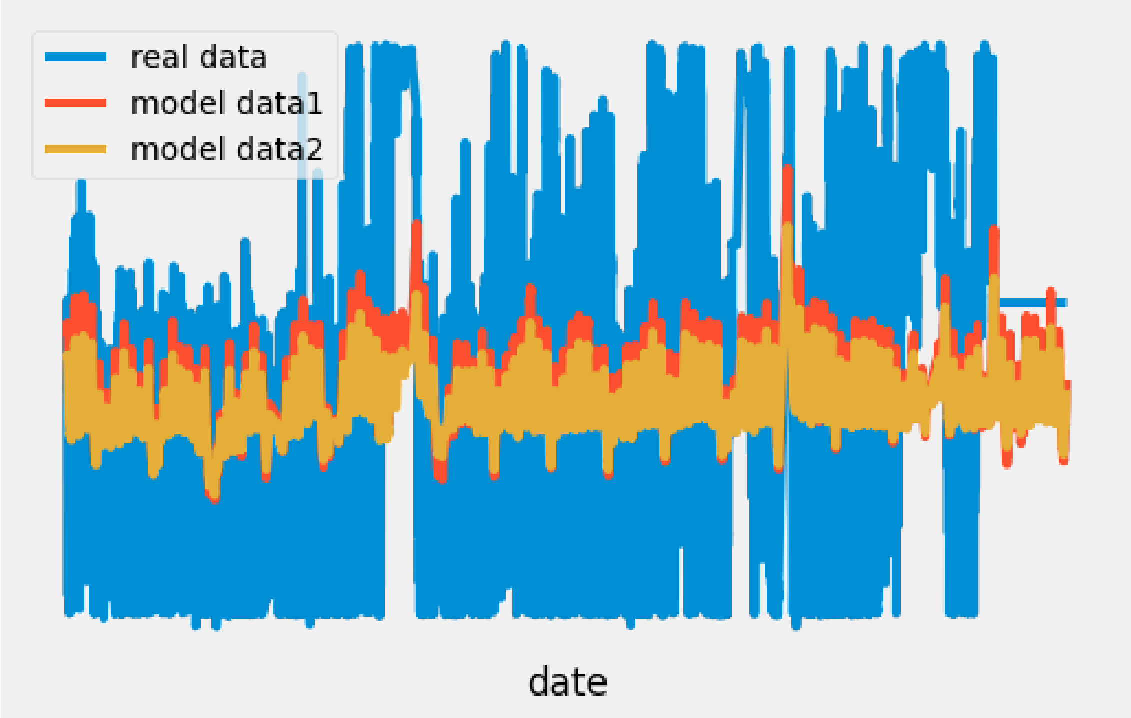
```
                                 OLS Regression Results
=======================================================================================
Dep. Variable:                      y   R-squared (uncentered):                   0.730
Model:                            OLS   Adj. R-squared (uncentered):              0.729
Method:                 Least Squares   F-statistic:                          1.088e+04
Date:                Tue, 25 Apr 2023   Prob (F-statistic):                        0.00
Time:                        17:30:03   Log-Likelihood:                         -28286.
No. Observations:                8066   AIC:                                  5.658e+04
Df Residuals:                    8064   BIC:                                  5.659e+04
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
x1             0.0068      0.082      0.083      0.934      -0.154       0.168
x2             0.5424      0.246      2.207      0.027       0.061       1.024
==============================================================================
Omnibus:                     2588.315   Durbin-Watson:                   0.500
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              384.616
Skew:                           0.070   Prob(JB):                     3.03e-84
Kurtosis:                       1.939   Cond. No.                         214.
==============================================================================
```



ML models: relation between (Exterior temp, interior temp) and HVAC

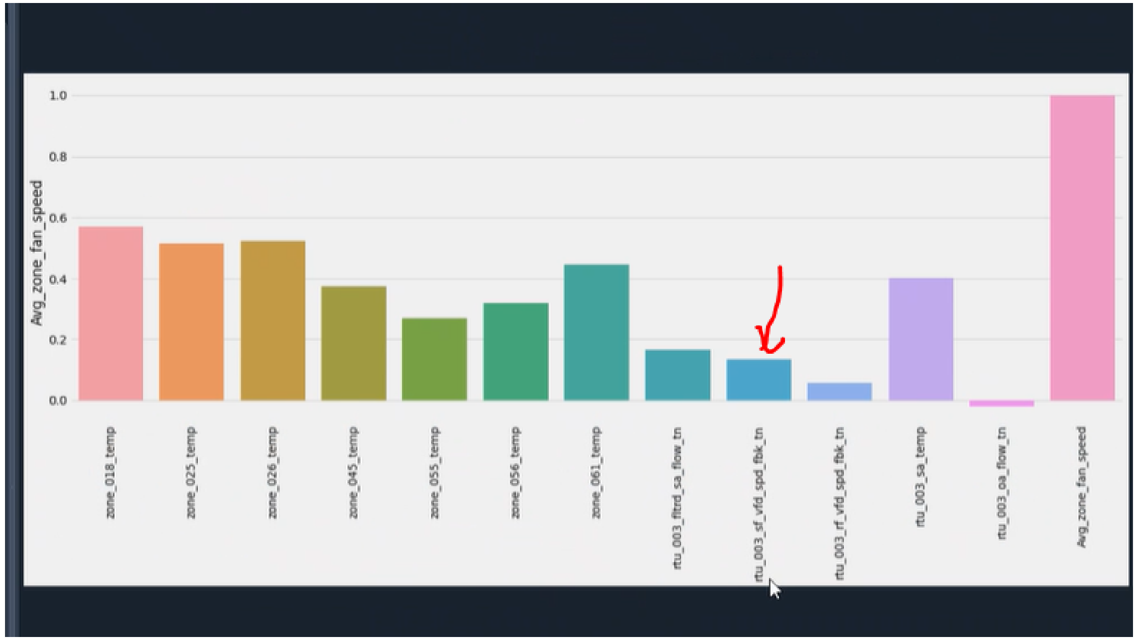


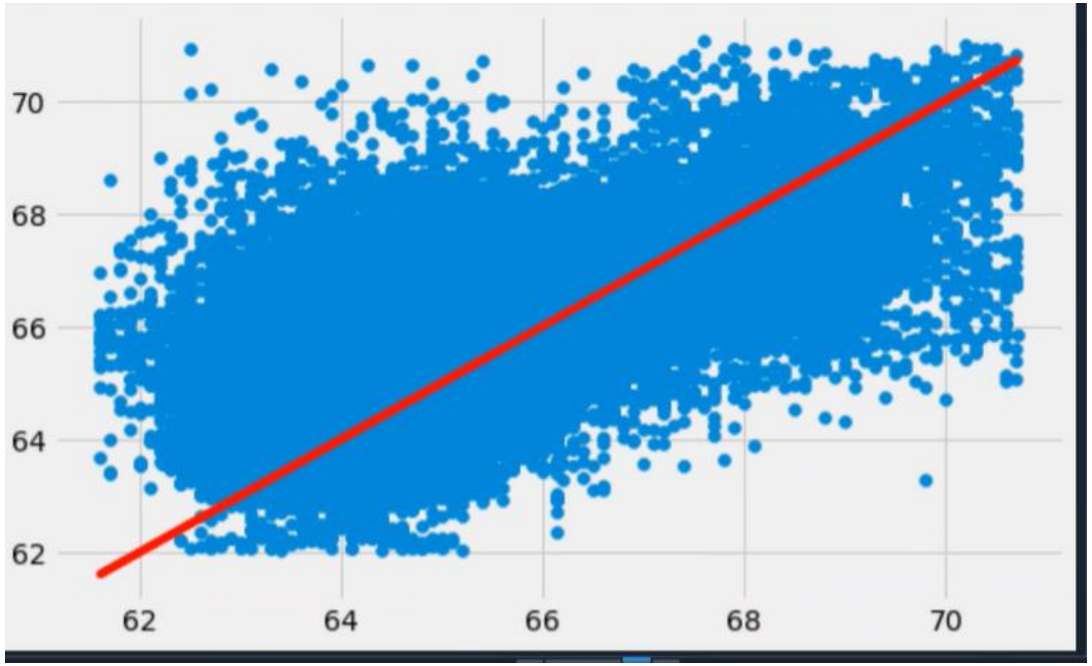




Extension of ML models

Performance of the ML model used

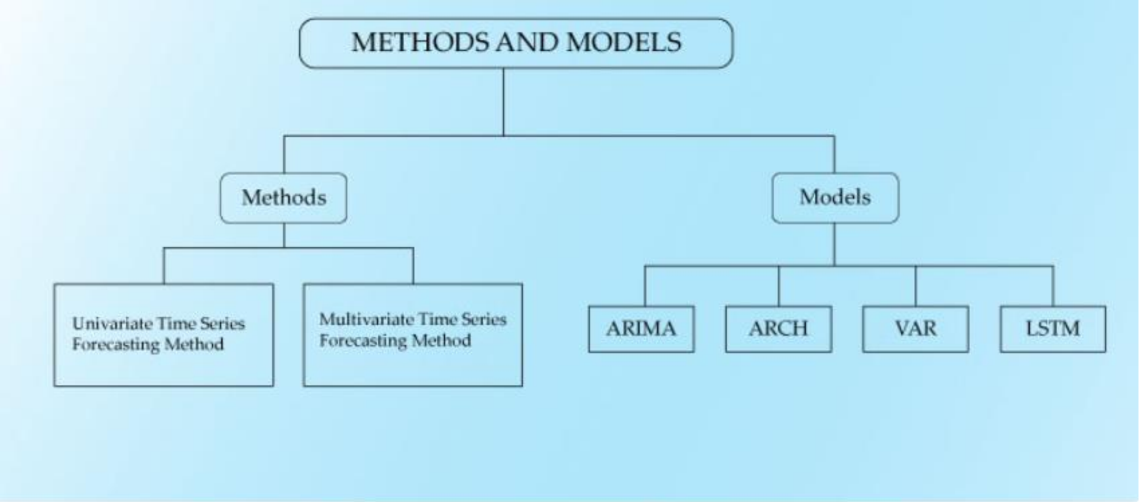




Improvements required to the ML model

- Time dependency is an important factor for learning time-series data characteristics.

- **Suitable models for forecasting time series data:**
  - Autoregressive Integrated Moving Average (**ARIMA**): It's a statistical model for time series data analysis or predicting future trends
  - Moving Average: Uses previous forecast error rather than directly using previous values as in regression.
  - Exponential Smoothening: forecasting method only for univariate data
  - Long Short-Term Memory (**LSTM**) it's a variety of Deep neural network model



Time-Series Forecasting: Methods and Models in Machine Learning

References

1. Sensorless Air Flow Control in an HVAC System through Deep Learning, Junseo Son et. al. Appl. Sci. 2019, 9, 3293
2. Hong, Tianzhen; Luo, Na; Blum, David; Wang, Zhe (2022), A three-year building operational performance dataset for informing energy efficiency , Dryad, Dataset, https://doi.org/10.7941/D1N33Q
3. https://www.analyticssteps.com/blogs/introduction-time-series-analysis-time-series-forecasting-machine-learning-methods-models
4. https://builtin.com/data-science/time-series-forecasting-python

Thank you

Ayan Nath: ayannath23@gmail.com
Ravikiran Anapagaddi: rkiranmat@gmail.com
Rajat Nandan NS: rajat.nandan.ns@gmail.com

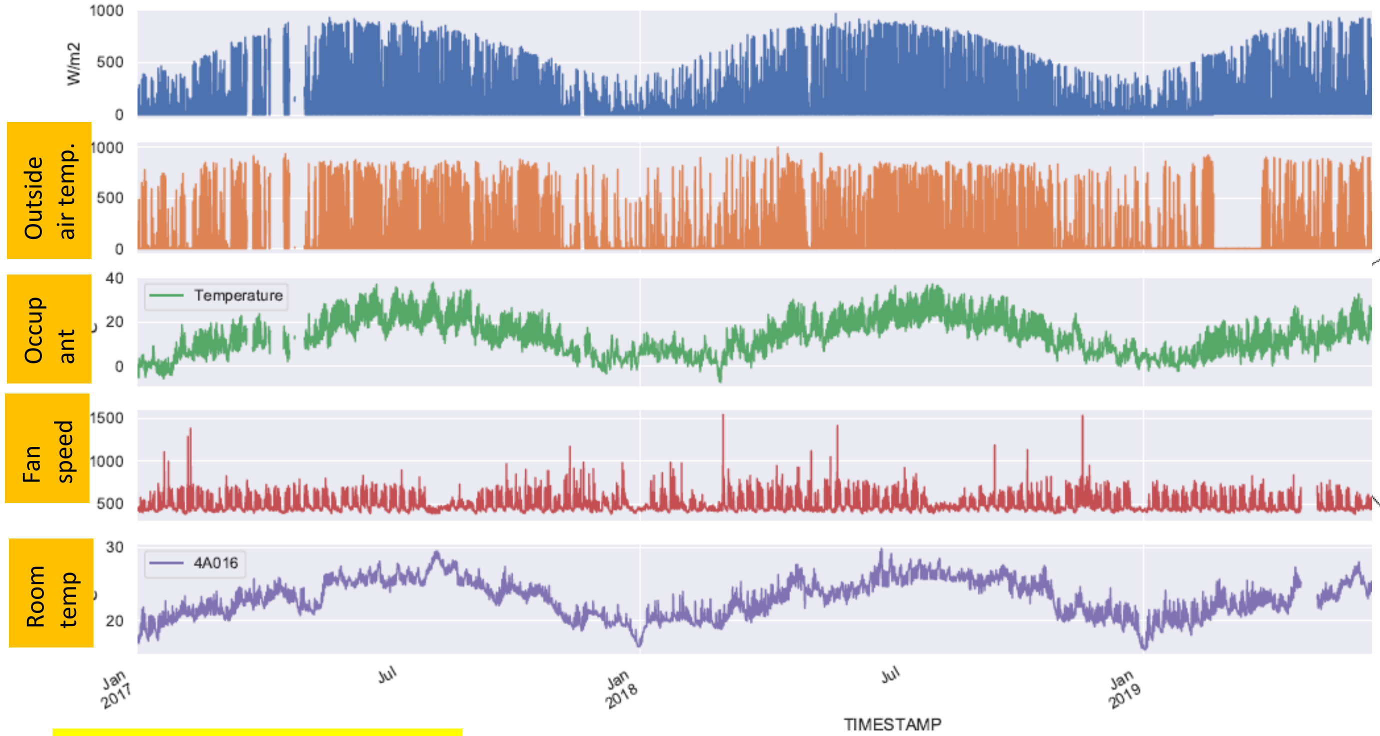
Relate Zone temp (18,25,..) HVAC-3 flow rate

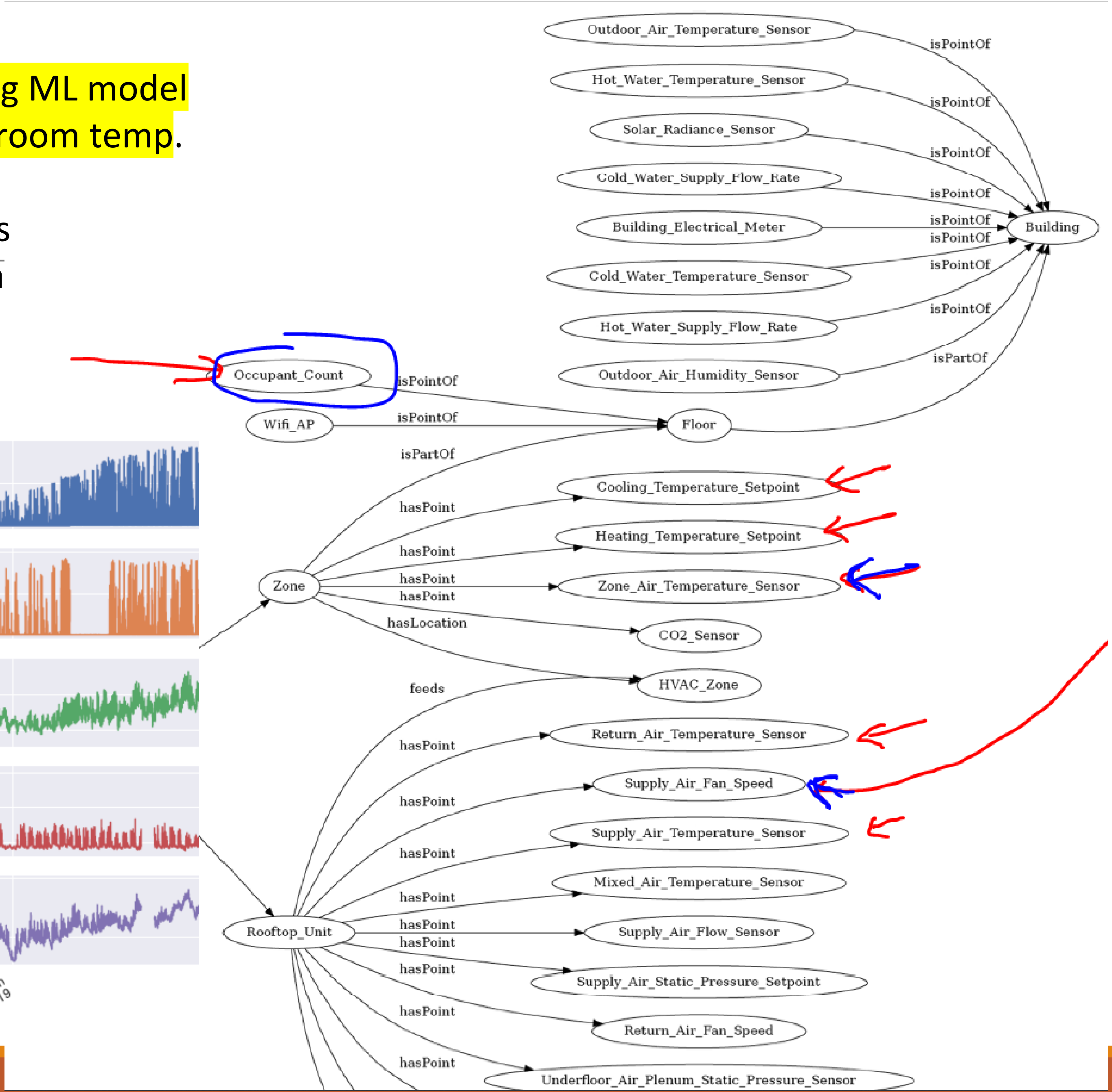
South Wing	3	18, 25, 26, 45, 48, 55, 56, 61

Target: Predict variation in fan speed forecasting using ML model

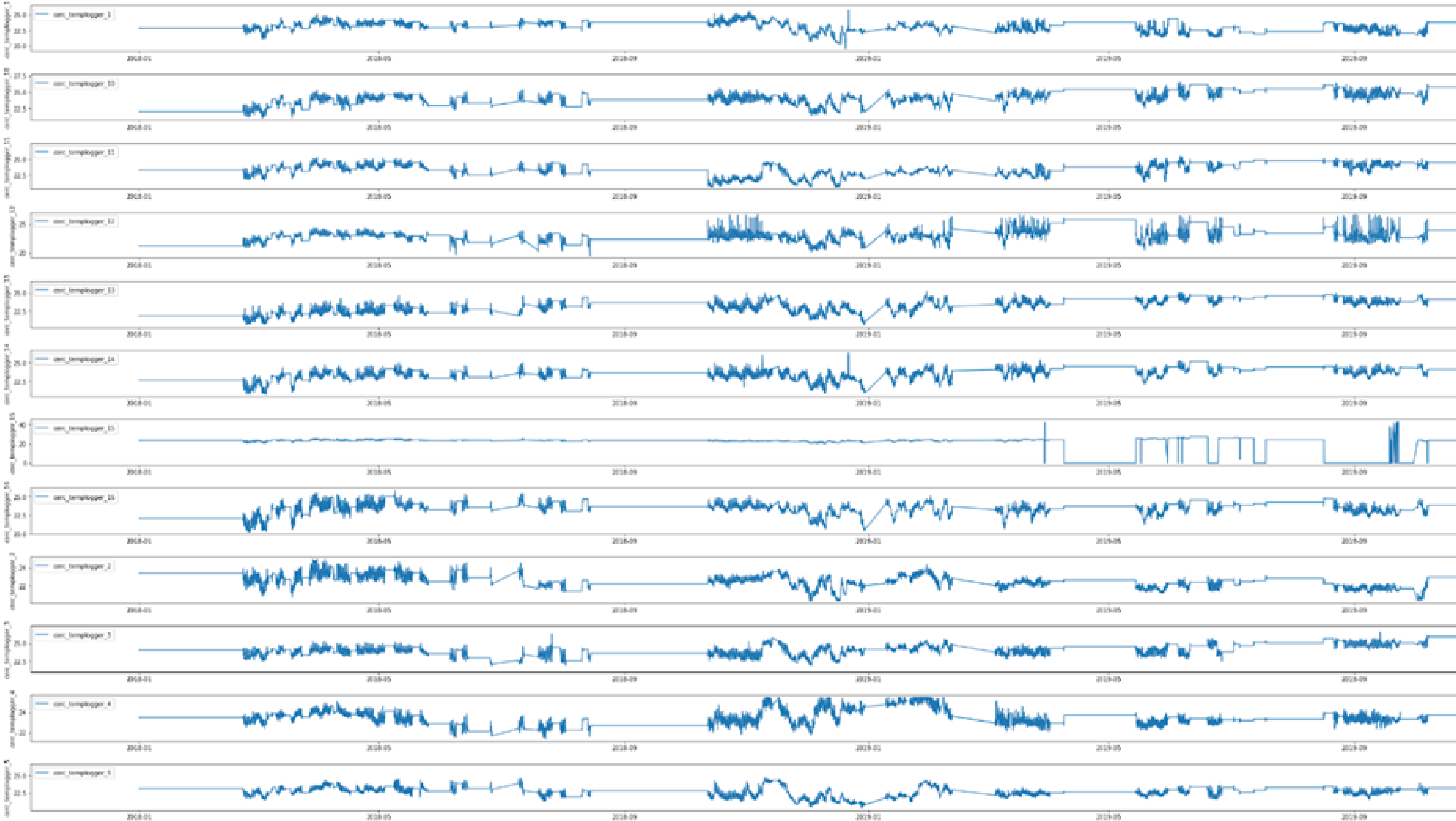
- Relate fan speed with ~~building occupancy &~~ room temp.

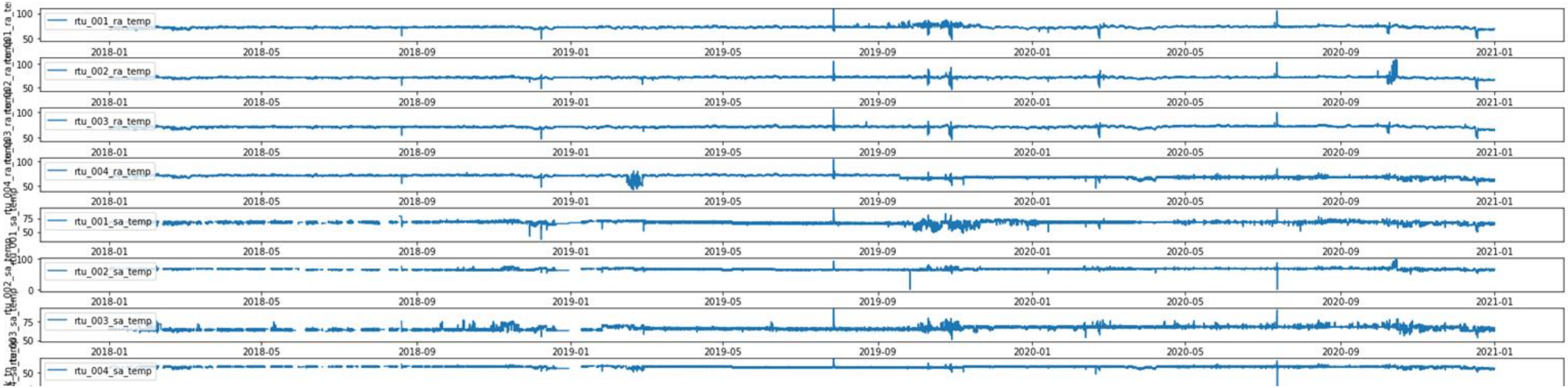
1. Activity: Plotting – relating different variables
   - Average out to remove the missing data
2. ML model :

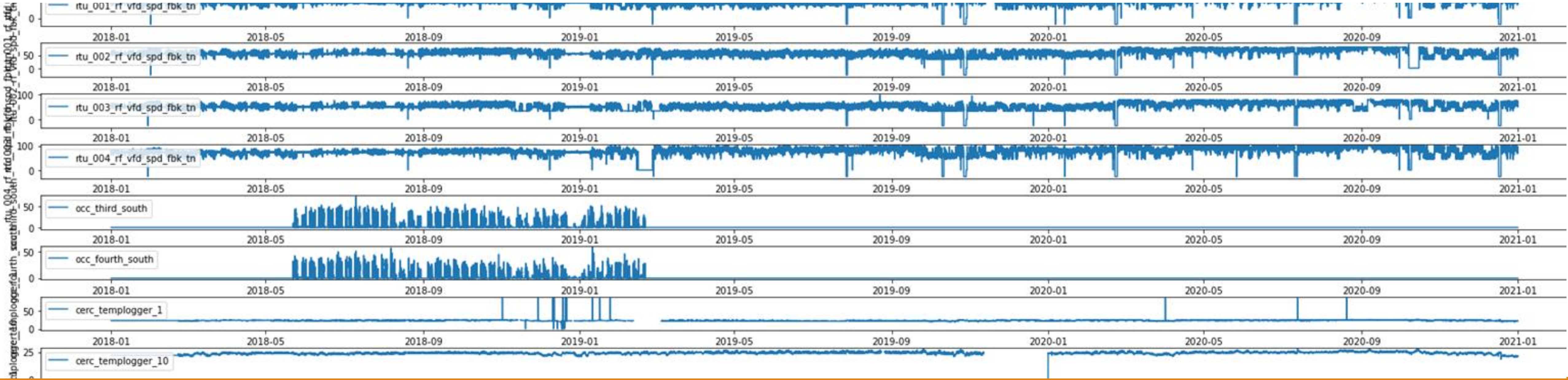


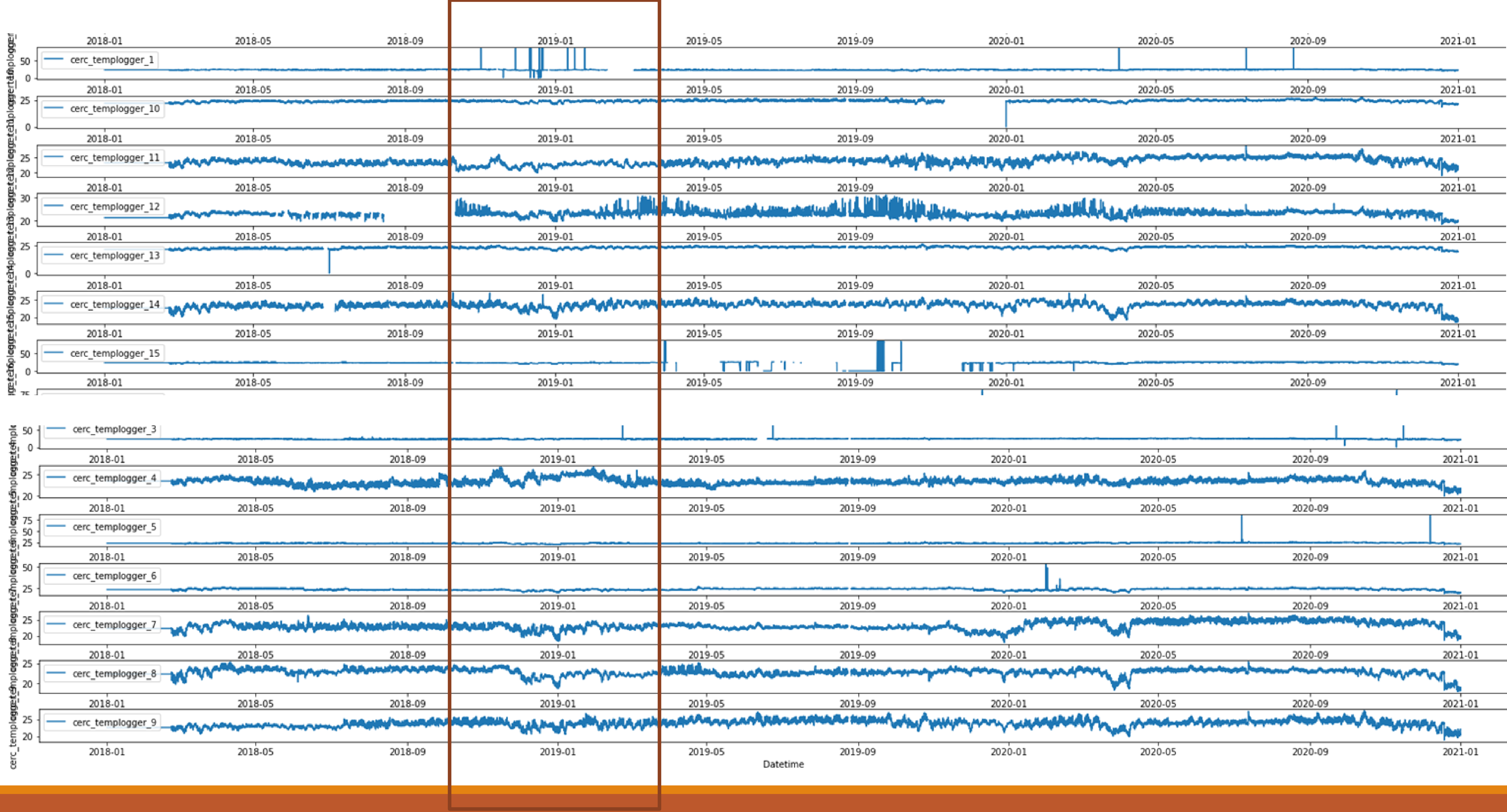


Multi-variable model




rtu_001_ra_temp
rtu_002_ra_temp
rtu_003_ra_temp
rtu_004_ra_temp
rtu_001_sa_temp
rtu_002_sa_temp
rtu_003_sa_temp
rtu_004_sa_temp
2018-01
2018-05
2018-09
2019-01
2019-05
2019-09
2020-01
2020-05
2020-09
2021-01


rtu_001_rf_vfd_spd_fbk_tn
rtu_002_rf_vfd_spd_fbk_tn
rtu_003_rf_vfd_spd_fbk_tn
rtu_004_rf_vfd_spd_fbk_tn
occ_third_south
occ_fourth_south
cerc_templogger_1
cerc_templogger_10
2018-01
2018-05
2018-09
2019-01
2019-05
2019-09
2020-01
2020-05
2020-09
2021-01



cerc_templogger_1
cerc_templogger_10
cerc_templogger_11
cerc_templogger_12
cerc_templogger_13
cerc_templogger_14
cerc_templogger_15
cerc_templogger_3
cerc_templogger_4
cerc_templogger_5
cerc_templogger_6
cerc_templogger_7
cerc_templogger_8
cerc_templogger_9
2018-01
2018-05
2018-09
2019-01
2019-05
2019-09
2020-01
2020-05
2020-09
2021-01
Datetime

TA

Thivin Anandh 05-04 20:04 Edited

Project Presentation

CCE-AIML-Jan2023

Total Duration : **12 min per presentation** (10 min presentation - 2 min Q&A)

Date of Presentation : **22-Apr : 6:00 pm to 8:00 pm**

Its mandatory for all people involved in the project to be present during the presentation session.

- Please keep any introductions to the project as minimum as possible (what ever was discussed during previous presentation)
- Directly start with the progress (Data cleanup, ML model description and results)

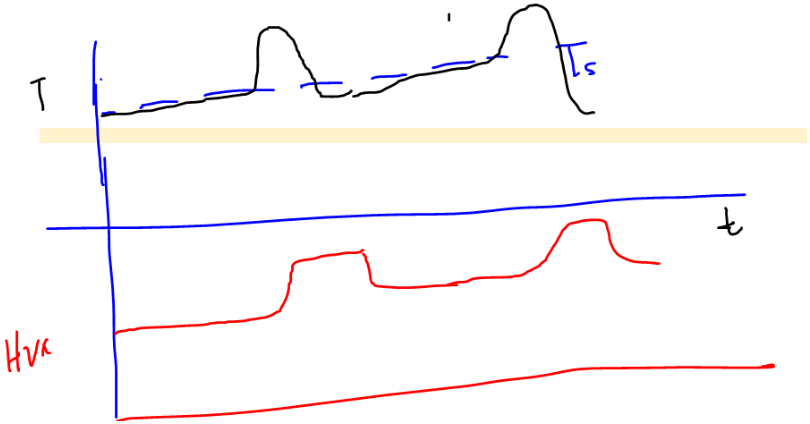
Deliverables:

- Fully working version of the code on github with proper readme file
- a 3 page IEEE format report of your project
- Presentation in PDF form (also uploaded to github, if there is no confidential information on the slides , else kindly ignore)

You can download the report template at this URL : https://www.ieee.org/conferences/publishing/templates.html

Thanks,
Thivin D

See less



Data	File name	Column name	Description	Number of data points	Unit	Sampling frequency	Missing rate of the raw data (2018 - 2020)	Specific available time period (default is 3 years from 2018 to 2020)
HVAC operational data		hp_hws_temp.csv	Heat pump heating water supply temperature	1	°F	1 min	0.14	
		Roof Top Unit	Roof Top Unit * supply air temperature setpoint (*: 001, 002, 003, 004)	4	°F	1 min	0.15	
			Roof Top Unit * supply air temperature (*: 001, 002, 003, 004)	4	°F	1 min	0.14	
			Roof Top Unit * return air temperature (*: 001, 002, 003, 004)	4	°F	1 min	0.14	
			Roof Top Unit * mixed air temperature (*: 001, 002, 003, 004)	4	°F	1 min	0.14	
			Roof Top Unit * outdoor air temperature (*: 001, 002, 003, 004)	4	°F	1 min	0.14	
			Roof Top Unit * filtered supply air flow rate (*: 001, 002, 003, 004)	4	CFM	1 min	0.14	
			Roof Top Unit * outdoor air flow rate (*: 001, 002, 003, 004)	4	CFM	1 min	0.02	Apr -- Dec 2020
			Roof Top Unit * outdoor air damper position (*: 001, 002, 003, 004)	4	%	1 min	0.15	
			Roof Top Unit * economizer setpoint (*: 001, 002, 003, 004)	4	°F	1 min	0.14	
			Roof Top Unit * air pressure static setpoint (*: 001, 002, 003, 004)	4	psi	1 min	0.15	
			Roof Top Unit * plenum air pressure at floor ** (*: 001, 002, 003, 004; **: gnd_lvl, lvl2)	8	psi	1 min	0.14	
			Roof Top Unit * supply fan speed (*: 001, 002, 003, 004)	4	%	1 min	0.14	
			Roof Top Unit * return fan speed (*: 001, 002, 003, 004)	4	%	1 min	0.14	
			Heat meter for air source heat pump	1	mbtuph	5 min	0.29	Aug -- Dec 2020
		Chilled water	Chilled water supply temperature	1	°F	5 min	0.01	
			Chilled water return temperature	1	°F	5 min	0.01	
			Chilled water fow rate	1	CFM	5 min	0.02	
		Hot water	Hot water supply temperature	1	°F	5 min	0.16	Oct 2019 -- Dec 2020
			Hot water return temperature	1	°F	5 min	0.16	
			Hot water fow rate	1	CFM	5 min	0.01	
		Supply air fan speed	Supply air fan speed of Zone *	44	%	1 min	0.15-0.23	
			Heating water valve position of Zone *	51	%	1 min	0.15-0.25	

Time Series Analysis and Forecasting with Python – single variable

https://towardsdatascience.com/an-end-to-end-project-on-time-series-analysis-and-forecasting-with-python-4835e6bf050b

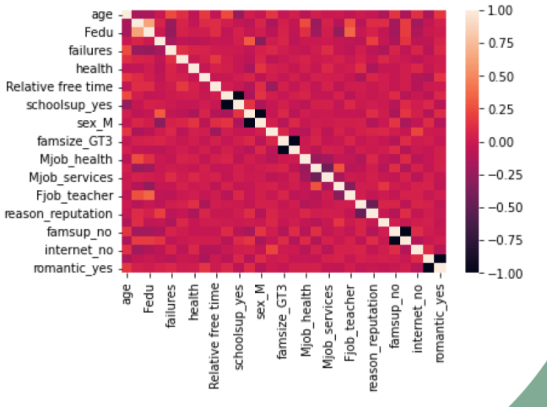
Multivariate Multi-Step Time Series Forecasting Models

https://machinelearningmastery.com/how-to-develop-machine-learning-models-for-multivariate-multi-step-air-pollution-time-series-forecasting/

Sample slides:

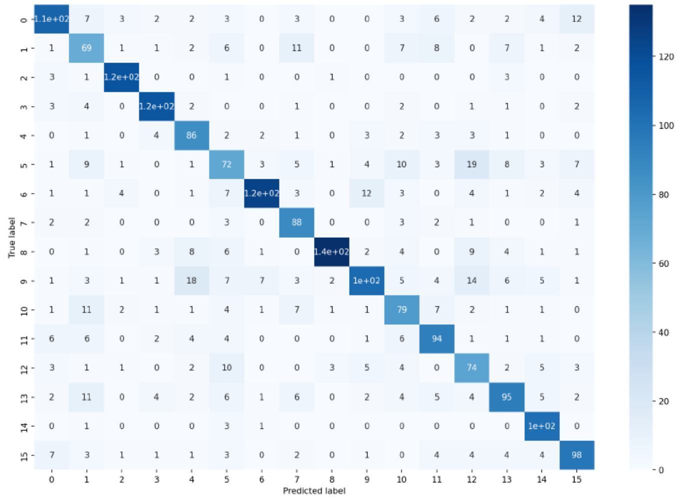
Feature selection using correlation

- For The features that had 0.95 or more correlation, only one such feature was kept. The rest were dropped
- Absolute values of correlation were used
- Dropped features:
- ['schoolsup_yes', 'famsup_yes', 'paid_yes', 'higher_yes', 'romantic_yes']



Confusion Matrix

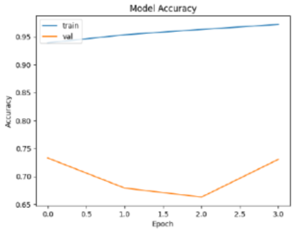
	precision	recall	f1-score	support
0	0.69	0.78	0.73	139
1	0.59	0.53	0.56	131
2	0.93	0.89	0.91	131
3	0.88	0.86	0.87	134
4	0.80	0.66	0.72	130
5	0.49	0.53	0.51	137
6	0.74	0.89	0.81	141
7	0.86	0.68	0.76	130
8	0.77	0.94	0.85	143
9	0.57	0.77	0.66	135
10	0.66	0.60	0.63	132
11	0.75	0.69	0.72	136
12	0.65	0.54	0.59	138
13	0.64	0.70	0.67	136
14	0.95	0.76	0.85	134
15	0.74	0.74	0.74	133
accuracy			0.72	2160
macro avg	0.73	0.72	0.72	2160
weighted avg	0.73	0.72	0.72	2160

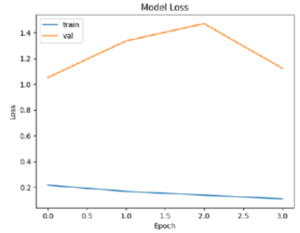


Result

- Training was done in 4 epochs by adopting early stop.
- Learning rate was updated once after the 3rd epoch.
- Train accuracy: 0.9718
- Validation accuracy: 0.7306
- Test accuracy: 0.7226851851851852

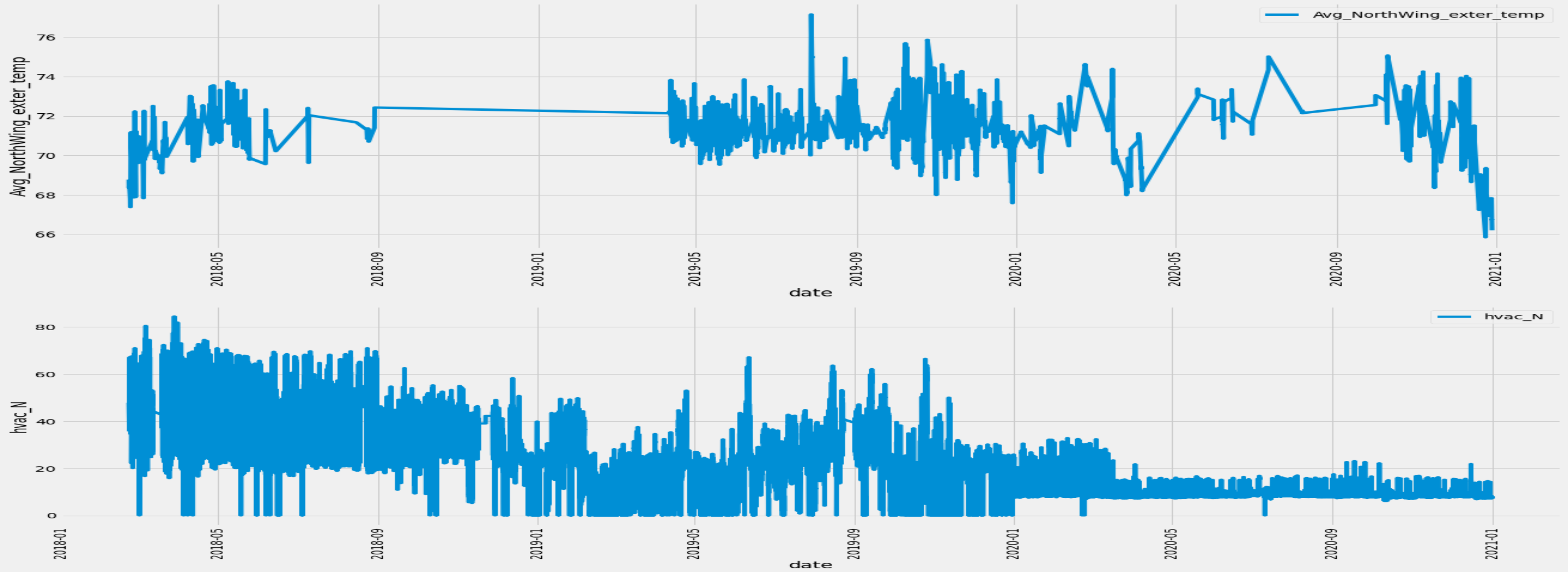
```
Epoch 1/20
191/191 [==============================] - ETA: 0s - loss: 0.2170 - accuracy: 0.9388WARNING:tensorflow:Can save best model only with val_acc available, skipp
191/191 [==============================] - 338s 2s/step - loss: 0.2170 - accuracy: 0.9388 - val_loss: 1.0519 - val_accuracy: 0.7331 - lr: 2.0000e-04
Epoch 2/20
191/191 [==============================] - ETA: 0s - loss: 0.1692 - accuracy: 0.9535WARNING:tensorflow:Can save best model only with val_acc available, skipp
191/191 [==============================] - 340s 2s/step - loss: 0.1692 - accuracy: 0.9535 - val_loss: 1.3360 - val_accuracy: 0.6794 - lr: 2.0000e-04
Epoch 3/20
191/191 [==============================] - ETA: 0s - loss: 0.1395 - accuracy: 0.9629WARNING:tensorflow:Can save best model only with val_acc available, skipp

Epoch 3: ReduceLROnPlateau reducing learning rate to 4.0000001899898055e-05.
191/191 [==============================] - 348s 2s/step - loss: 0.1395 - accuracy: 0.9629 - val_loss: 1.4697 - val_accuracy: 0.6631 - lr: 2.0000e-04
Epoch 4/20
191/191 [==============================] - ETA: 0s - loss: 0.1122 - accuracy: 0.9718Restoring model weights from the end of the best epoch: 1.
WARNING:tensorflow:Can save best model only with val_acc available, skipping.
191/191 [==============================] - 343s 2s/step - loss: 0.1122 - accuracy: 0.9718 - val_loss: 1.1237 - val_accuracy: 0.7306 - lr: 4.0000e-05
Epoch 4: early stopping
```





Lighting zone	RTU	Thermal zones
North Wing	1	36, 37, 38, 39, 40, 41, 42, 64, 65, 66, 67, 68, 69, 70
North Wing	2	19, 20, 27, 28, 29, 30, 31, 32, 33, 34, 35, 43, 44, 49, 50, 57, 58, 59, 60, 62, 63, 71, 72
South Wing	3	18, 25, 26, 45, 48, 55, 56, 61
South Wing	4	16, 17, 21, 22, 23, 24, 46, 47, 51, 52, 53, 54

Table 1. Key Electrical Panels.



Introduction

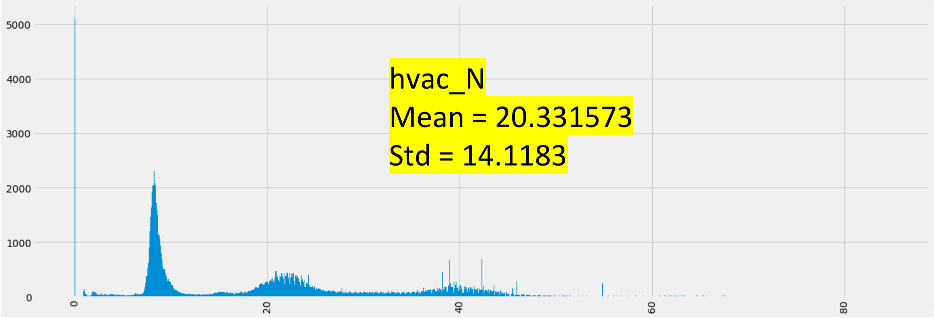
Worldwide building consumes lot of energy and contributes to emissions
Recently multiple building data are available from the diff. sensor data connected to building management systems

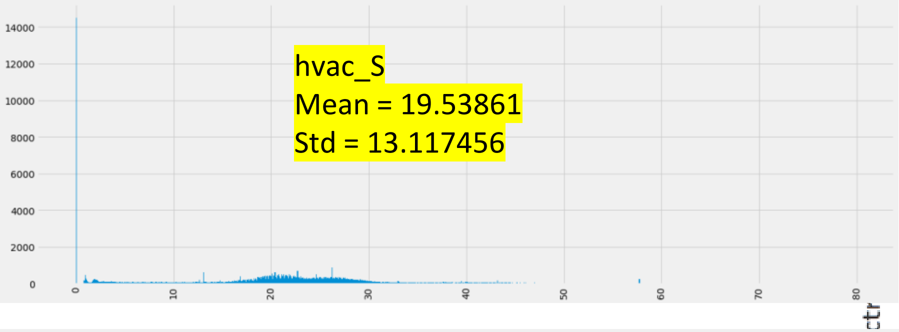
ML techniques can be analyzed to evaluate various sensor data and relate with its energy consumption pattern
These data/techniques can help to develop strategies for reducing energy consumption of buildings

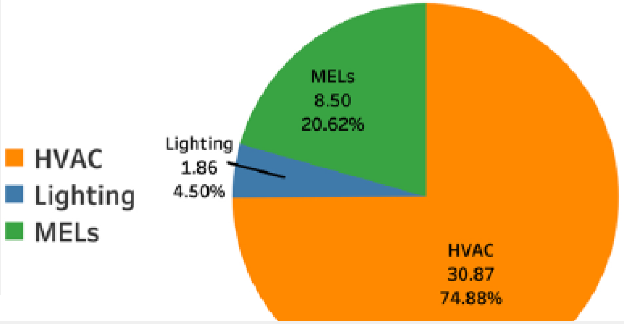
Objective:

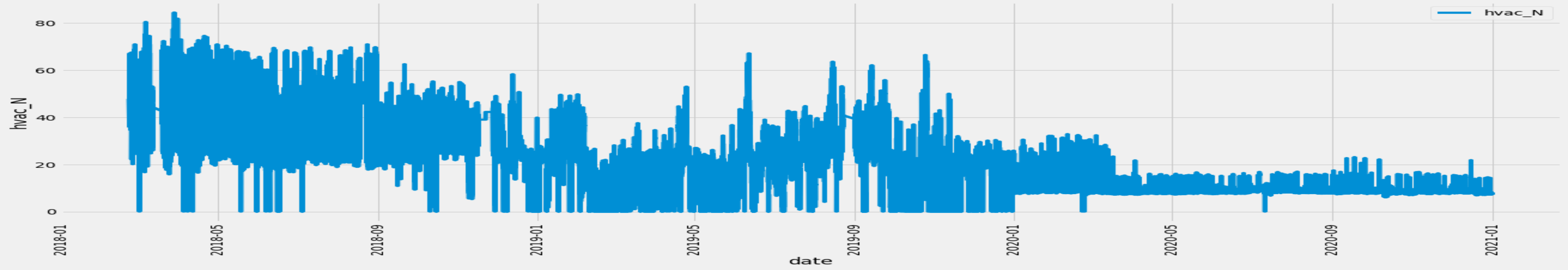
We want to analyze the energy consumption pattern of building using ML techniques

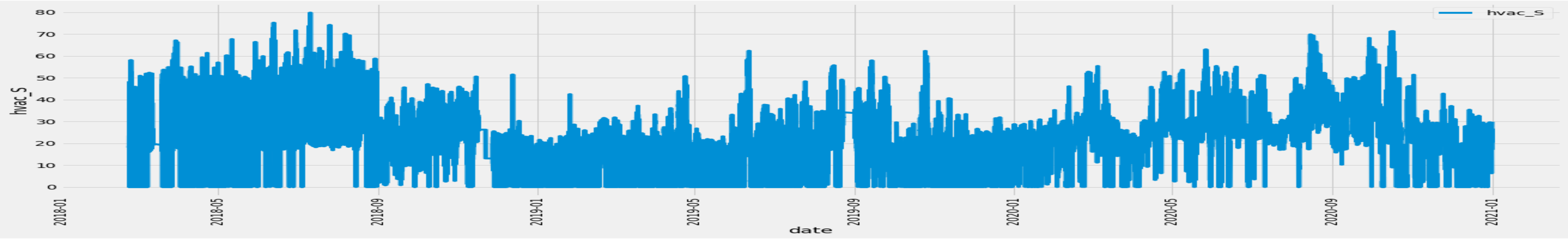
Visualizing the dataset: Hvac power consumption



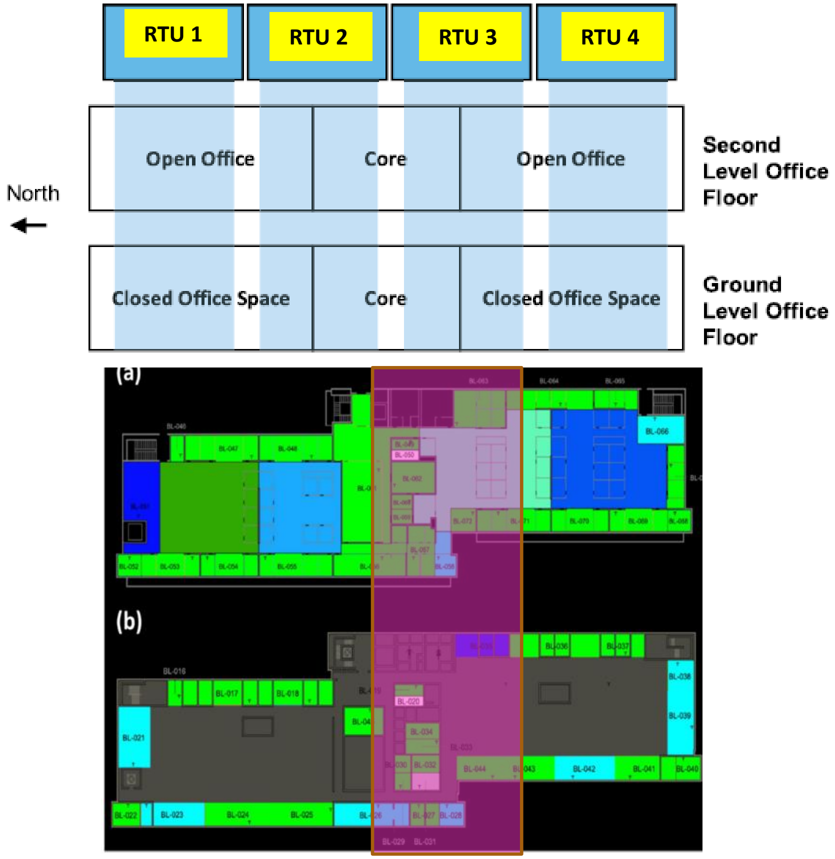








Plan for using ML model



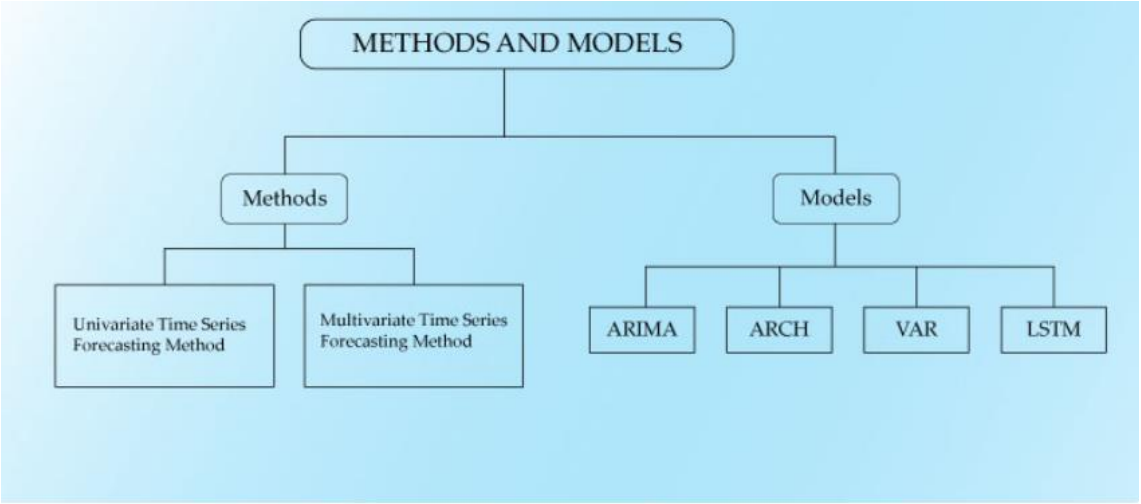
Lighting zone	RTU	Thermal zones
North Wing	1	36, 37, 38, 39, 40, 41, 42, 64, 65, 66, 67, 68, 69, 70
North Wing	2	19, 20, 27, 28, 29, 30, 31, 32, 33, 34, 35, 43, 44, 49, 50, 57, 58, 59, 60, 62, 63, 71, 72
South Wing	3	18, 25, 26, 45, 48, 55, 56, 61
South Wing	4	16, 17, 21, 22, 23, 24, 46, 47, 51, 52, 53, 54

1. Relate energy consumption of HVAC with temp. data

Improvements required to the ML model

- Time dependency is an important factor for learning time-series data characteristics.

- **Suitable models for forecasting time series data:**
    - Autoregressive Integrated Moving Average (**ARIMA**): It's a statistical model for time series data analysis or predicting future trends
    - Moving Average: Uses previous forecast error rather than directly using previous values as in regression.
    - Exponential Smoothening: forecasting method only for univariate data
    - Long Short-Term Memory (**LSTM**) it's a variety of Deep neural network model



Time-Series Forecasting: Methods and Models in Machine Learning

References

1. Hong, Tianzhen; Luo, Na; Blum, David; Wang, Zhe (2022), A three-year building operational performance dataset for informing energy efficiency , Dryad, Dataset, https://doi.org/10.7941/D1N33Q
2. https://www.analyticssteps.com/blogs/introduction-time-series-analysis-time-series-forecasting-machine-learning-methods-models
3. https://builtin.com/data-science/time-series-forecasting-python