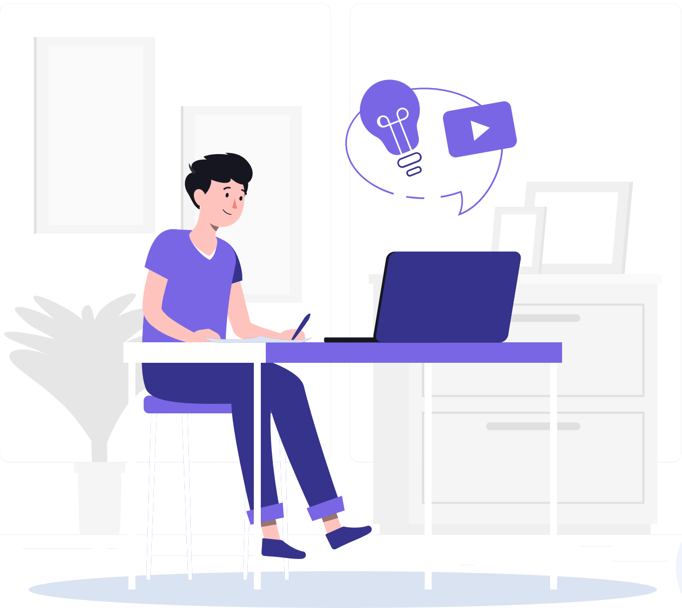
Occupancy Detection

Created and Presented
By
Cristina Sahoo



Applications

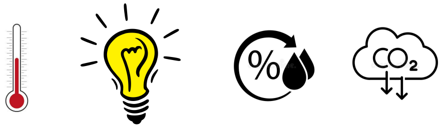
- Recent studies and measurements *[13-15]* report energy savings of 30% to 42% with accurate occupancy determination

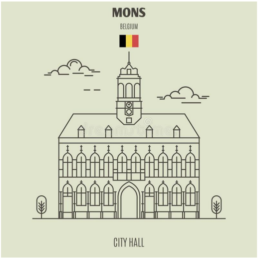
- When occupancy data was used as an input for HVAC control algorithms, energy savings were as high as 80% *[17]*

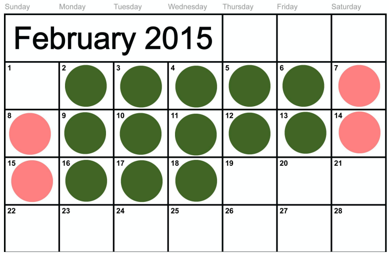
- A system that could accurately detect the presence of the occupants without using a camera is very interesting due to privacy concerns

- Other applications for occupancy detection include security and determination of building occupant behaviors

Problem Statement



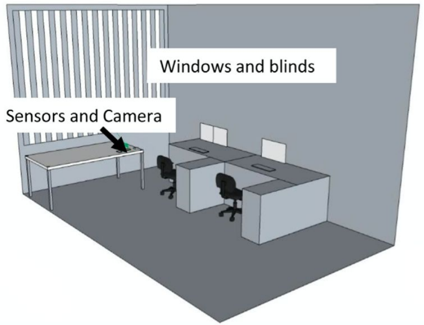


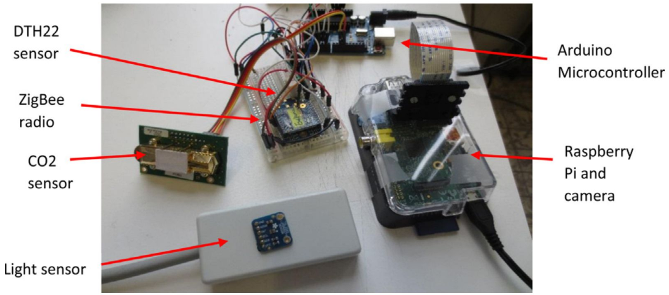


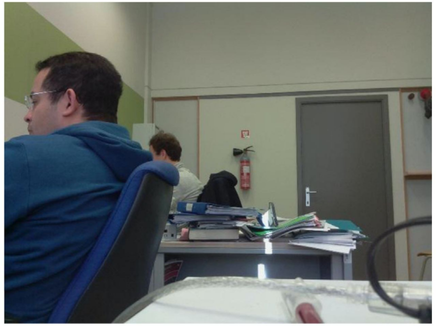
- Identify trends over time and correlation between environment variables and occupancy
- Identify algorithms/models with best accuracy score for predicting occupancy

Data Source and Collection

Occupancy Detection Dataset - UCI Machine Learning Repository







Data Cleaning and Preprocessing

Dataset	Description	Dimensions	Class Distribution (class 0 to class 1)
occupancy	combined data	17895, 7	76.89% to 23.10%
train	training data	8143, 7	78.76% to 21.23%
test	testing data, door open	2665, 7	63.52% to 36.47%
test2	testing data, door closed	9752, 7	78.98% to 21.01%

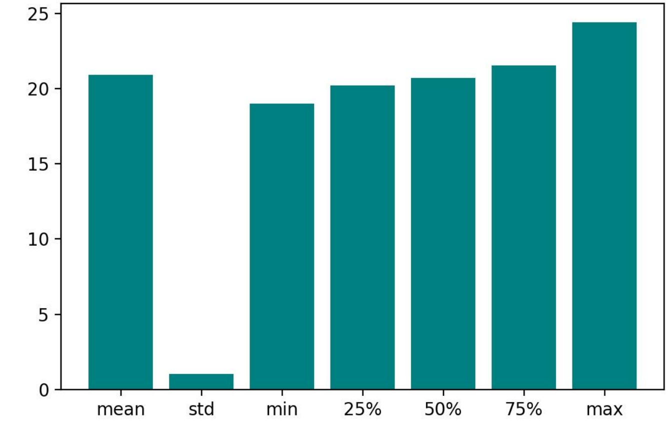
- Remove/impute null values
- Remove/impute outliers
- Remove duplicate observations
- Rename columns

- Round dates to nearest minute
- Sort by date, chronologically
- Index by date to allow time series type exploration of the data
- Create the "weekday" feature

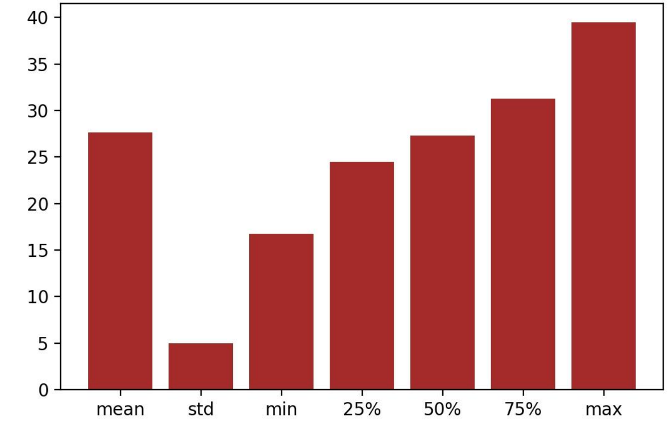
Data Dictionary

Feature Name	Feature Description	Units of Measurement or Format
date	time the observation was recorded	year-month-day hour:minute:second
temperature	temperature recorded	Celsius
humidity	relative humidity recorded	%
light	light recorded at time of observation	Lux
co2	CO2 measured at the time of observation	ppm, parts per million
humidity_ratio	derived quantity from temperature and relative humidity	kgwater-vapor/kg-air
occupancy	status of room occupancy	0 for not occupied, 1 for occupied status
weekday	indicates if the timestamp is weekday or weekend	1 for weekday, 0 for weekend

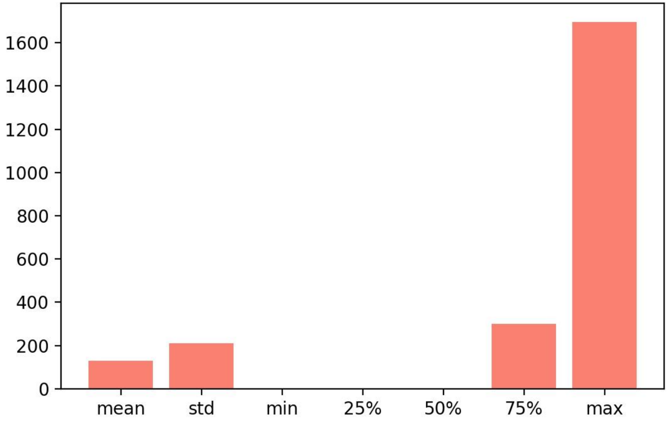
Descriptive Statistics: Temperature



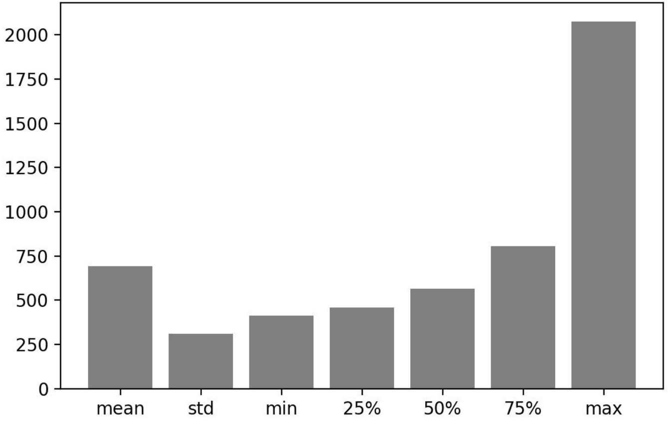
Descriptive Statistics: Humidity



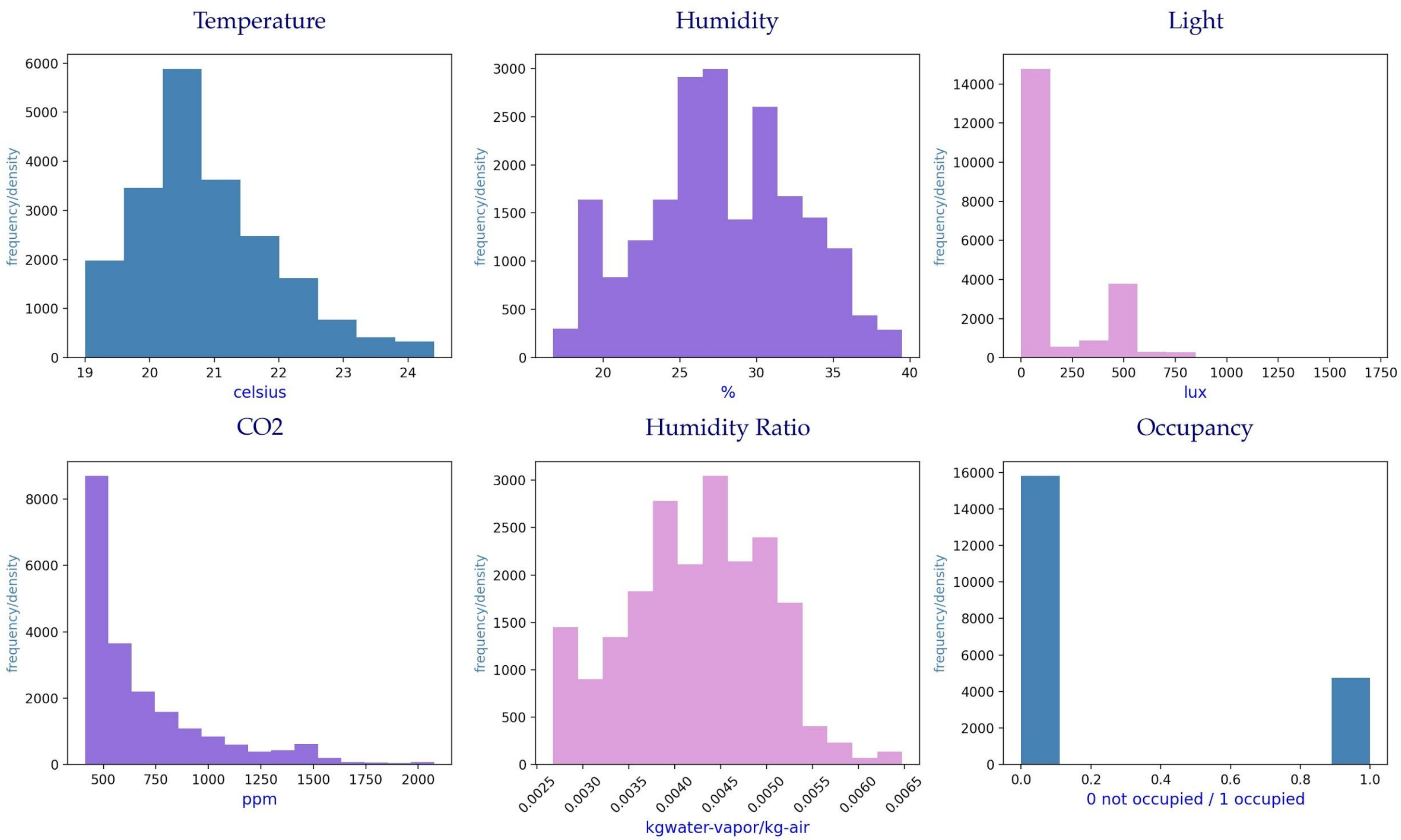
Descriptive Statistics: Light

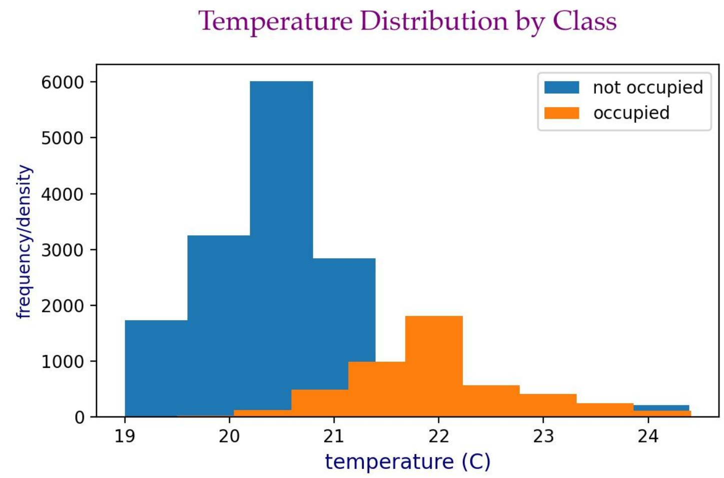


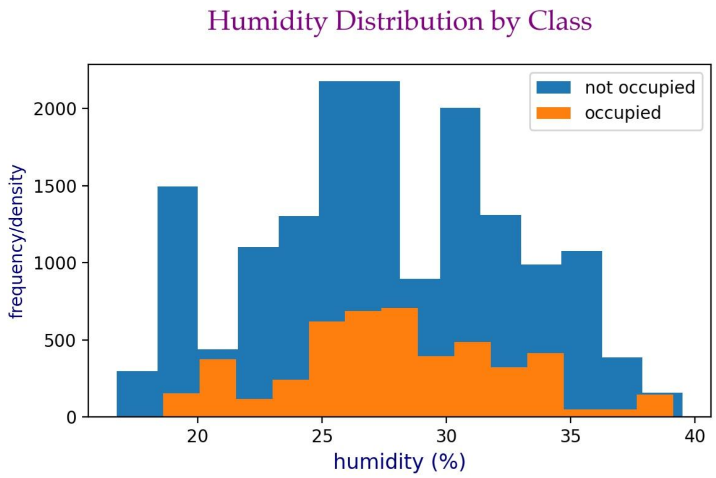
Descriptive Statistics: CO2

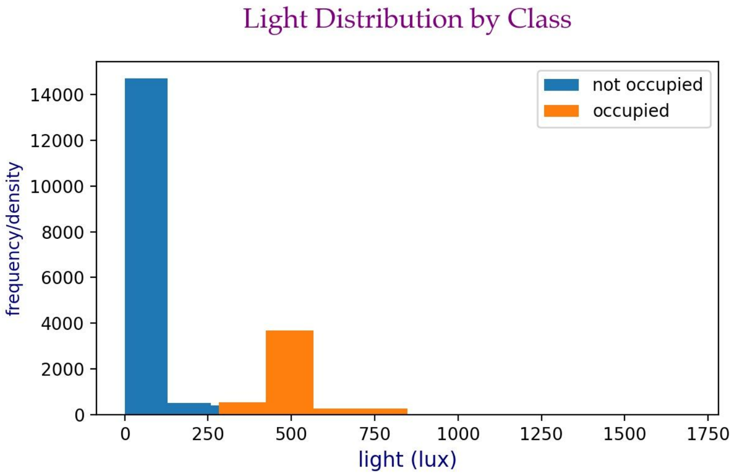


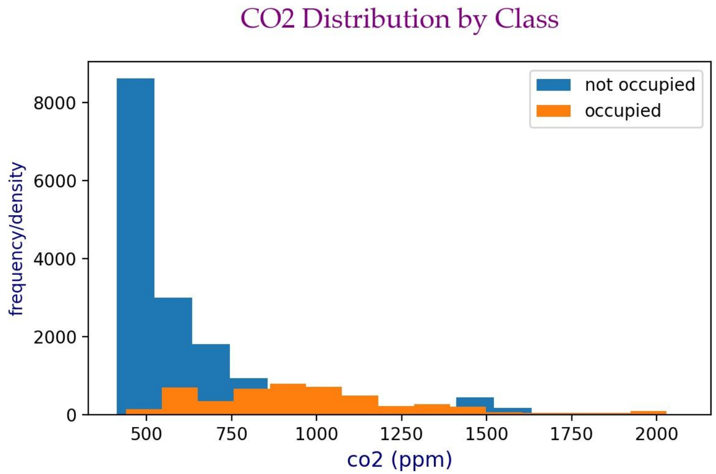
Data Distributions



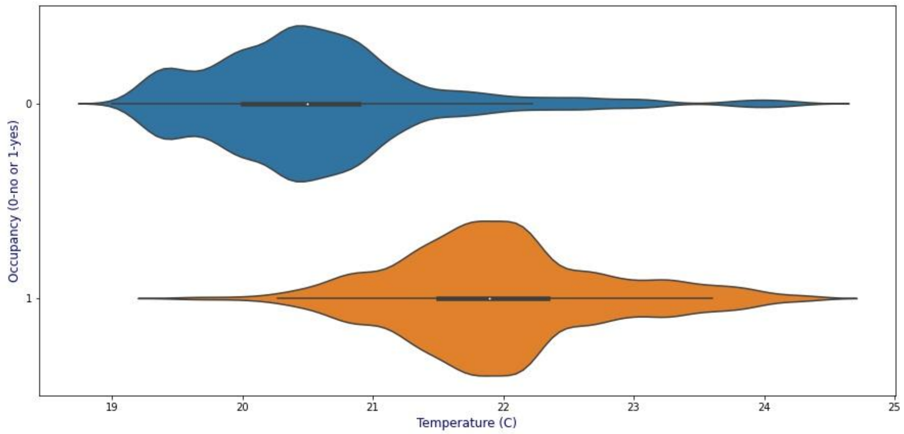

Temperature Distribution by Class
not occupied
occupied
frequency/density
temperature (C)
0
1000
2000
3000
4000
5000
6000
19
20
21
22
23
24


Humidity Distribution by Class
not occupied
occupied
frequency/density
humidity (%)
0
500
1000
1500
2000
20
25
30
35
40

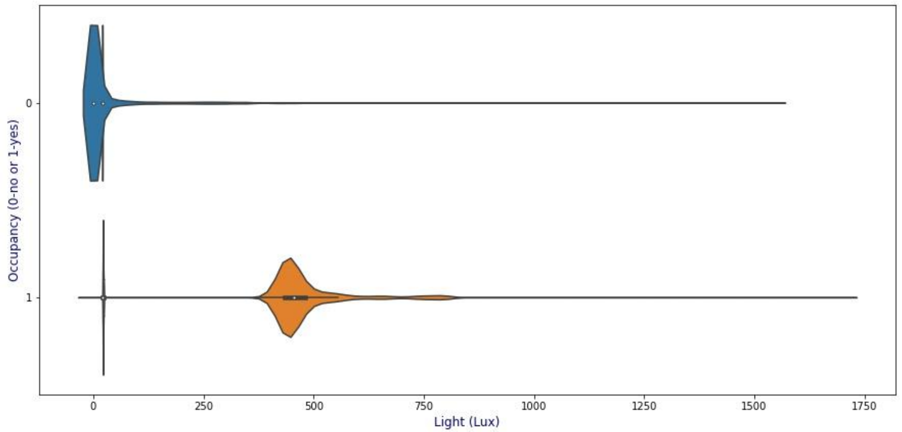

Light Distribution by Class
not occupied
occupied
frequency/density
light (lux)
0
2000
4000
6000
8000
10000
12000
14000
0
250
500
750
1000
1250
1500
1750

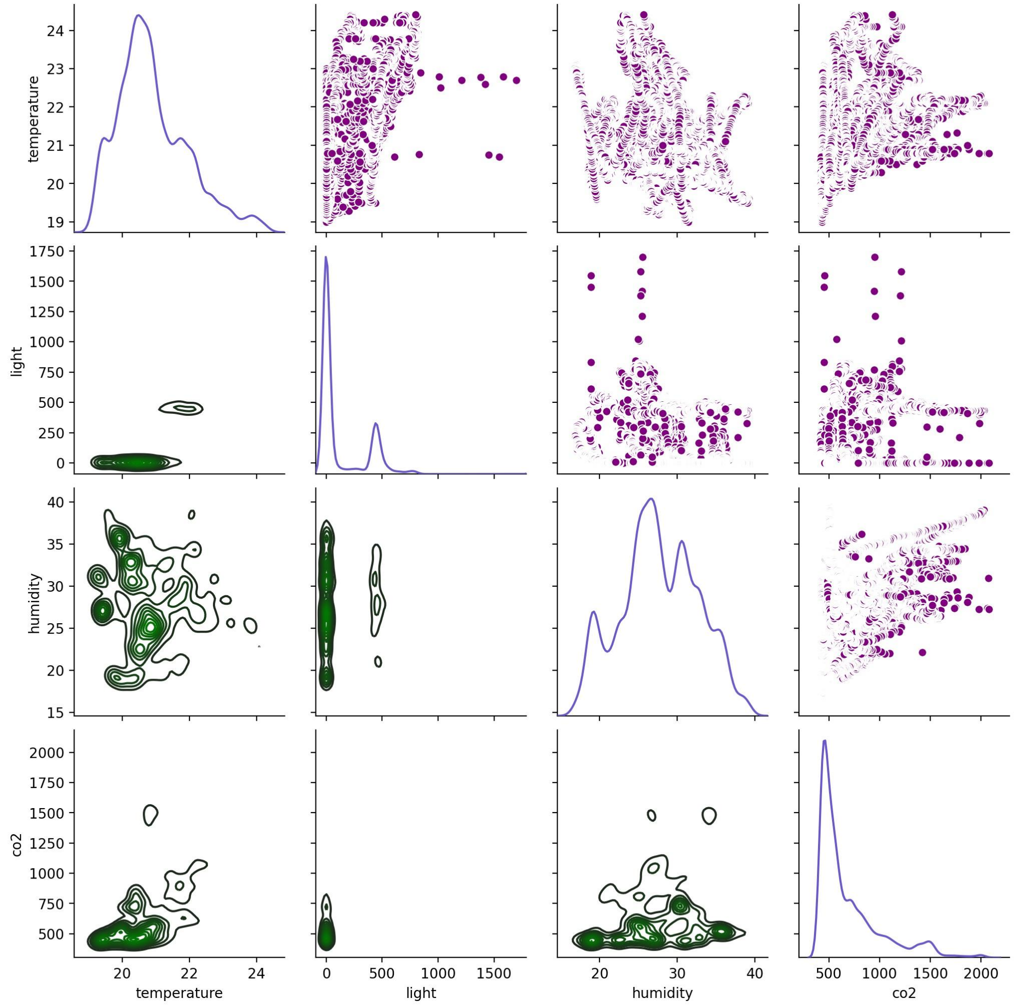

CO2 Distribution by Class
not occupied
occupied
frequency/density
co2 (ppm)
0
2000
4000
6000
8000
500
750
1000
1250
1500
1750
2000

Temperature Distribution by Class

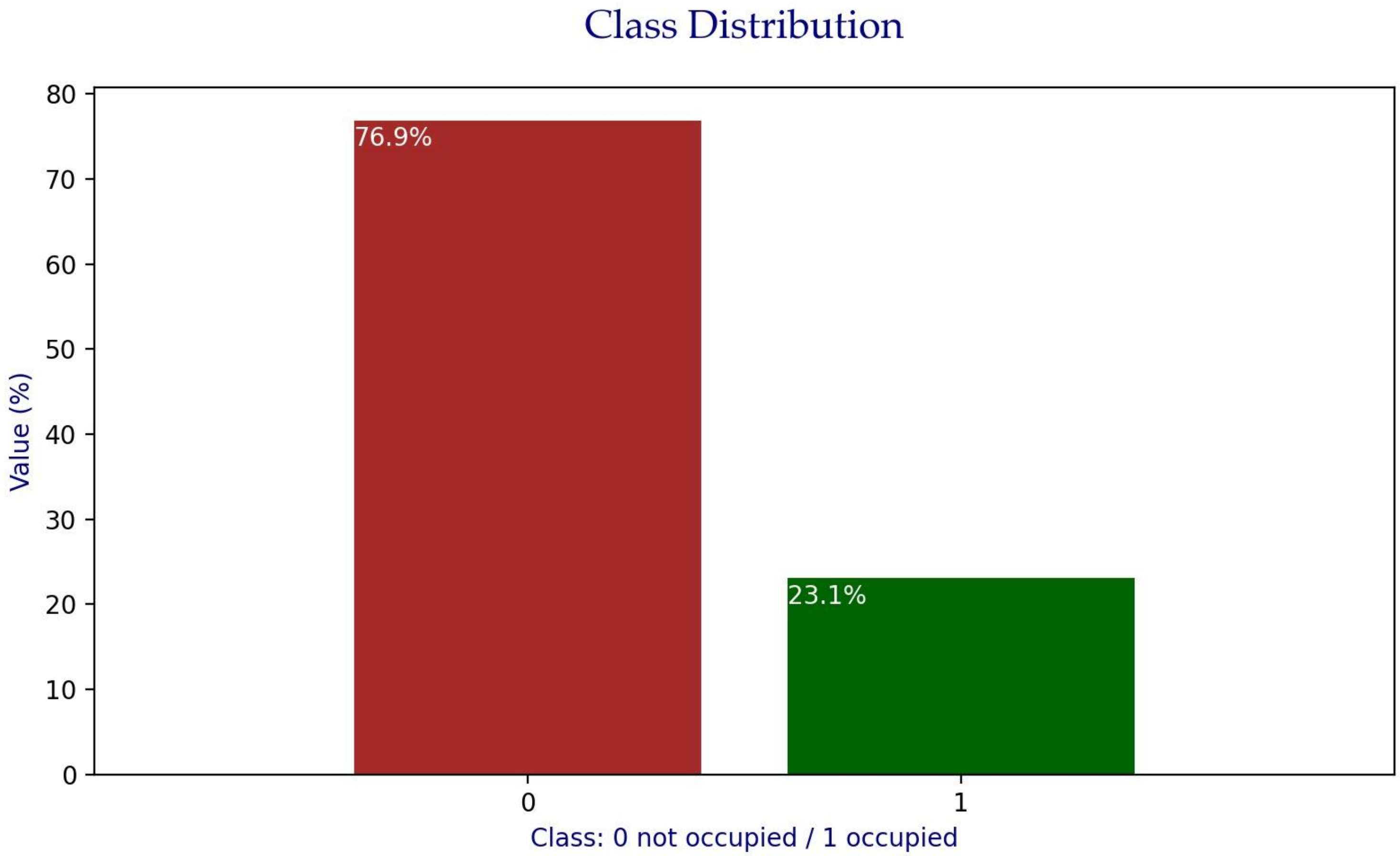


Light Distribution by Class

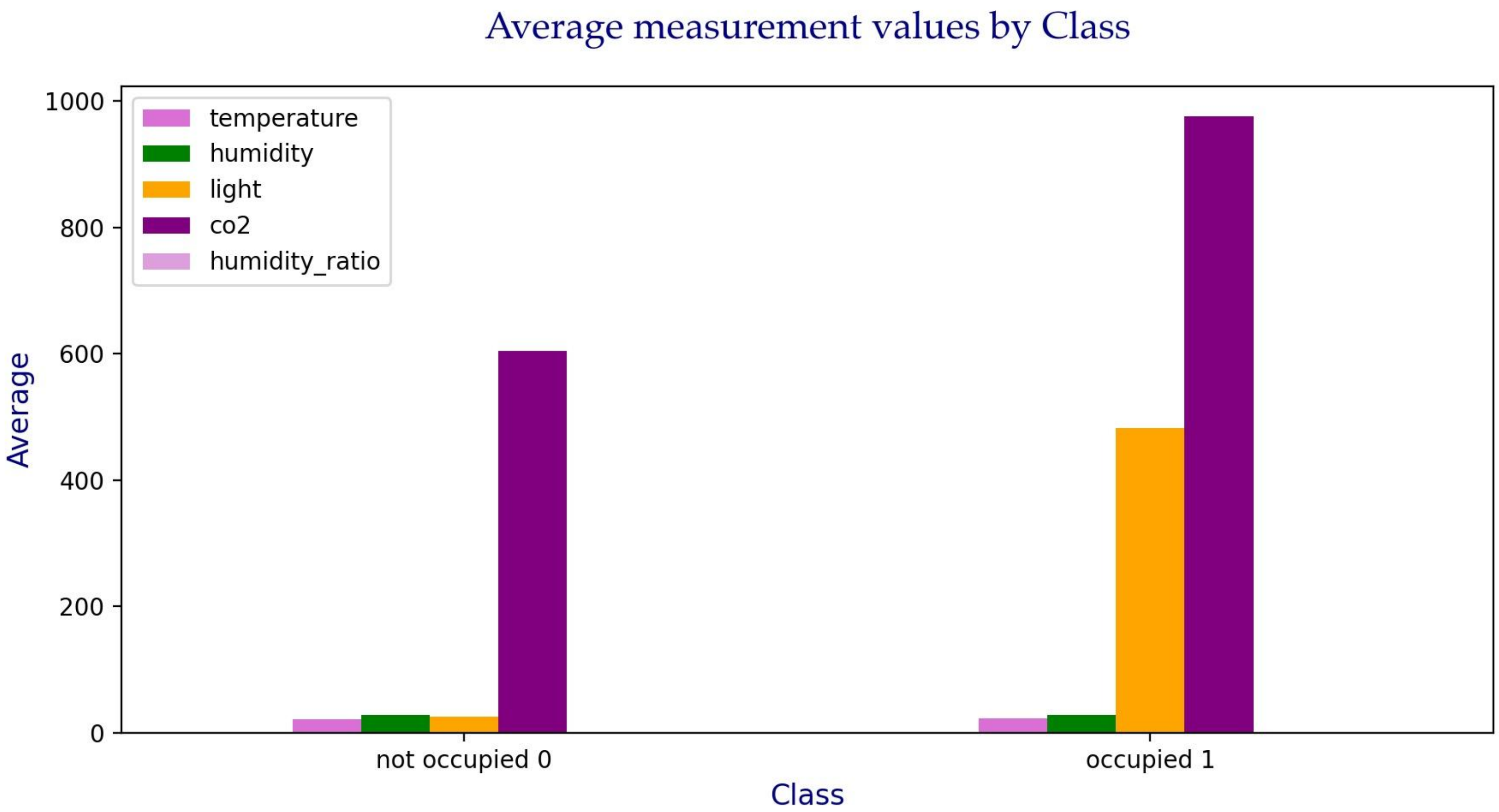


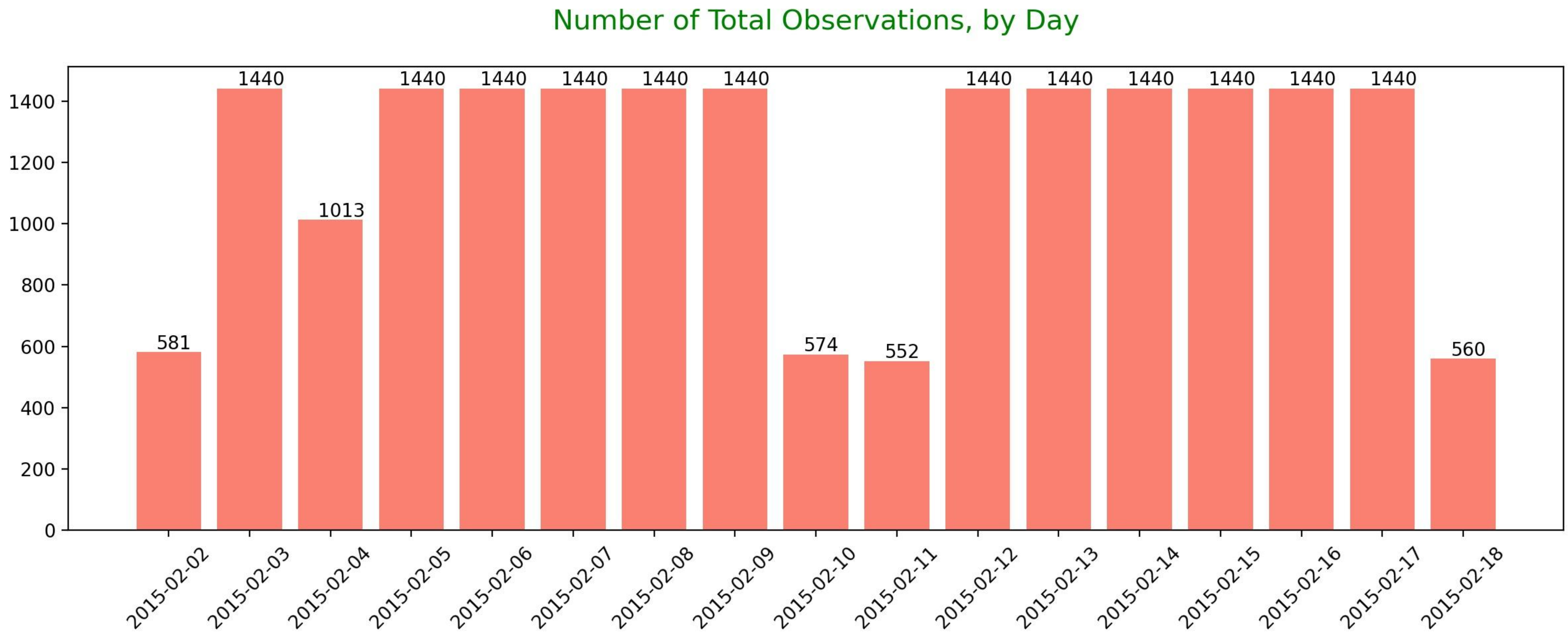


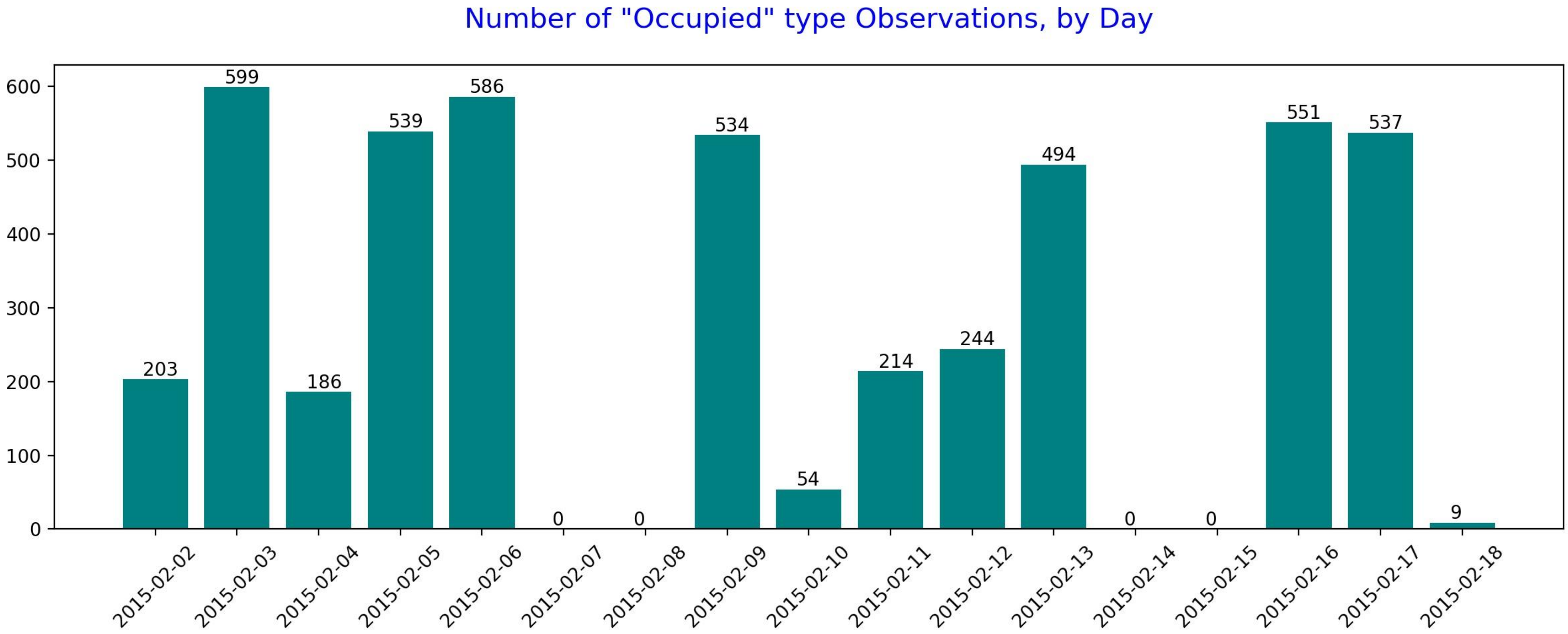
temperature
light
humidity
co2

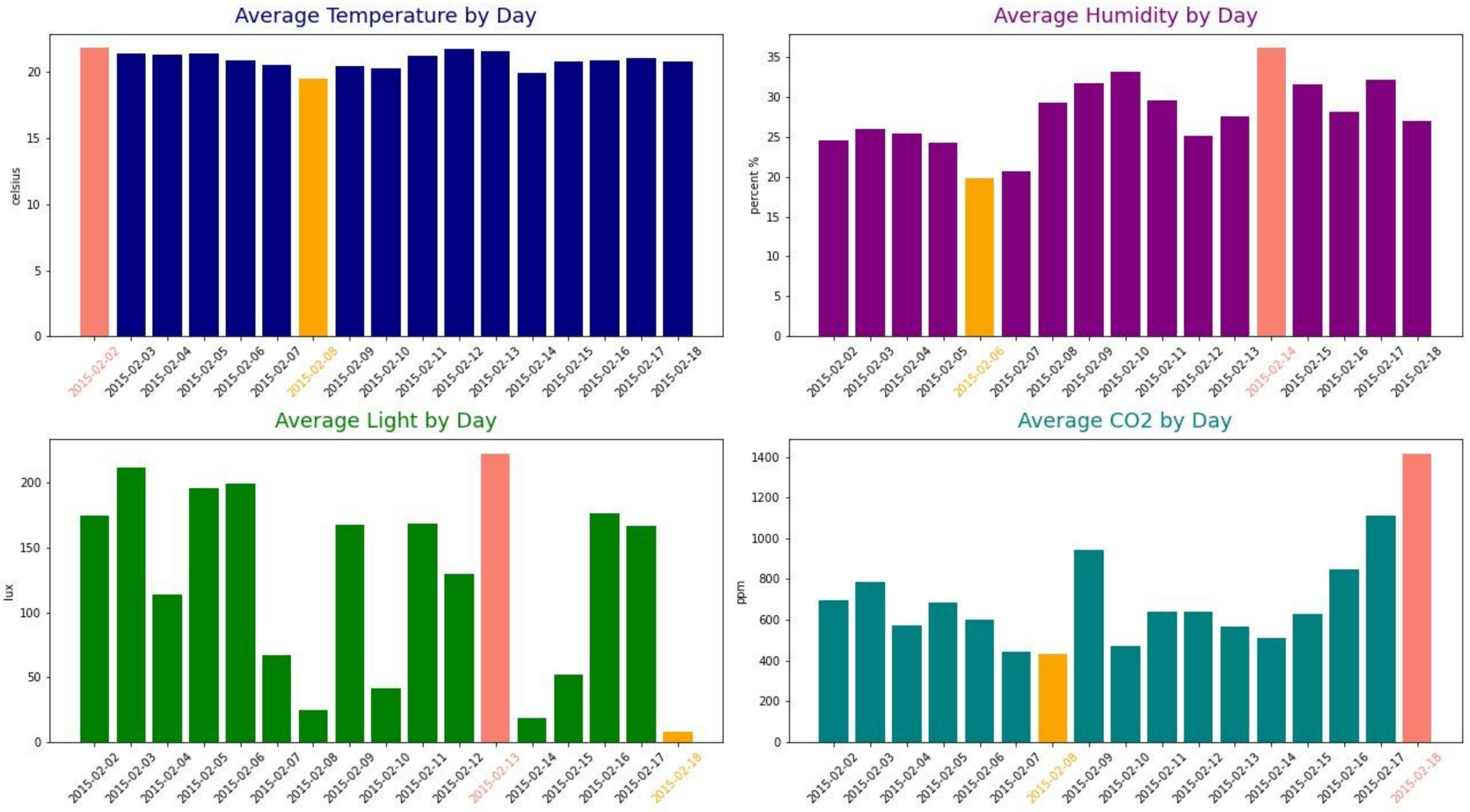


Class Distribution
Value (%)
80
70
60
50
40
30
20
10
0
76.9%
23.1%
0
1
Class: 0 not occupied / 1 occupied

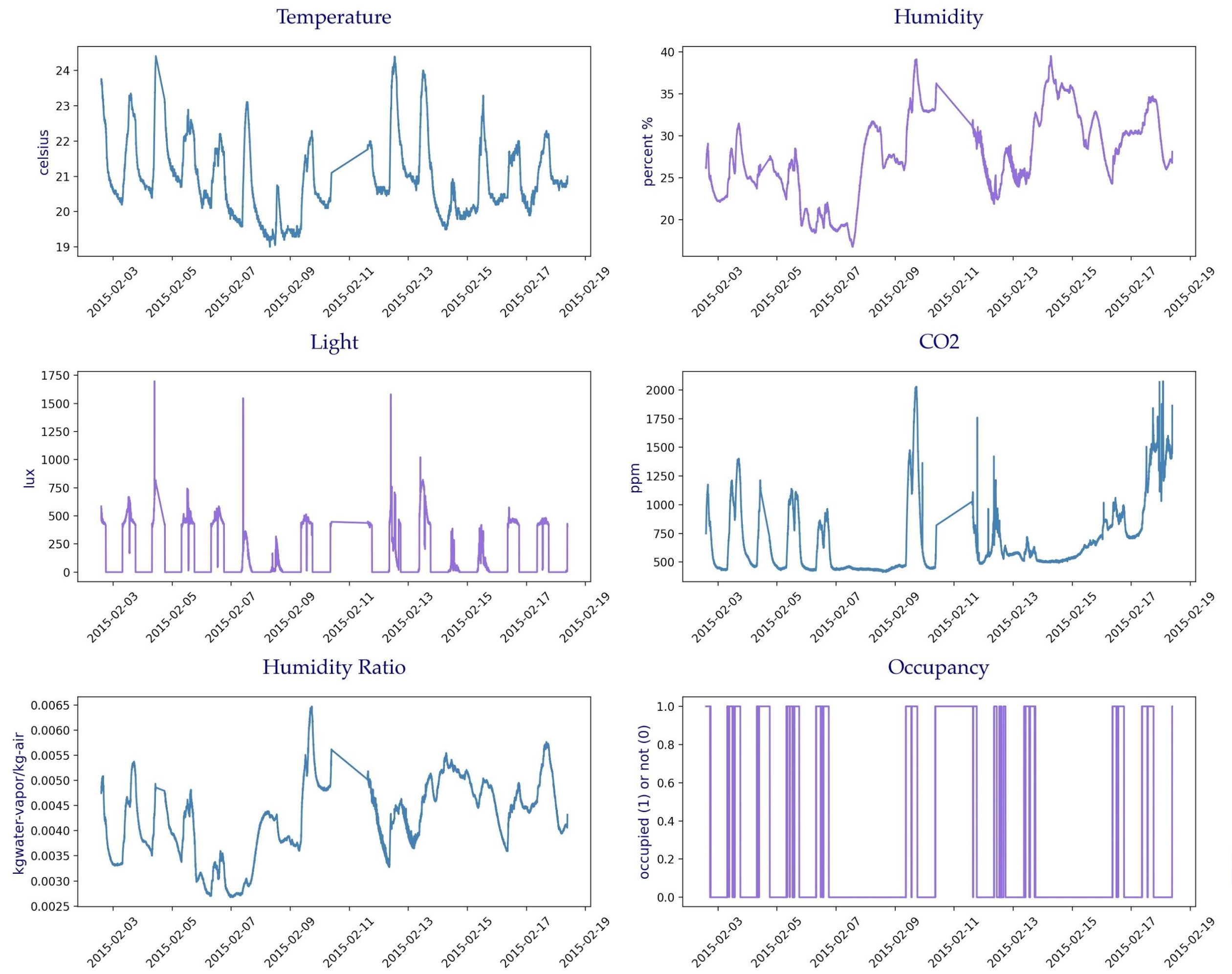

Average measurement values by Class
temperature
humidity
light
co2
humidity_ratio
1000
800
600
400
200
0
Average
not occupied 0
occupied 1
Class

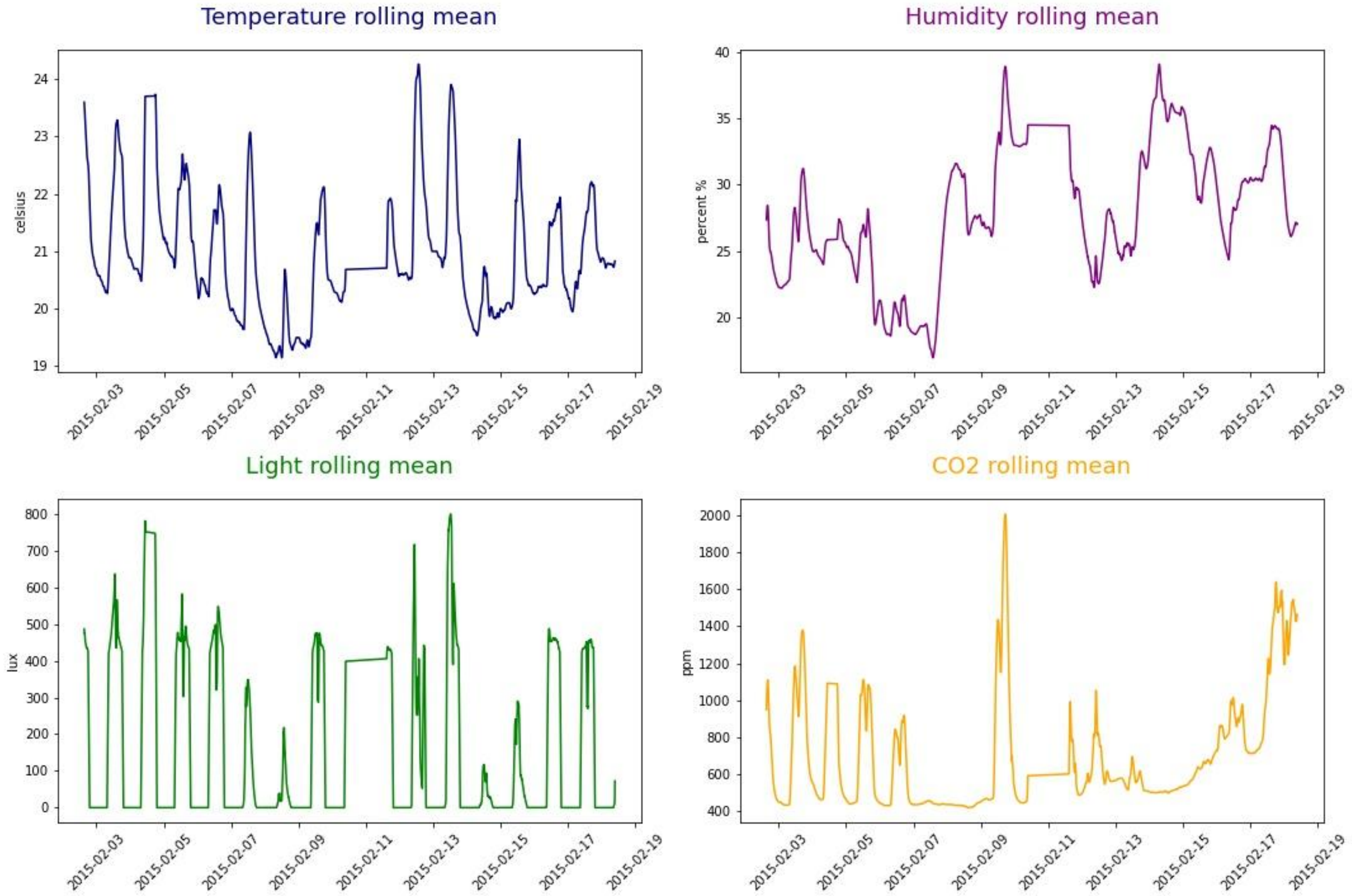

Number of Total Observations, by Day
0
200
400
600
800
1000
1200
1400
581
1440
1013
1440
1440
1440
1440
1440
574
552
1440
1440
1440
1440
1440
1440
560
2015-02-02
2015-02-03
2015-02-04
2015-02-05
2015-02-06
2015-02-07
2015-02-08
2015-02-09
2015-02-10
2015-02-11
2015-02-12
2015-02-13
2015-02-14
2015-02-15
2015-02-16
2015-02-17
2015-02-18


Number of "Occupied" type Observations, by Day
0
100
200
300
400
500
600
203
599
186
539
586
0
0
534
54
214
244
494
0
0
551
537
9
2015-02-02
2015-02-03
2015-02-04
2015-02-05
2015-02-06
2015-02-07
2015-02-08
2015-02-09
2015-02-10
2015-02-11
2015-02-12
2015-02-13
2015-02-14
2015-02-15
2015-02-16
2015-02-17
2015-02-18

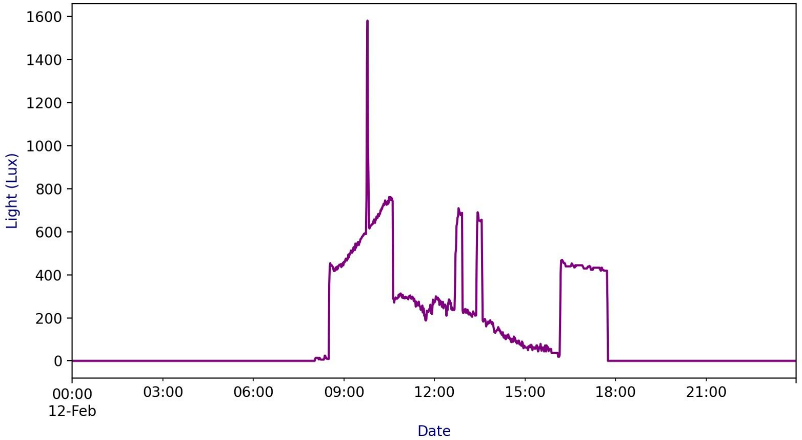


Average Temperature by Day
celsius
Average Humidity by Day
percent %
Average Light by Day
lux
Average CO2 by Day
ppm
2015-02-02
2015-02-03
2015-02-04
2015-02-05
2015-02-06
2015-02-07
2015-02-08
2015-02-09
2015-02-10
2015-02-11
2015-02-12
2015-02-13
2015-02-14
2015-02-15
2015-02-16
2015-02-17
2015-02-18

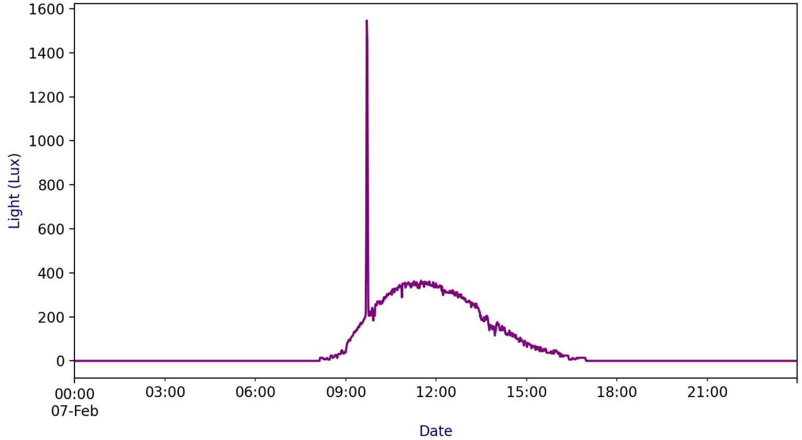

Temperature
celsius
Humidity
percent %
Light
lux
CO2
ppm
Humidity Ratio
kgwater-vapor/kg-air
Occupancy
occupied (1) or not (0)
2015-02-03
2015-02-05
2015-02-07
2015-02-09
2015-02-11
2015-02-13
2015-02-15
2015-02-17
2015-02-19


Temperature rolling mean
celsius
Humidity rolling mean
percent %
Light rolling mean
lux
CO2 rolling mean
ppm
2015-02-03
2015-02-05
2015-02-07
2015-02-09
2015-02-11
2015-02-13
2015-02-15
2015-02-17
2015-02-19

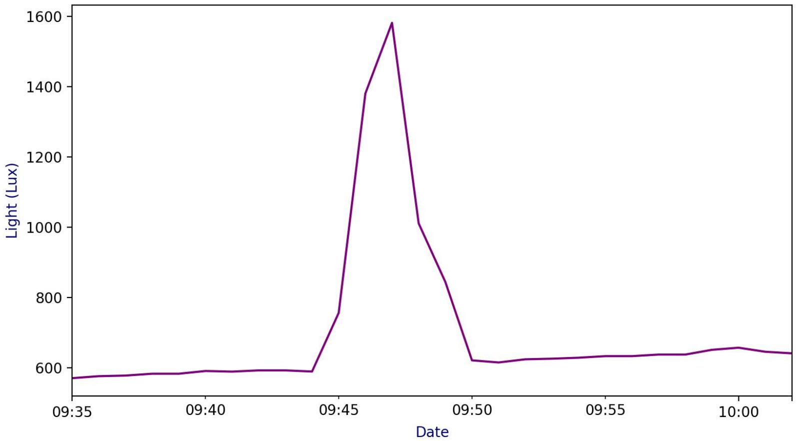
Light measurements on Thursday 2015-02-12



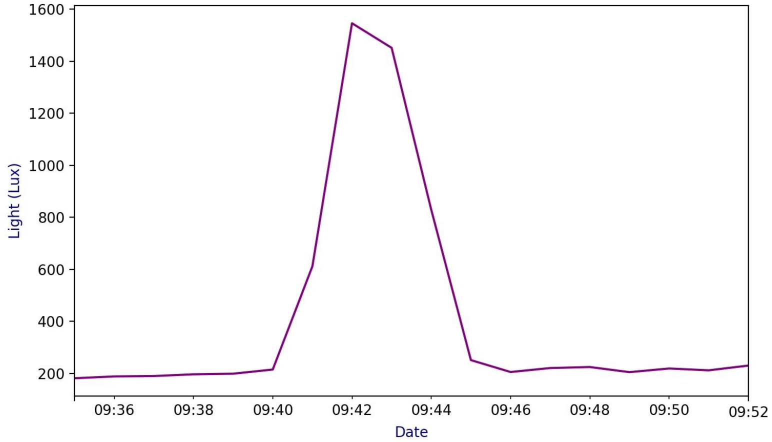
Light measurements on Saturday 2015-02-07



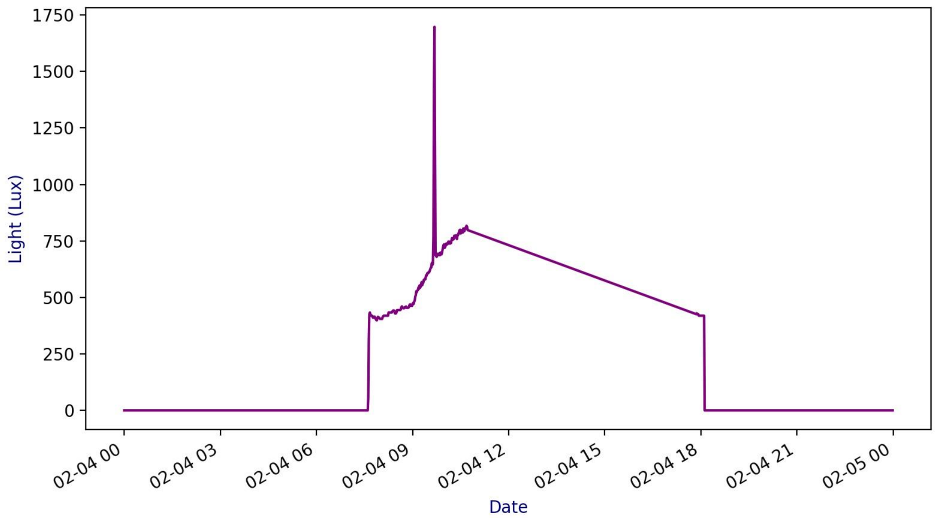
Light measurements on Thursday 2015-02-12 9:35am to 10:02am



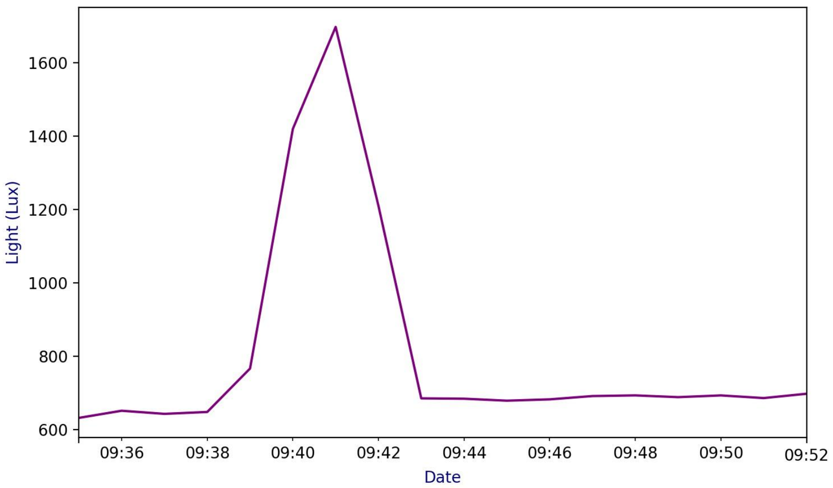
Light measurements on Saturday 2015-02-07 between 9:35am and 9:52am



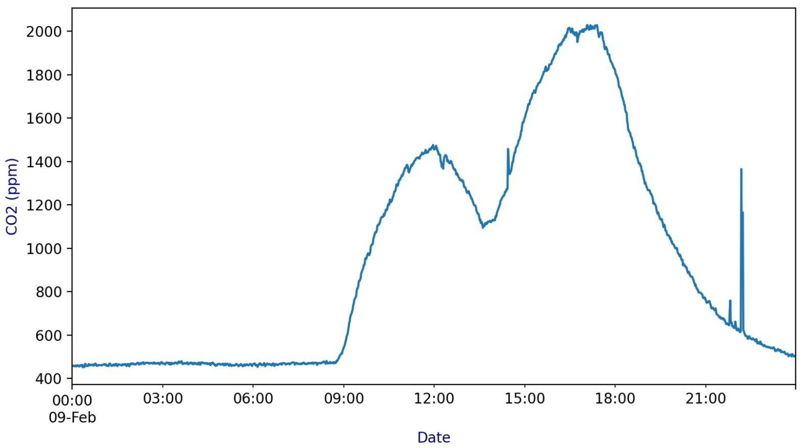
Light measurements on Wednesday 2015-02-04



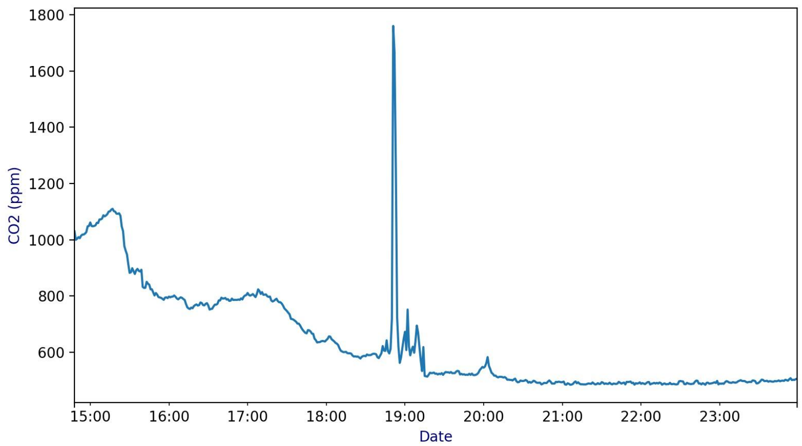
Light measurements on Wednesday 2015-02-04 between 9:35am and 9:52am



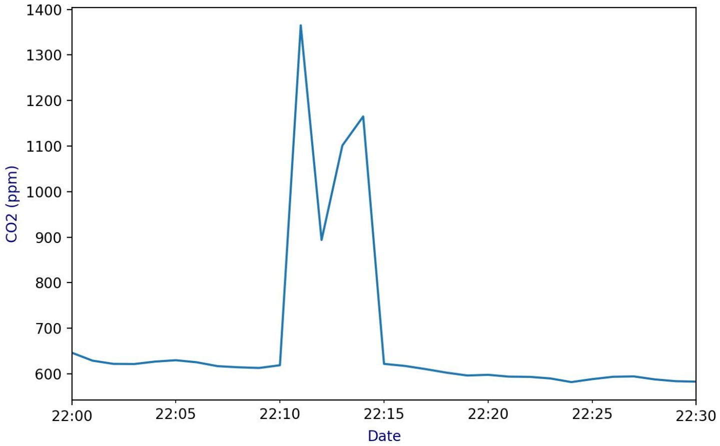
CO2 measurements on Monday 2015-02-09



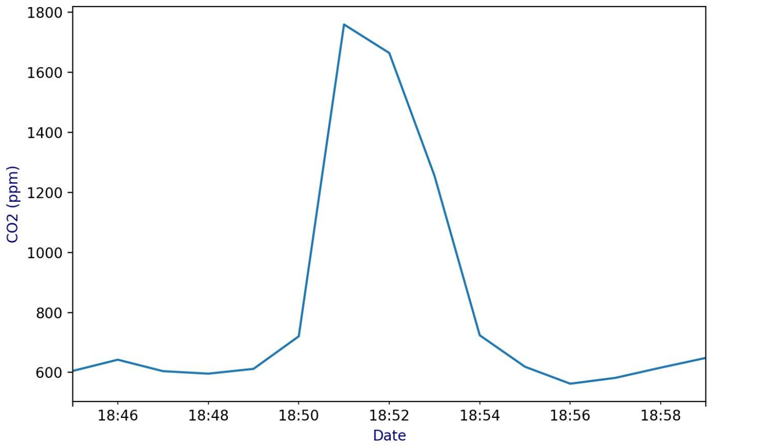
CO2 measurements on Wednesday 2015-02-11

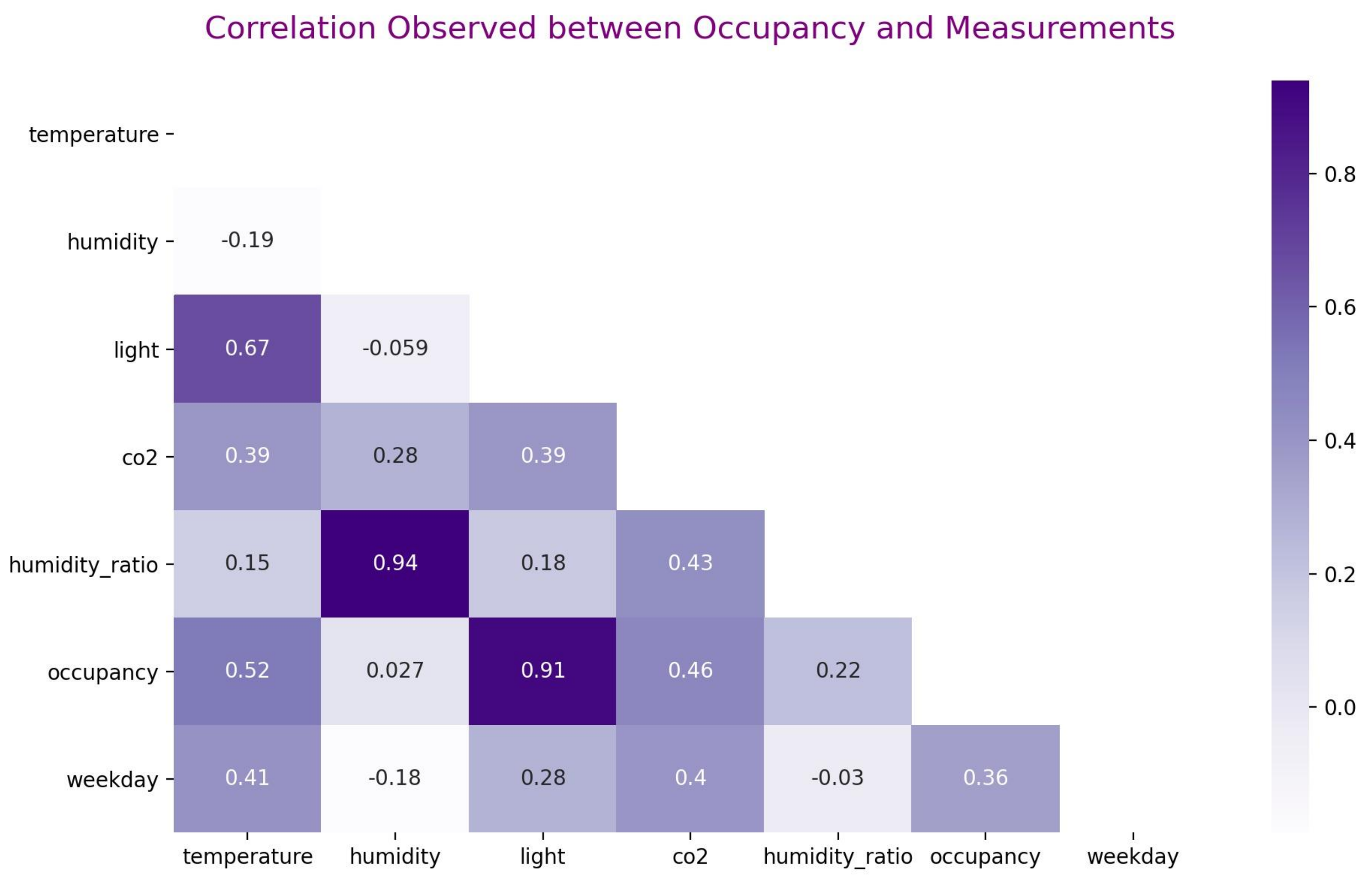


CO2 measurements on Monday 2015-02-09 between 22:00 and 22:30



CO2 measurements on Wednesday 2015-02-11 between 18:45 and 18:59




Correlation Observed between Occupancy and Measurements
temperature
humidity
light
co2
humidity_ratio
occupancy
weekday
-0.19
0.67
-0.059
0.39
0.28
0.39
0.15
0.94
0.18
0.43
0.52
0.027
0.91
0.46
0.22
0.41
-0.18
0.28
0.4
-0.03
0.36
temperature
humidity
light
co2
humidity_ratio
occupancy
weekday
0.8
0.6
0.4
0.2
0.0

Modeling

- Models were trained on 75% of the combined data, with 25% left for testing.
- Models were run on combinations of the [temperature, light, co2, humidity, humidity_ratio, weekday] features.
- A total of 154 models were run.
- The best model was Random Forest [temperature, humidity, light, co2, weekday]. At training time, the model reported 99.36% accuracy on a 75/25 train/test split of combined data, 96.74% accuracy on test data with the door open, and 99.87% accuracy on test data with the door closed.

Abbreviation	Meaning
RF	Random Forest
LDA	Linear Discriminant Analysis
GBM	Gradient Boosting Machine
AdaBoost	Adaptive Boosting
KNN	K-Nearest Neighbor
SVC	Support Vector Classification

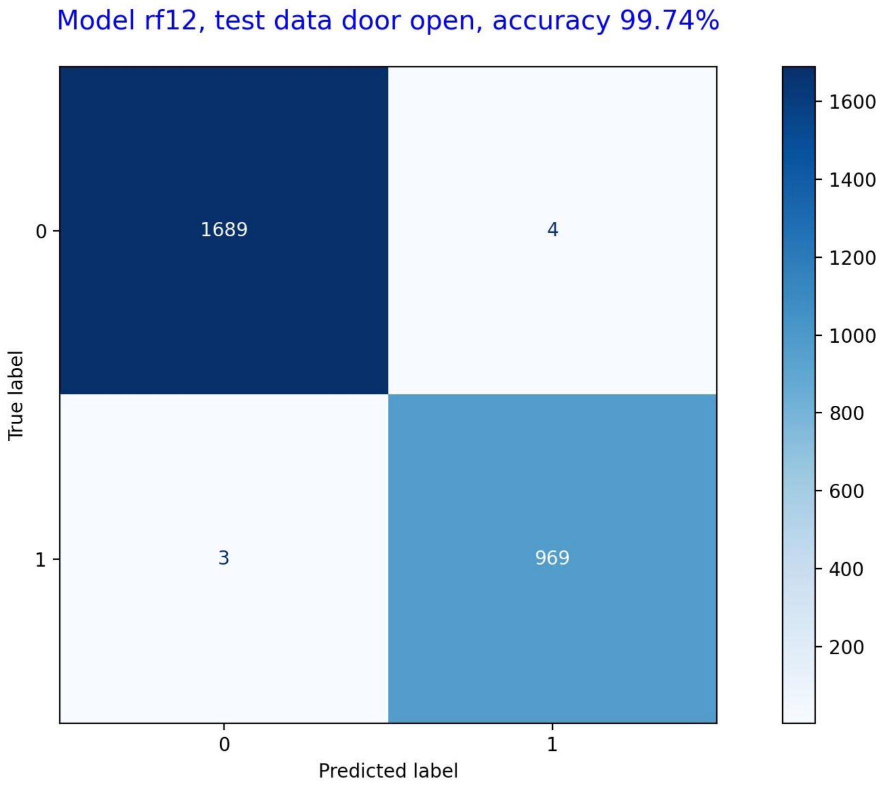
Dataset	Class 0 Not-occupied	Class 1 Occupied
Combined	77%	23%
Training	79%	21%
Door Open	64%	36%
Door Closed	79%	21%

Model Id	Model name	Features	Best score	Train score	Test score	Sensitivity	Specificity	Precision	Accuracy	F1-score
12	**rf12**	**temperature, humidity, light, co2, weekday**	**0.9923**	**1.0000**	**0.9936**	**0.9874**	**0.9954**	**0.9849**	**0.9936**	0.9861
11	rf11	temperature, humidity, light, co2, humidity_ratio, weekday	0.9925	1.0000	0.9934	0.9865	0.9954	0.9849	**0.9934**	0.9857
0	rf0	temperature, humidity, light, co2, humidity_ratio	0.9922	1.0000	0.9932	0.9865	0.9952	0.9840	**0.9932**	0.9853
1	rf1	temperature, humidity, light, co2	0.9921	1.0000	0.9930	0.9865	0.9949	0.9832	**0.9930**	0.9849
3	rf3	temperature, humidity, light, humidity_ratio	0.9919	0.9997	0.9928	0.9857	0.9949	0.9832	**0.9928**	0.9844
56	cart12	temperature, humidity, light, co2, weekday	0.9896	1.0000	0.9928	0.9823	0.9960	0.9865	**0.9928**	0.9844
78	gbm12	temperature, humidity, light, co2, weekday	0.9915	1.0000	0.9926	0.9832	0.9954	0.9848	**0.9926**	0.9840
14	rf14	temperature, humidity, light, humidity_ratio, weekday	0.9923	0.9997	0.9926	0.9857	0.9947	0.9824	**0.9926**	0.9840
77	gbm11	temperature, humidity, light, co2, humidity_ratio, weekday	0.9919	1.0000	0.9926	0.9857	0.9947	0.9824	**0.9926**	0.9840
80	gbm14	temperature, humidity, light, humidity_ratio, weekday	0.9915	0.9995	0.9924	0.9848	0.9947	0.9824	**0.9924**	0.9836

Model name	Accuracy Door Open	Accuracy Door Closed
rf12	99.74%	99.87%
rf11	99.74%	99.86%
rf0	99.74%	99.85%
rf1	99.74%	99.85%
rf3	99.77%	99.79%
cart12	99.7%	99.88%
gbm12	99.66%	99.88%
rf14	99.74%	99.82%
gbm11	99.59%	99.87%
gbm14	99.7%	99.81%

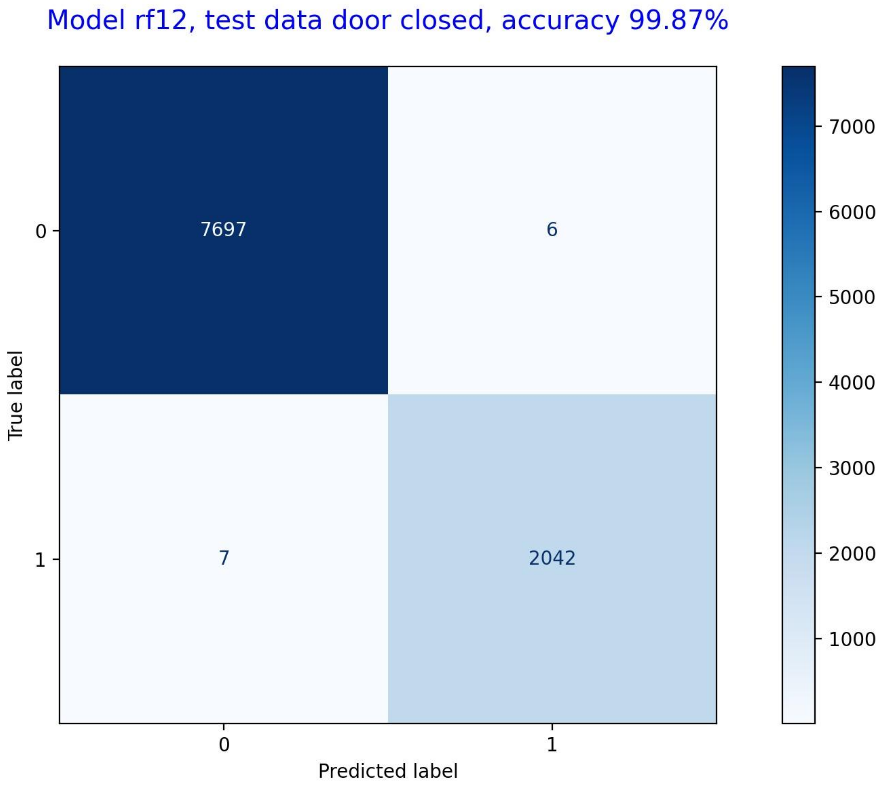
Random Forest

temperature, humidity, light, co2, weekday
Door Open



Random Forest

temperature, humidity, light, co2, weekday
Door Closed



Random Forest

temperature, humidity, light, co2, weekday

Door Open

date	temperature	humidity	light	co2	humidity_ratio	occupancy	weekday	predictions
2015-02-02 17:24:00	22.525	24.890000	426.000000	814.250000	0.004200	1	1	0
2015-02-02 17:34:00	22.600	25.066667	428.333333	849.333333	0.004250	0	1	1
2015-02-02 18:02:00	22.390	24.912000	418.600000	782.800000	0.004169	1	1	0
2015-02-03 07:43:00	20.310	23.200000	415.000000	470.800000	0.003414	1	1	0
2015-02-03 13:36:00	23.200	25.533333	634.000000	922.166667	0.004491	0	1	1
2015-02-03 13:37:00	23.200	25.500000	629.000000	899.000000	0.004485	0	1	1
2015-02-04 07:52:00	20.600	24.200000	399.500000	528.000000	0.003627	0	1	1

Random Forest

temperature, humidity, light, co2, weekday
Door Closed

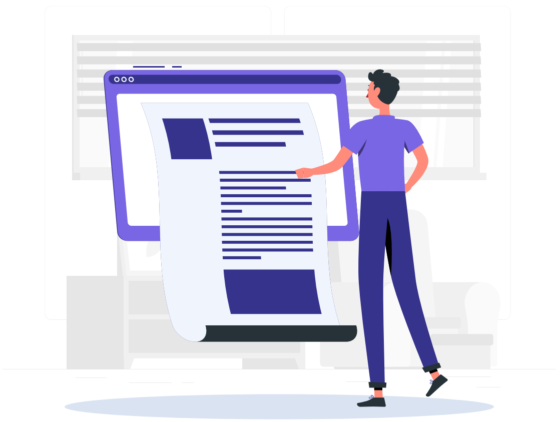
date	temperature	humidity	light	co2	humidity_ratio	occupancy	weekday	predictions
2015-02-12 08:34:00	20.600000	22.200000	442.750000	681.750000	0.003325	0	1	1
2015-02-12 12:56:00	24.390000	23.392500	236.500000	852.500000	0.004419	1	1	0
2015-02-12 13:26:00	24.200000	23.745000	690.500000	729.000000	0.004435	0	1	1
2015-02-12 16:10:00	22.390000	26.000000	191.500000	534.500000	0.004353	1	1	0
2015-02-12 17:44:00	21.890000	27.890000	279.333333	603.666667	0.004530	1	1	0
2015-02-13 08:59:00	21.290000	25.463333	510.333333	528.666667	0.003984	0	1	1
2015-02-13 09:59:00	22.650000	24.897500	726.750000	585.000000	0.004234	0	1	1
2015-02-13 10:04:00	22.722500	25.047500	714.500000	613.000000	0.004278	0	1	1
2015-02-13 13:47:00	23.745000	25.972500	659.250000	567.500000	0.004722	1	1	0
2015-02-13 18:05:00	21.290000	32.790000	0.000000	577.000000	0.005139	1	1	0
2015-02-16 08:52:00	20.600000	25.200000	405.000000	875.750000	0.003777	1	1	0
2015-02-17 08:39:00	20.566667	30.500000	301.000000	785.666667	0.004568	1	1	0
2015-02-17 13:28:00	21.675000	33.190000	431.000000	1139.500000	0.005328	0	1	1

Conclusions

- Light, temperature, and CO2 are higher when the room is occupied
- Light provides good class separation and should be included as a feature when modeling
- Weekday status improved the RF and GBM model scores
- High accuracies can be obtained with Random Forest, Gradient Boosting Machine, and CART models
- Lowest accuracies were observed with the LDA and AdaBoost models

Recommendations

- Incorporate the best model using a microprocessor
- Another option is to remotely process the data and only transmit the control signal for the HVAC system
- The light sensor appears to be very important in the classification task (example)
- The CO2 sensor can be very useful for demand control ventilation applications (article)



Next Steps

- Collect new data with up-to-date sensors
- Collect additional information, such as door open or closed, number of occupants, outdoor temperature, blinds closed or open, etc.
- Collect similar data for longer periods of time, i.e. months or years
- Create a model that can predict how many occupants are in the room
- Time series modeling

Resources

(1) Sensitivity and Specificity
(2) Accuracy
(3) Precision
(4) Executive Summary
(5) How to Predict Room Occupancy Based on Environmental Factors
(6) Technical Report and Project
(7) Improving Prediction of Office Room Occupancy Through Random Sampling
(8) Occupancy
(9) Room Occupancy Detection
(10) Classroom Occupancy Project
(11) Accurate occupancy detection of an office room from light, temperature, humidity and CO2 measurements using statistical learning models. Luis M. Candanedo, VÃ©ronique Feldheim. Energy and Buildings. Volume 112, 15 January 2016, Pages 28-39.
(12) Calculate day in the past
(13) V.L.Erickson, M.Á.Carreira-Perpinán, A.E.Cerpa, OBSERVE:Occupancy-based system for efficient reduction of HVAC energy, in: Proceedings of the 10th International Conference on, IEEE, Information Processing in Sensor Networks (IPSN), Chicago, IL, 2011, pp. 258–269.
(14) V.L.Erickson, M.Á.Carreira-Perpinán, A.E.Cerpa,Occupancy modeling and prediction for building energy management, ACM Trans. Sensor Netw. (TOSN) 10 (3) (2014) 42.
(15) Dong B., Andrews B., (2009). Sensor-based occupancy behavioral pattern recognition for energy and comfort management in intelligent buildings. Proceedings of Building Simulation.
(16) J. Brooks, S. Goyal, R. Subramany, Y. Lin, T. Middelkoop, L. Arpan, L. Carloni, P. Barooah, An experimental investigation of occupancy-based energy-efficient control of commercial building indoor climate, in: Proceeding of the IEEE 53rd Annual Conference on, IEEE, Decision and Control (CDC), Los Angeles, CA, 2014, pp. 5680–5685.
(17) J. Brooks, S. Kumar, S. Goyal, R. Subramany, P. Barooah, Energy-efficient control of under-actuated HVAC zones in commercial buildings, Energy Build. 93 (2015) 160–168.

Questions?

Thank you!