

Introduction to Data Science for PI Data for PI Professionals

Ahmad Fattahi – Manager, Data Science Enablement, OSIsoft
Kleanthis Mazarakis – Data Scientist, OSIsoft

Objective

- Gain a basic understanding of data science process
- Understand how PI System fits in the picture

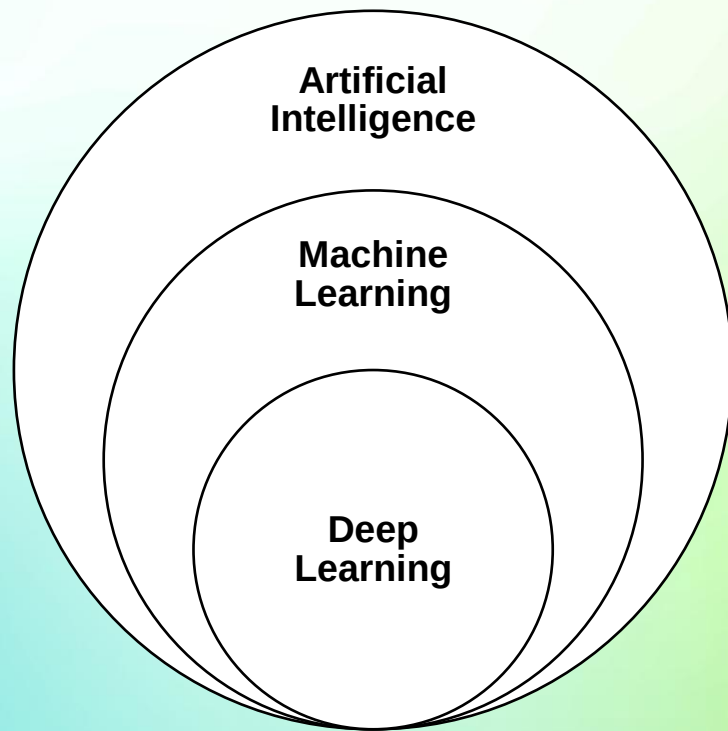
Agenda

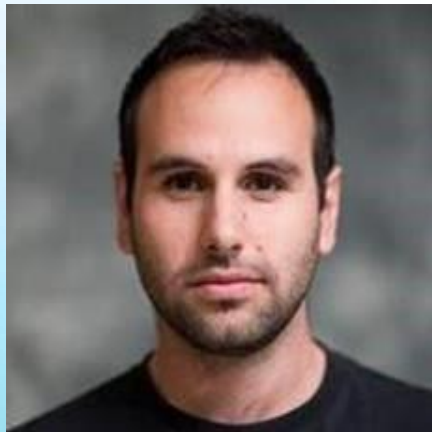
- Nomenclature
- A primer on data science practice
- CRISP-DM
- Data science for PI data: a case study

Nomenclature

Data Science is an interdisciplinary field of scientific methods, processes, algorithms and systems to extract knowledge or insights from data in various forms.

-wikipedia





- Q: Why did you become data scientists?
- A: Because “Superhero” is not a job title!

How do you tell myth from fact?

It's all about **VALUE!**

Data Science should be a core **organizational capability** as opposed to a technical practice.

Always start with a sharp question.

Not so sharp: how can we improve our bottom line?

Sharp: how can we save on energy to cool our building and what factors contribute to that?

Start Small

How Do You Pick the Right Analysis?

Exploratory

- What is the number of days in which the set point was never reached?
- What is the average power usage across all those days?

Predictive

- Which features can **predict** the power usage on a certain day?

Causal

- What **causes** power consumption to fluctuate?

Is the goal of the project to...



... predict?



... control?

What Functions Do You Need For Success?

Data Engineer

- Architecture, infrastructure, data governance, deployment

Data Scientist

- Cleaning, analysis and communication



Subject Matter Expert

- Intuition, what matters and what doesn't

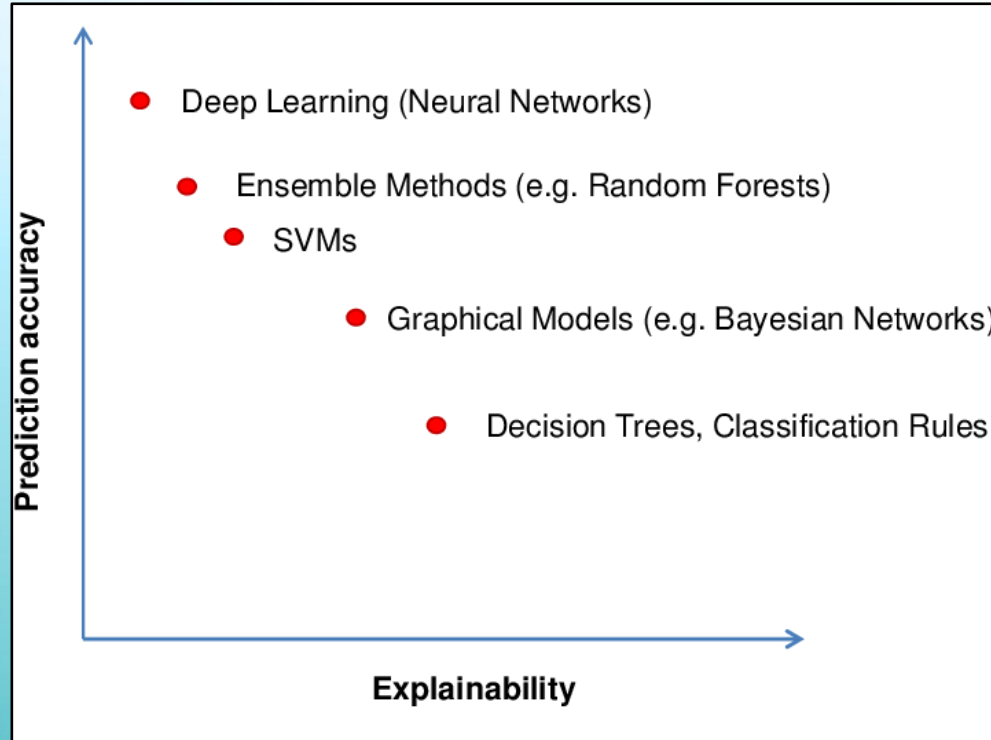
Data and Subject Matter Management

- Ownership, support, communication

Communication is key.

Explainability

The Explainability Tradeoff in Machine Learning



Source: ResearchGate GmbH

Reproducible Work Is the Differentiator

Assume your work is going to be repeated and tweaked frequently

Over time:

- Models veer off
- Physical systems change
- Priorities evolve
- New business owners come
- You get reassigned!

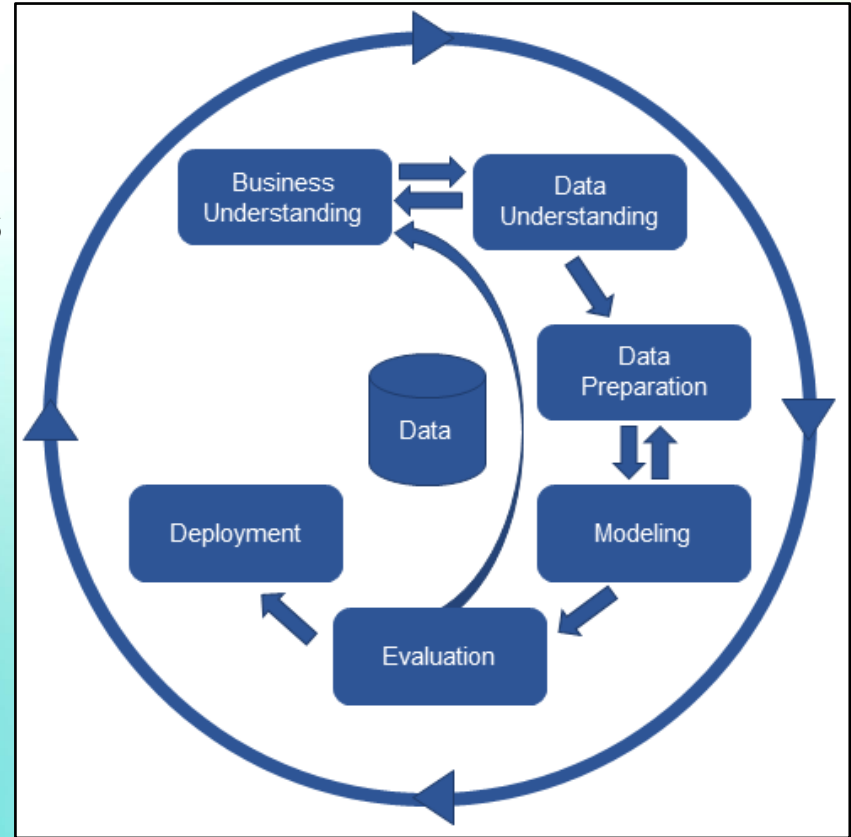
Leverage tools such as Jupyter Notebooks or other commercial platforms

Story: Optimize Building Energy Consumption



CRISP-DM

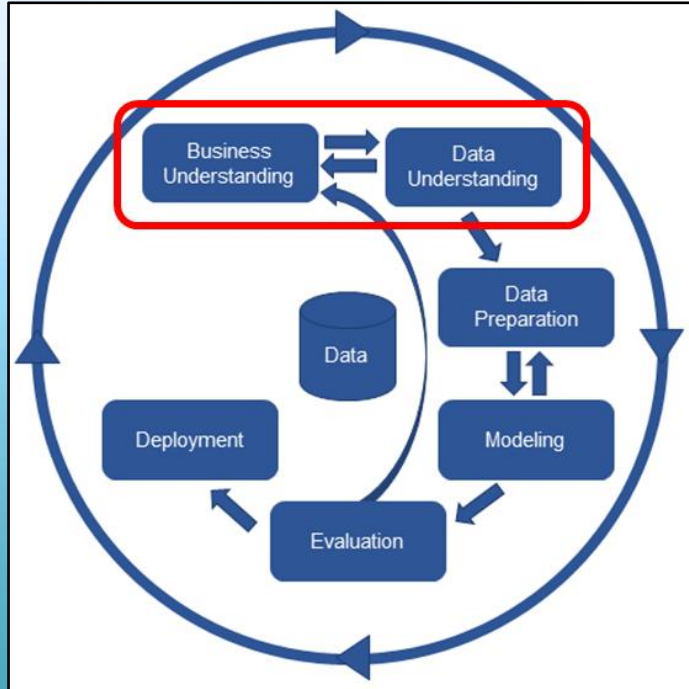
- **C**Ross **I**ndustry **S**tandard **P**rocess for **D**ata **M**ining
- Among most popular methodologies
- Emphasizes cycles and iterations



Source: KDnuggets

Case study: Interacting with PI System data

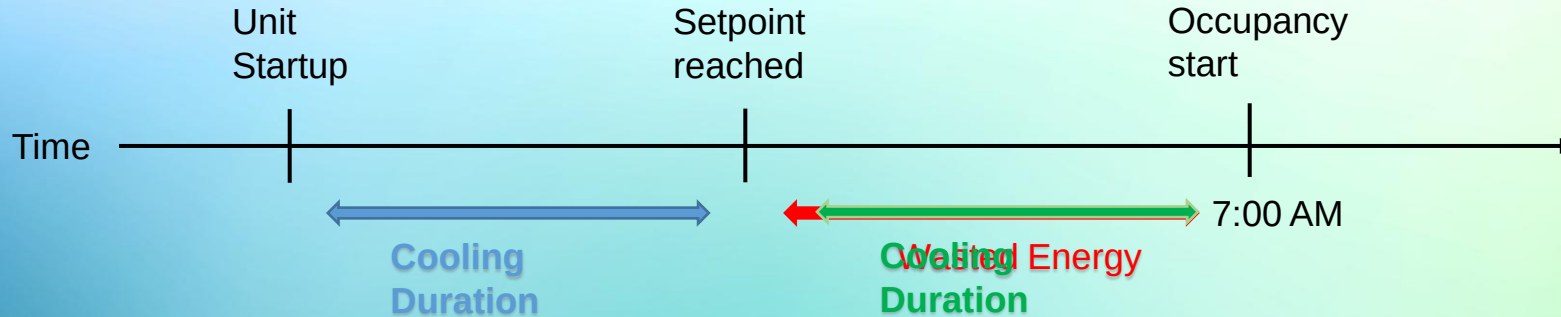
Business and Data Understanding



- What is the Business Objective?
- What data is available?
- What data is missing?

Case Study: Facilities Optimization

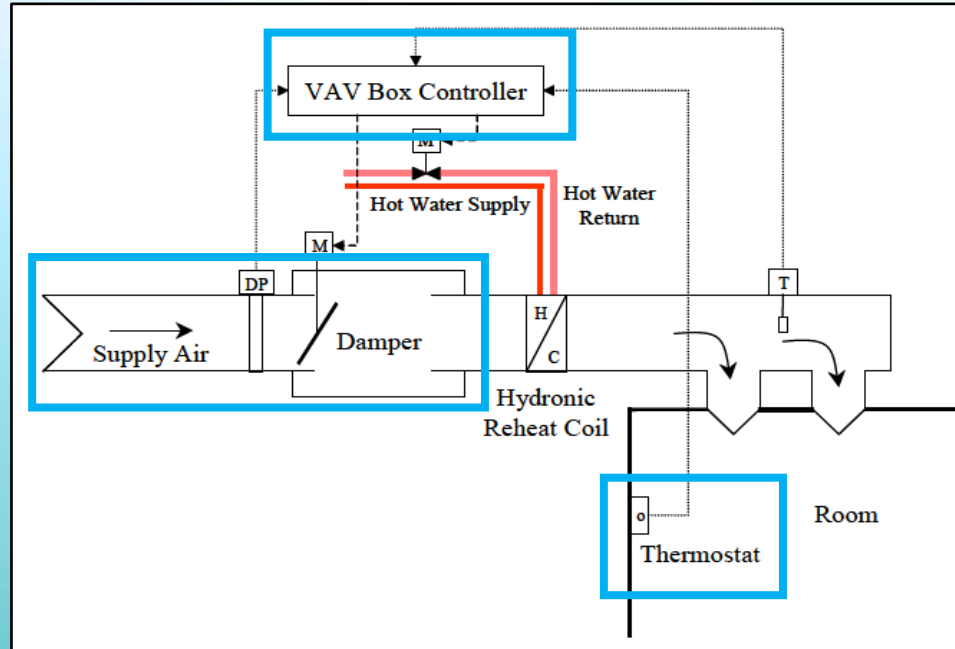
Improve Energy Efficiency of the building, by optimizing the startup of the **Variable Air Volume Cooling (VAVCO)** units



Value

- “If you saved 1 hour a day for 261 Working Days that is a significant amount of energy” – Subject Matter Expert

How does a VAV unit operate?



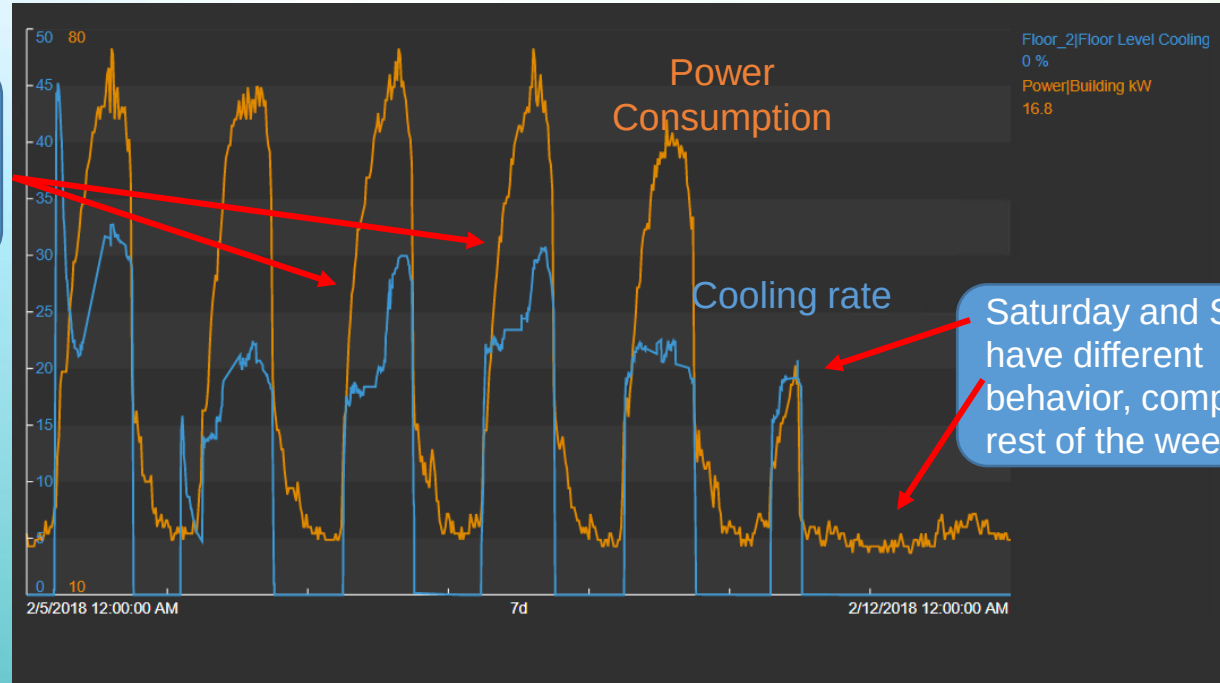
Available Data

The screenshot displays the OSIsoft PIWorld interface. On the left, a tree view shows the hierarchy of elements: Building, AC Units, Floor_2, Floor_3, Floor_4, Floor_5, Floor_6, VAVCO 6-06, VAVCO 6-07, VAVCO 6-08, VAVCO 6-09, VAVCO 6-10, VAVCO 6-11, Heating, IDF, Power, and Weather. The right pane shows the 'VAVCO 6-07' element details, with tabs for General, Child Elements, Attributes, Ports, Analyses, Notification Rules, and Version. The 'Attributes' tab is active, showing a table of data points categorized by 'Control'.

Name	Value
Category: <None>	
Attribute1	0
Cooling failure (alarm)	0
Cooling Failure (warning)	0
heating and cooling	0
Heating failure (alarm)	No Data
Heating Failure (warning)	No Data
VAV signal call conflict	0
Category: Control	
% cooling	0 %
% heating	0 %
% need more air signal (cooling)	0 %
% need more air signal (heating)	-0.799995422363281 %
Actual Airflow	40 ft3/min
Afterhour Status	0
Damper Command	35.5275382995605 %
Damper Position	35.444389343262 %
Desired Airflow	35 ft3/min
Force Damper Closed (cooling)	0
Force Damper Closed (heating)	0
Force Damper Open (cooling)	0
Force Damper Open (heating)	0
Lock Damper Position (cooling)	0
Lock Damper Position (heating)	0

Data understanding – Power Consumption

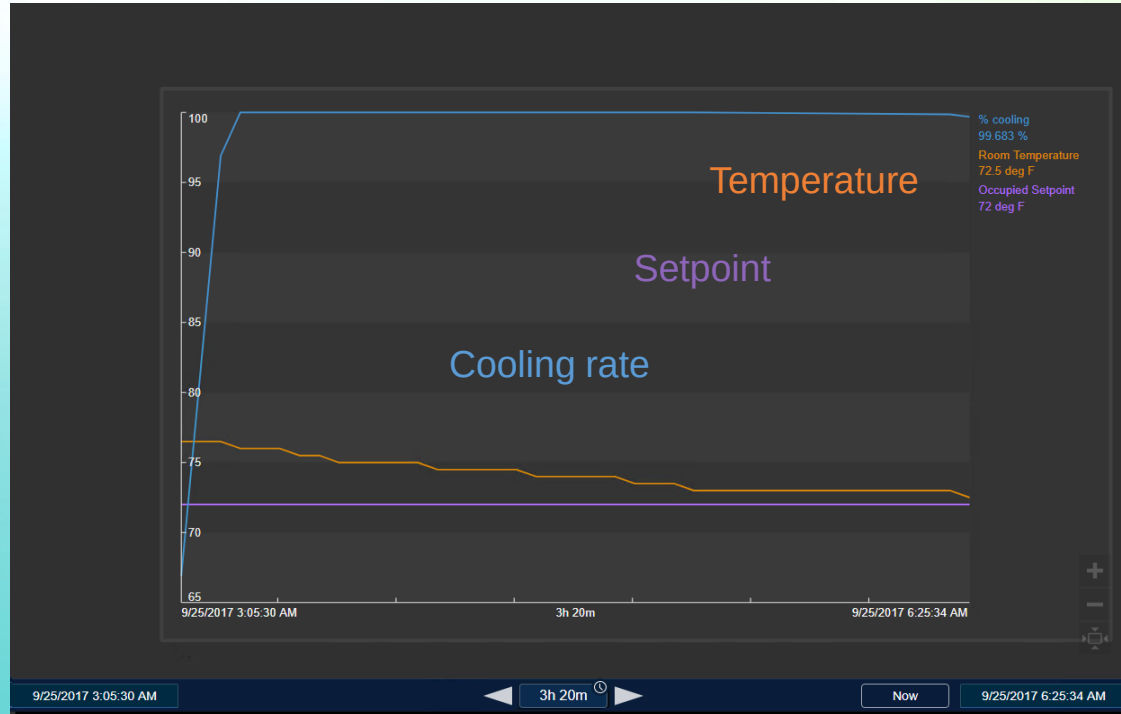
Daily power usage
aligned with **Average
% Cooling**



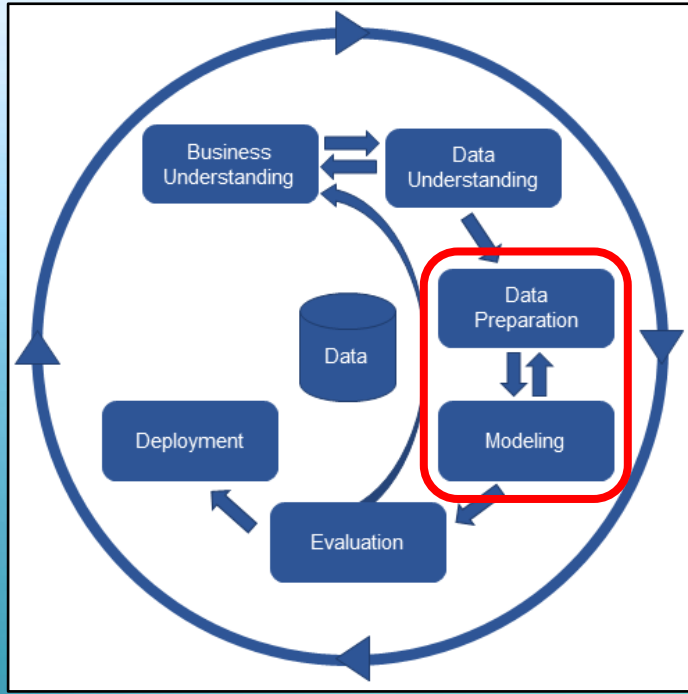
Saturday and Sunday
have different
behavior, compared to
rest of the week

Data understanding – Daily Cooling operation

- Morning startup event
- Feature extraction with Event Frames
- Predict Duration of Event Frame



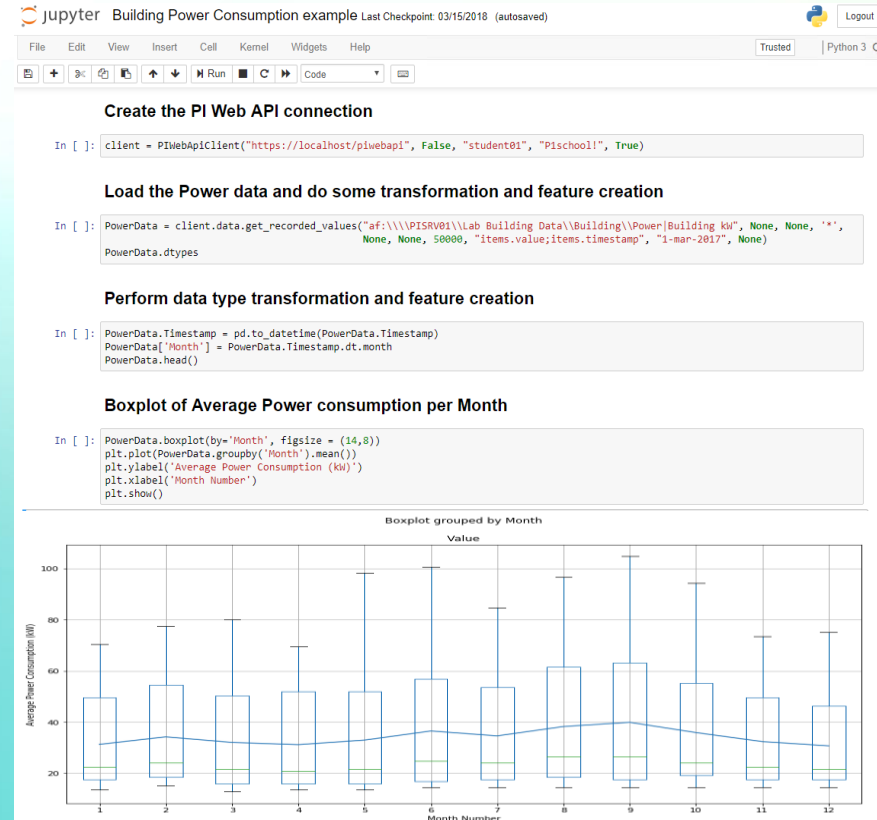
Data Preparation and Modelling



- How do I access PI Data from Advanced Analytics platforms?
- Which features are relevant for model prediction and can I identify any relationships between them?
- Which model should I use?

Client libraries for Data Science (Python & R)

- Client libraries for R and Python available in GitHub
 - <https://github.com/osimloeff/PI-Web-API-Client-R>
 - <https://github.com/osimloeff/PI-Web-API-Client-Python>
- Jupyter Notebooks - Data Exploration with Python
- Asset Analytics - MATLAB integration



Event Frames published using PI Integrator for Business Analytics

Web based UI

Data Cleansing

Add / Remove features

Data shaping

Select Data > Modify View > Publish

Back Next

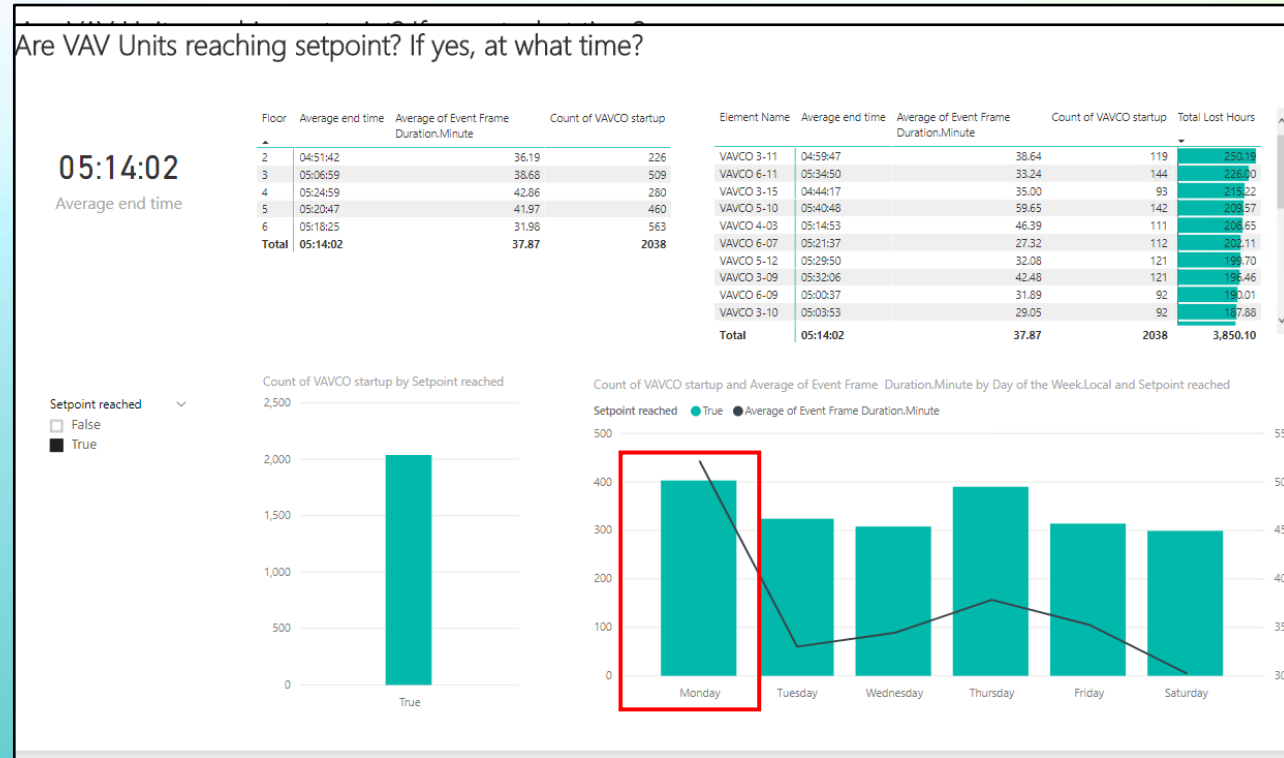
Apply

VAVCO startup	TimeStamp.Sta	TimeStamp.Enc	Month.Local	Day of the Wee	Event Frame D	% Cooling at Vi	Actual Airflow a	Damper Positio	Element Name	Outside Air Ten	Outside Relativ	Setpoint Offset	Setpoint reache	Space Humidi
VAVCO startu...	3/3/2017 7:0...	3/3/2017 7:2...	3	Friday	20	19.917	0	50	VAVCO 6-11	46.024	77.976	1	True	
VAVCO startu...	3/22/2017 7:...	3/22/2017 11:...	3	Wednesday	285	27.825	341	51.333	VAVCO 4-03	52.112	98.962	1.5	False	
VAVCO startu...	3/22/2017 7:...	3/22/2017 7:...	3	Wednesday	15	19.883	0	50	VAVCO 3-10	52.112	98.962	1	True	
VAVCO startu...	3/22/2017 7:...	3/22/2017 7:...	3	Wednesday	25	19.850	0	50	VAVCO 6-11	52.112	98.962	1	True	
VAVCO startu...	3/22/2017 7:...	3/22/2017 7:...	3	Wednesday	30	27.608	0	50	VAVCO 3-09	52.112	98.962	1.5	True	
VAVCO startu...	3/22/2017 7:...	3/22/2017 7:...	3	Wednesday	15	19.900	0	50	VAVCO 5-12	52.112	98.962	1	True	
VAVCO startu...	3/22/2017 7:...	3/22/2017 7:...	3	Wednesday	25	19.900	40	52.667	VAVCO 5-10	52.112	98.962	1	True	
VAVCO startu...	3/22/2017 7:...	3/22/2017 7:...	3	Wednesday	15	19.850	66	44	VAVCO 6-07	52.112	98.962	1	True	
VAVCO startu...	3/23/2017 7:...	3/23/2017 7:...	3	Thursday	10	19.900	96	35.333	VAVCO 6-11	45.794	85.969	1	True	
VAVCO startu...	3/24/2017 7:...	3/24/2017 7:...	3	Friday	10	19.850	343	24.556	VAVCO 4-03	52.999	63.596	1	True	
VAVCO startu...	3/24/2017 7:...	3/24/2017 7:...	3	Friday	10	19.850	94	41.444	VAVCO 6-11	53.000	63.597	1	True	
VAVCO startu...	3/24/2017 7:...	3/24/2017 7:...	3	Friday	10	19.900	79	33.333	VAVCO 3-09	53.000	63.597	1	True	
VAVCO startu...	3/24/2017 7:...	3/24/2017 7:...	3	Friday	10	19.850	87	44.556	VAVCO 6-07	53.000	63.598	1	True	
VAVCO startu...	3/25/2017 7:...	3/25/2017 7:...	3	Saturday	15	19.900	0	50	VAVCO 5-12	51.169	86.673	1	True	
VAVCO startu...	3/25/2017 7:...	3/25/2017 7:...	3	Saturday	25	19.850	38	52	VAVCO 5-10	51.169	86.673	1	True	
VAVCO startu...	3/25/2017 7:...	3/25/2017 8:...	3	Saturday	110	27.475	361	51.333	VAVCO 4-03	51.169	86.671	1.5	False	
VAVCO startu...	3/25/2017 7:...	3/25/2017 7:...	3	Saturday	15	19.850	0	50	VAVCO 3-09	51.169	86.671	1	True	
VAVCO startu...	3/27/2017 7:...	3/27/2017 7:...	3	Monday	5	19.625	69	48.444	VAVCO 5-12	49.936	83.103	1	True	
VAVCO startu...	3/27/2017 7:...	3/27/2017 7:...	3	Monday	30	27.058	47	37.556	VAVCO 5-10	49.936	83.103	1.5	True	
VAVCO startu...	3/27/2017 7:...	3/27/2017 8:...	3	Monday	115	20.25	313	23.556	VAVCO 4-03	49.935	83.103	1	False	
VAVCO startu...	3/27/2017 7:...	3/27/2017 7:...	3	Monday	20	19.850	60	33.778	VAVCO 3-09	49.935	83.102	1	True	
VAVCO startu...	3/27/2017 7:...	3/27/2017 7:...	3	Monday	15	19.850	105	47	VAVCO 6-11	49.935	83.102	1	True	
VAVCO startu...	3/28/2017 7:...	3/28/2017 7:...	3	Tuesday	10	19.850	101	47.333	VAVCO 5-12	48.355	80.228	1	True	

Data Exploration – Microsoft Power BI

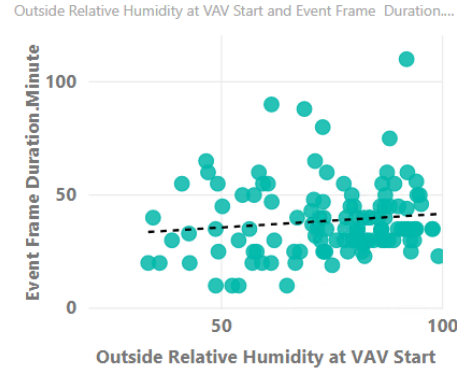
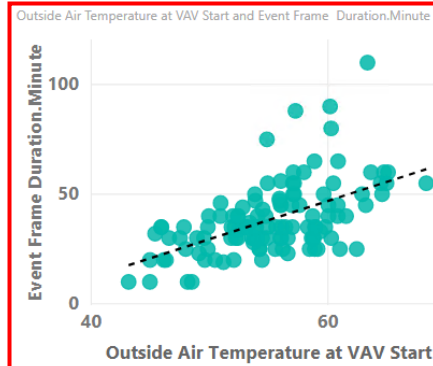
- **Target Average end time is 7 AM**
➔ **1.5 hours of wasted Energy**

- **Monday** has significantly higher Average Duration



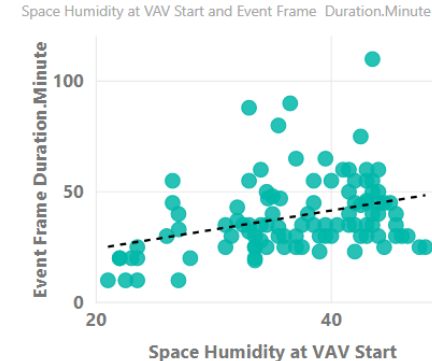
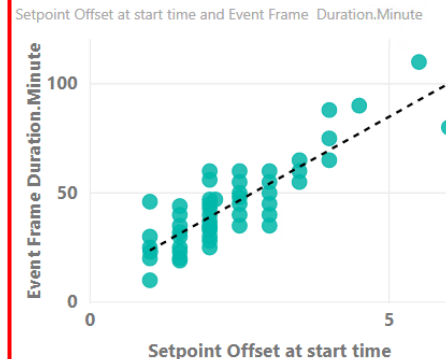
Bivariate Analysis – Microsoft Power BI

Bivariate Analysis



Floor
■ 3

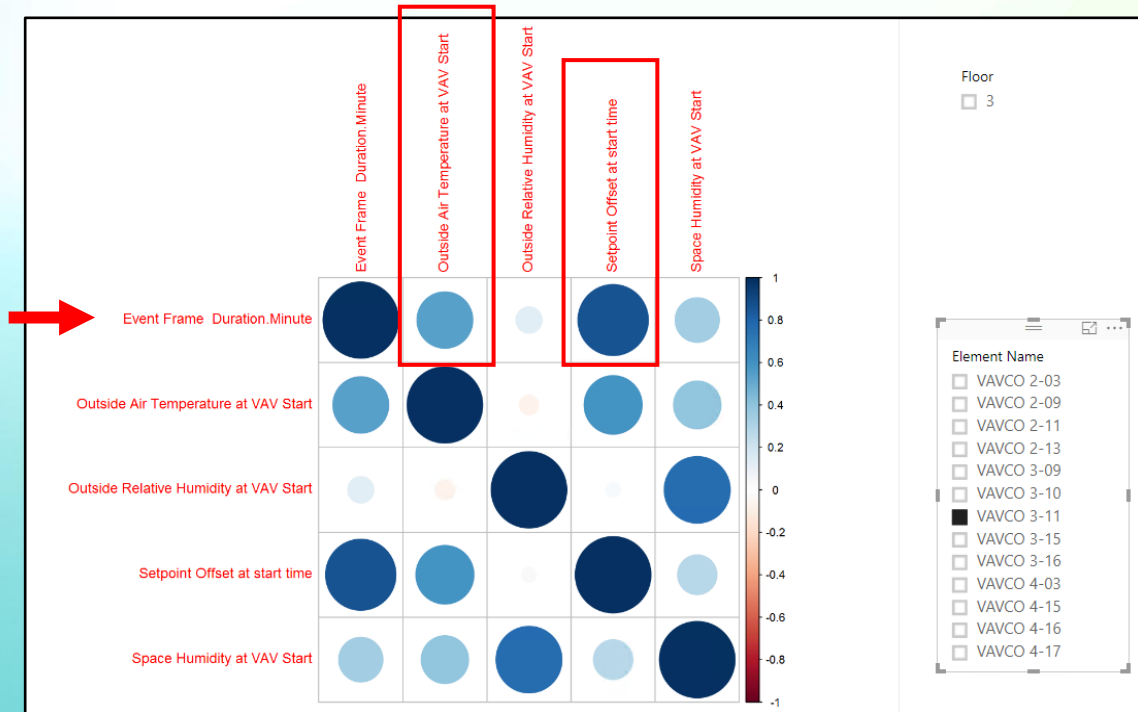
Element Name
□ VAVCO 3-09
□ VAVCO 3-10
■ VAVCO 3-11
□ VAVCO 3-15
□ VAVCO 3-16



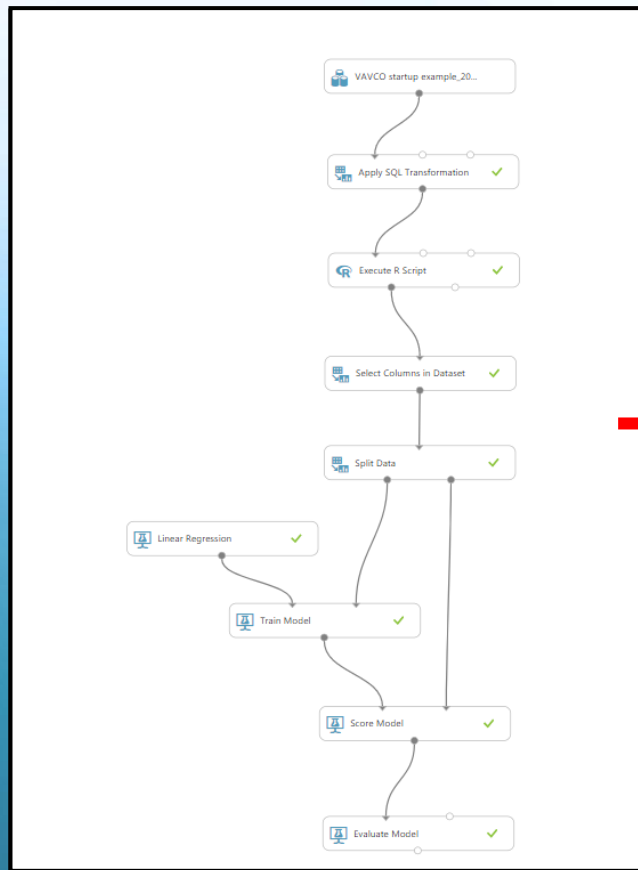
Correlation plot embedded in Microsoft Power BI

Confirms
Bivariate Analysis
Results

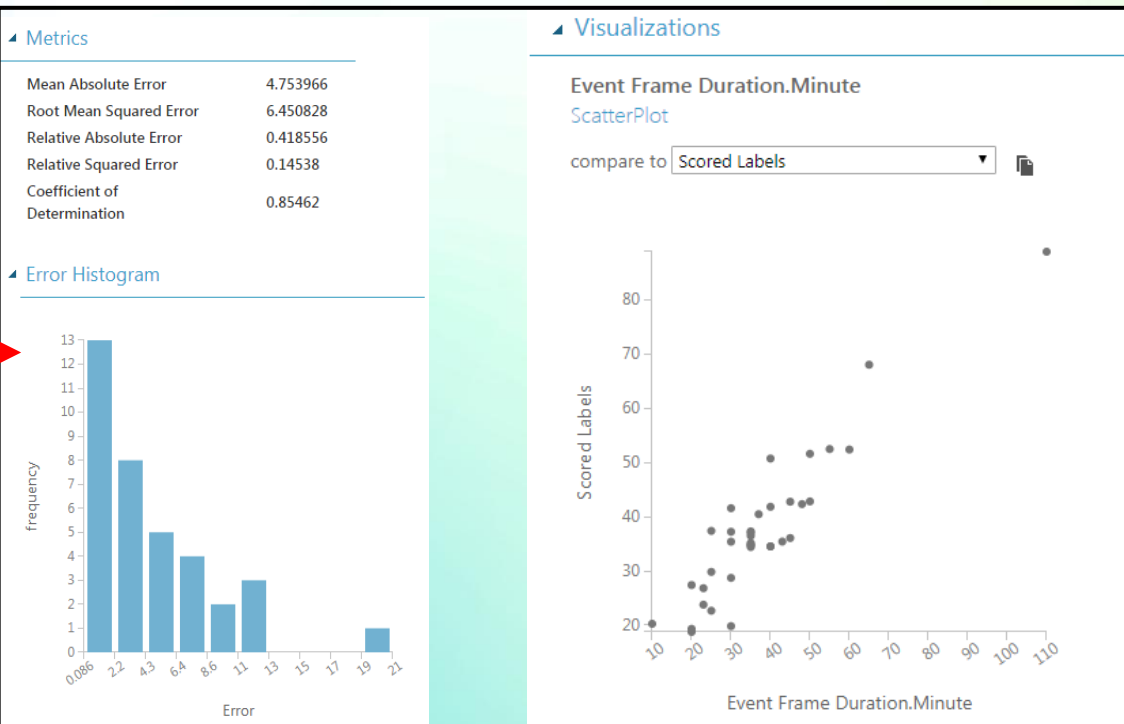
Strong correlation
of **Duration** with
Setpoint Offset
and **Outside
Temperature**



Azure ML model



Results



Deployment

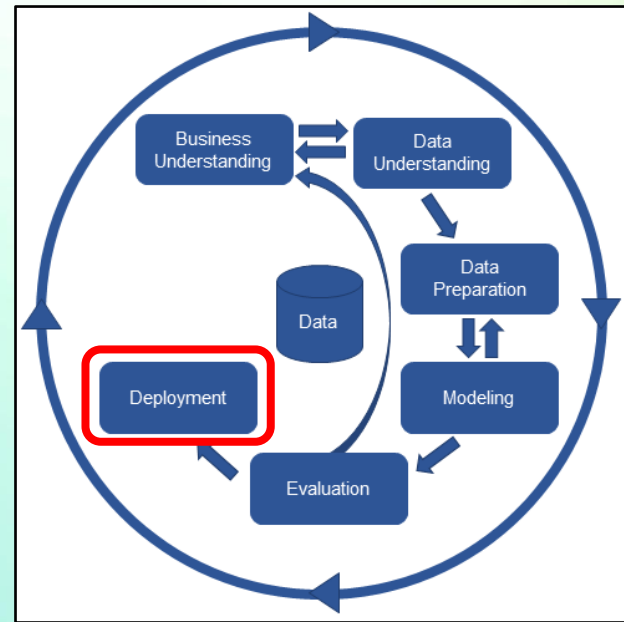
Productizing the model

- Have the model control the wakeup time of the building

Simpler models can be deployed in PI; some control models are built into the control network

- Regression or small Decision Trees can be efficiently deployed in AF
- Larger or more complex models need to go outside PI

Work in progress



Main Takeaways



Data Science delivers **Value**



Learn the process



Iterative process – Don't get discouraged



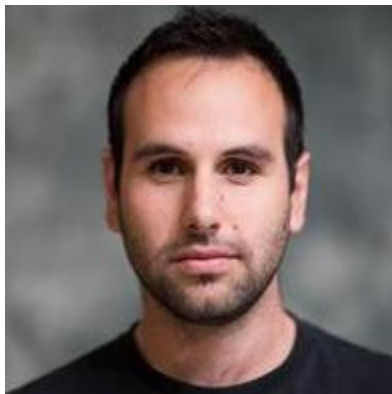
Leverage PI System And Subject Matter Expertise



Ahmad Fattahi

afattahi@osisoft.com

Manager, Data Science Enablement
OSIsoft



Kleanthis Mazarakis

kmazarakis@osisoft.com

Data Scientist
OSIsoft

Questions?

Please wait for
the **microphone**

State your
name & company



Please rate this session in the mobile app!

**DOWNLOAD
THE MOBILE APP**

- Rate sessions and provide feedback
- Meet and connect with other attendees

Join the conversation and SHARE
what you saw #osisoft #piworld

A black smartphone displaying the OSiSoft PIWorld app logo on its screen. The logo consists of a stylized white 'X' shape above the text 'OSiSoft. PIWorld'.

