Advanced Analytics for PI Data for Data Scientists

Ahmad Fattahi – Manager, Data Science Enablement, OSIsoft

Dallas Swift – Data Scientist, OSIsoft

 ©2018 OSIsoft, LLC

Goal: Gain a better understanding of data science practices for process data and the PI System

Agenda

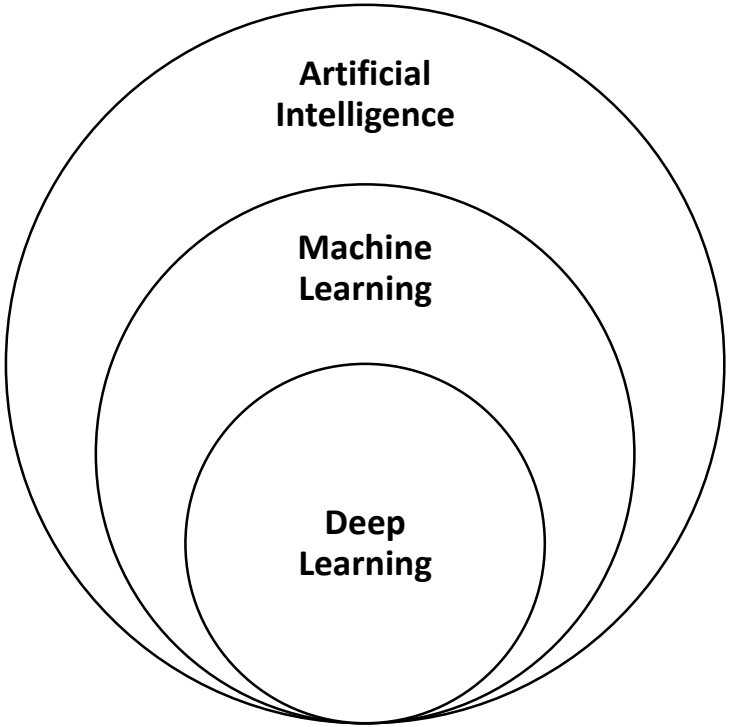
- Definitions and general concepts
- CRISP-DM Process
- Best practices and pitfalls
- Case study

©2018 OSIsoft, LLC

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



©2018 OSIsoft, LLC

CRISP-DM

- CRoss Industry Standard Process for Data Mining
- Among most popular methodologies
- Emphasizes cycles and iterations

Source: KDnuggets

©2018 OSIsoft, LLC

Story: Optimize Building Energy Consumption

©2018 OSIsoft, LLC

Inception: Management or SME

Start from a "Sharp Question"

- *"Can the building wake up later?"*

Business owner plays a key role

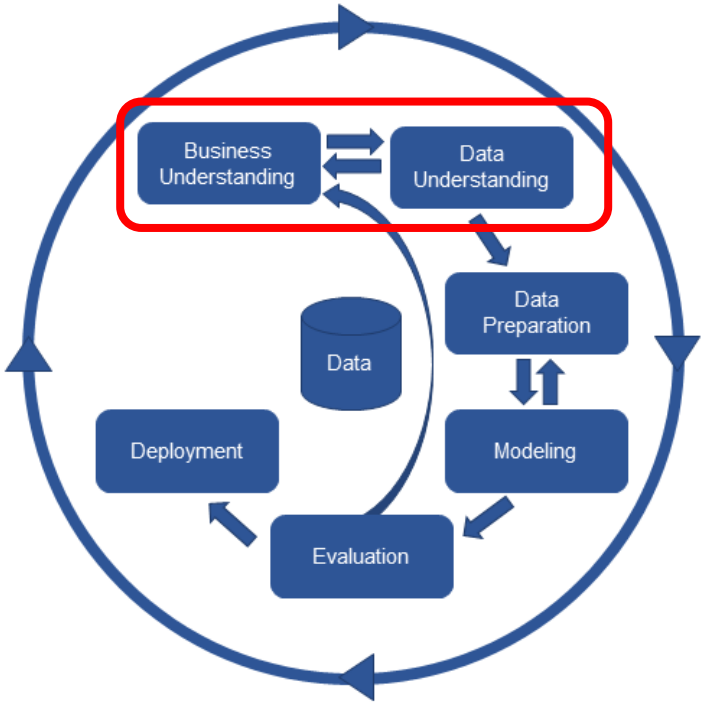
- *Facilities Manager*

Envision the delivery mechanism

- *"Recommendation engine? Direct control?"*

SME and data professionals start engaging

- Many conversations until they speak the same language



©2018 OSIsoft, LLC

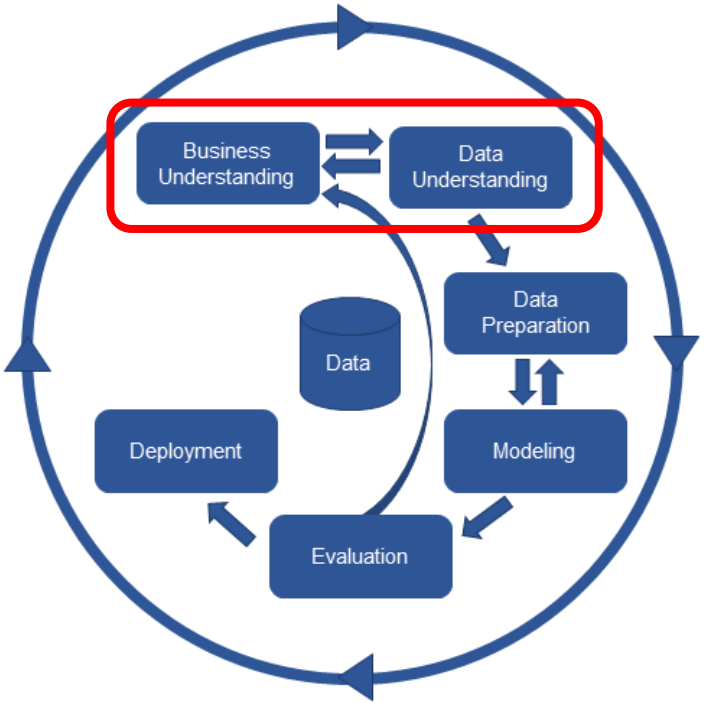
Pitfalls

Myth: The data scientist can do it all!

Targeting the wrong question

Losing sight of bottom line value to the business

Getting crushed between political gears



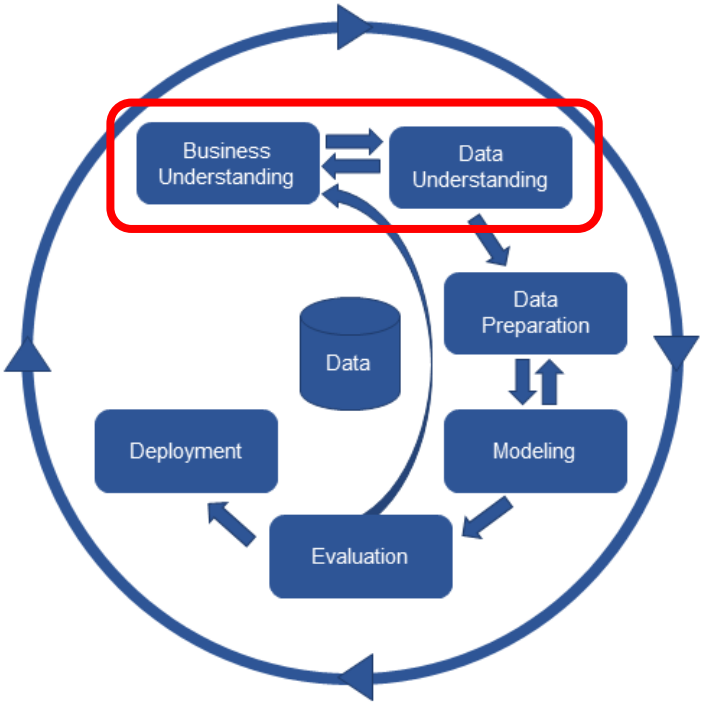
©2018 OSIsoft, LLC

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

©2018 OSIsoft, LLC

Pitfalls

- Myth: The data scientist can do it all!
- Targeting the wrong question
- Losing sight of bottom line value to the business
- Getting crushed between political gears



©2018 OSIsoft, LLC

Building the “Model”

Engage with data engineers, PI Admins

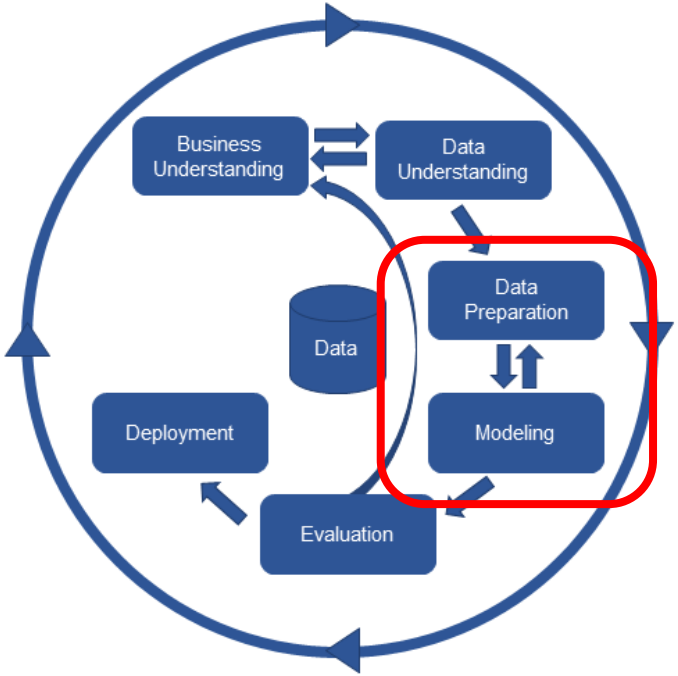
- Python and R libraries by OSIsoft, PI Web API, AF SDK, PI Integrators, PI SQL libraries

Build the features and the model

- Some features can be built in PI

Constantly ask for validation from the SME

- Does it make sense?



©2018 OSIsoft, LLC

Process Data Can Be Significantly Different!

Features typically have to be engineered from raw data

It is usually not the traditional “time-series” analysis

PI System can do a lot!

- Raw, summarized, or interpolated data
- Event Frames
- Hierarchy in AF is crucial

SME plays a key role

©2018 OSIsoft, LLC

Is the goal of the project to…

… predict?

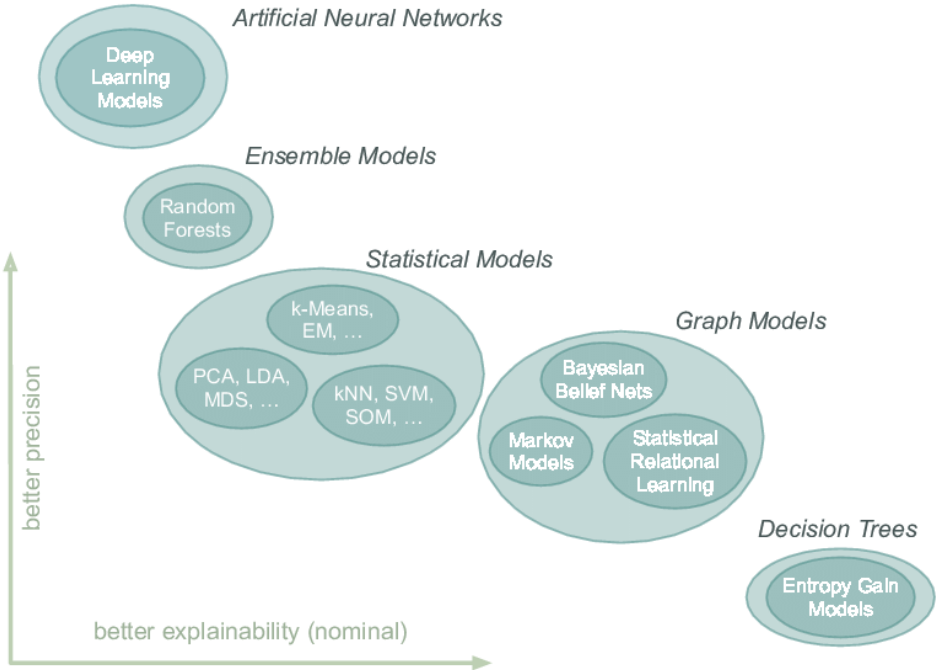
… control?

©2018 OSIsoft, LLC

Explainability

©2018 OSIsoft, LLC

Tradeoff

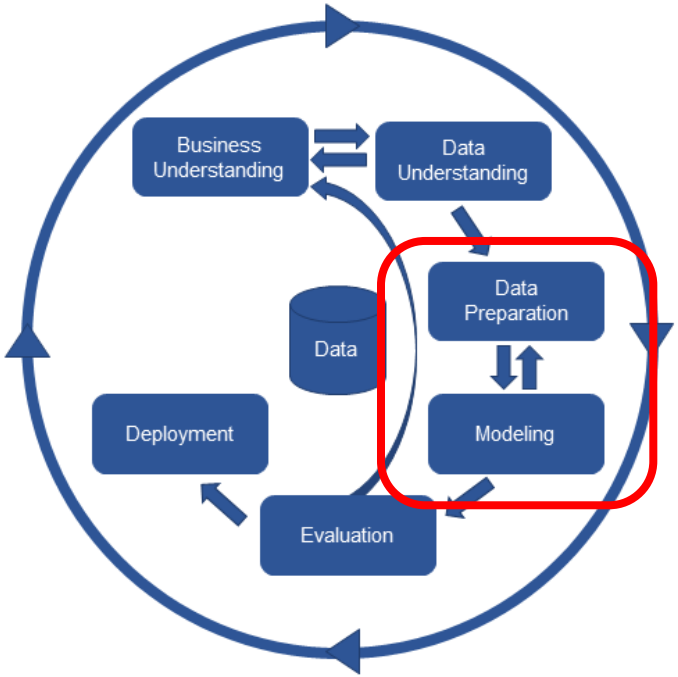


Source: ResearchGate GmbH

©2018 OSIsoft, LLC

Pitfalls – Veering off the process

- Building model for something uncontrollable
- Mixing correlation with causation
- Not including data engineering concerns for deployment
- Not leveraging PI capabilities in feature engineering



©2018 OSIsoft, LLC

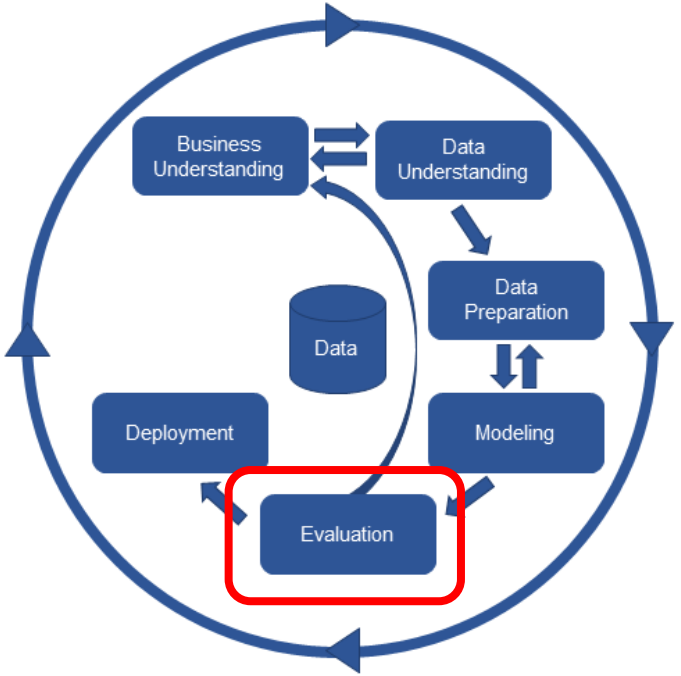
Evaluation – Loop back with the Business

Guarantees we answered the right question

Forces us to measure real value, often in dollars, man-hours, or other tangible resources

Not trivial!

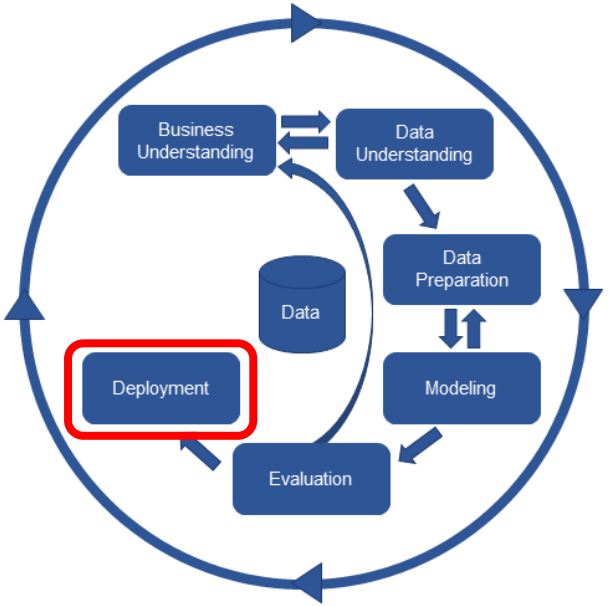
Caution: data scientists speak a different language than process people



©2018 OSIsoft, LLC

Deployment – Data Engineers Are Key

- Productizing the model
- Simpler models can be deployed in PI; some control models are built into the control network
- Consult with PI Admins and Data Engineers early
- Data Governance can pose challenges in production



©2018 OSIsoft, LLC

Reproducible Work Is the Differentiator

Assume your work is going to be repeated and tweaked frequently

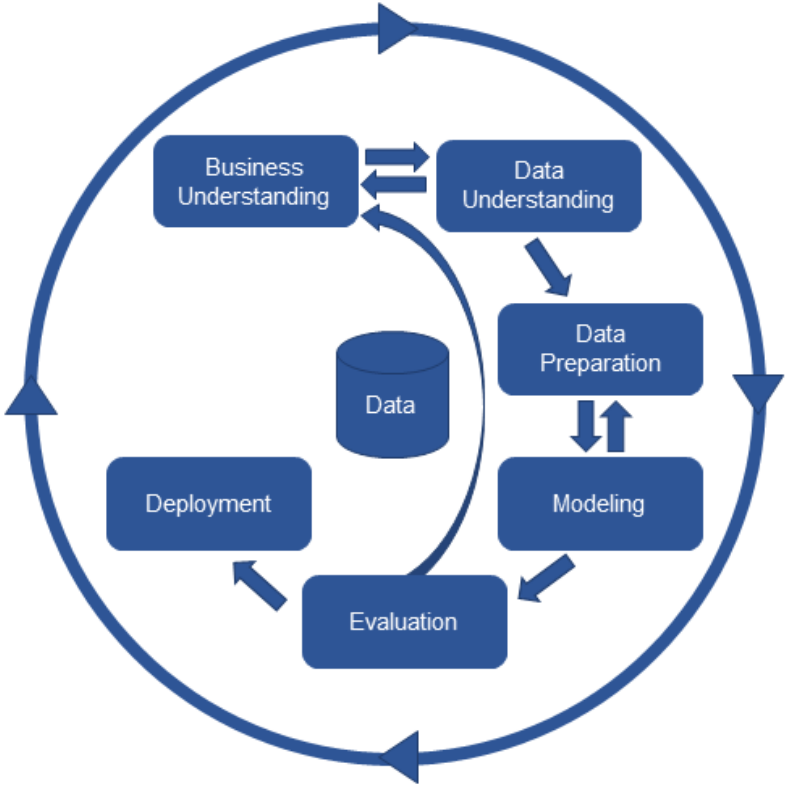
Over time:

- 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

©2018 OSIsoft, LLC

The Cycle Repeats



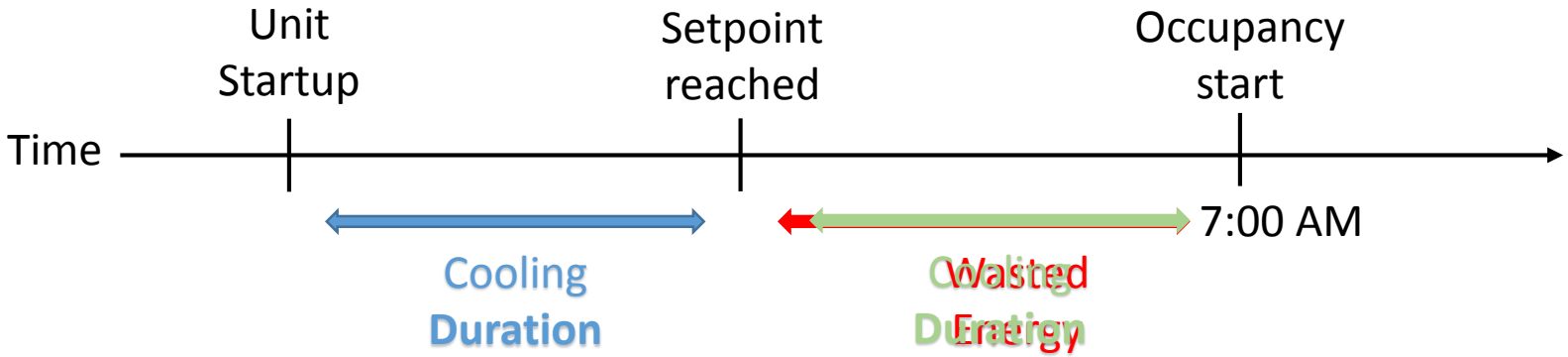
©2018 OSIsoft, LLC

Case study: Interacting with PI System data

©2018 OSIsoft, LLC

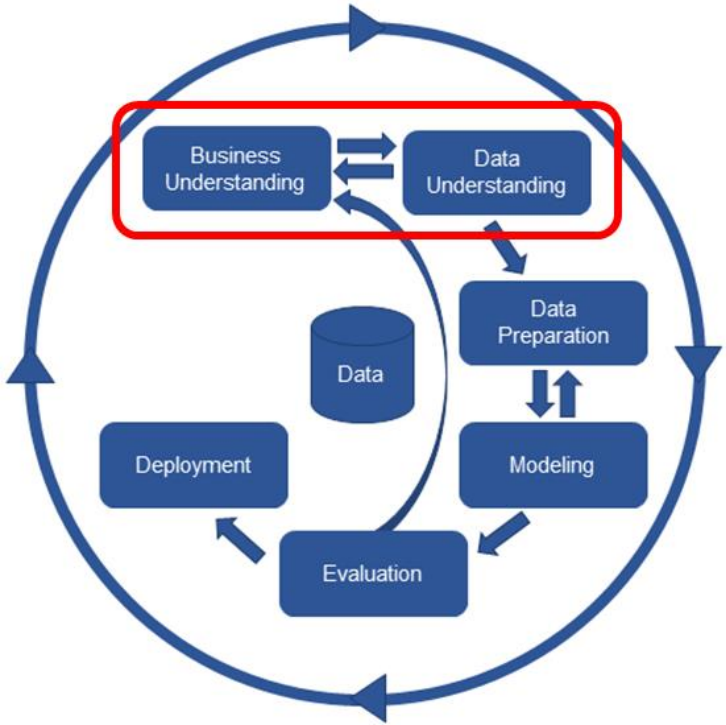
Reduce wasted cooling energy

Optimize the startup of the **Variable Air Volume Cooling (VAVCO)** units to improve the building's energy efficiency



©2018 OSIsoft, LLC

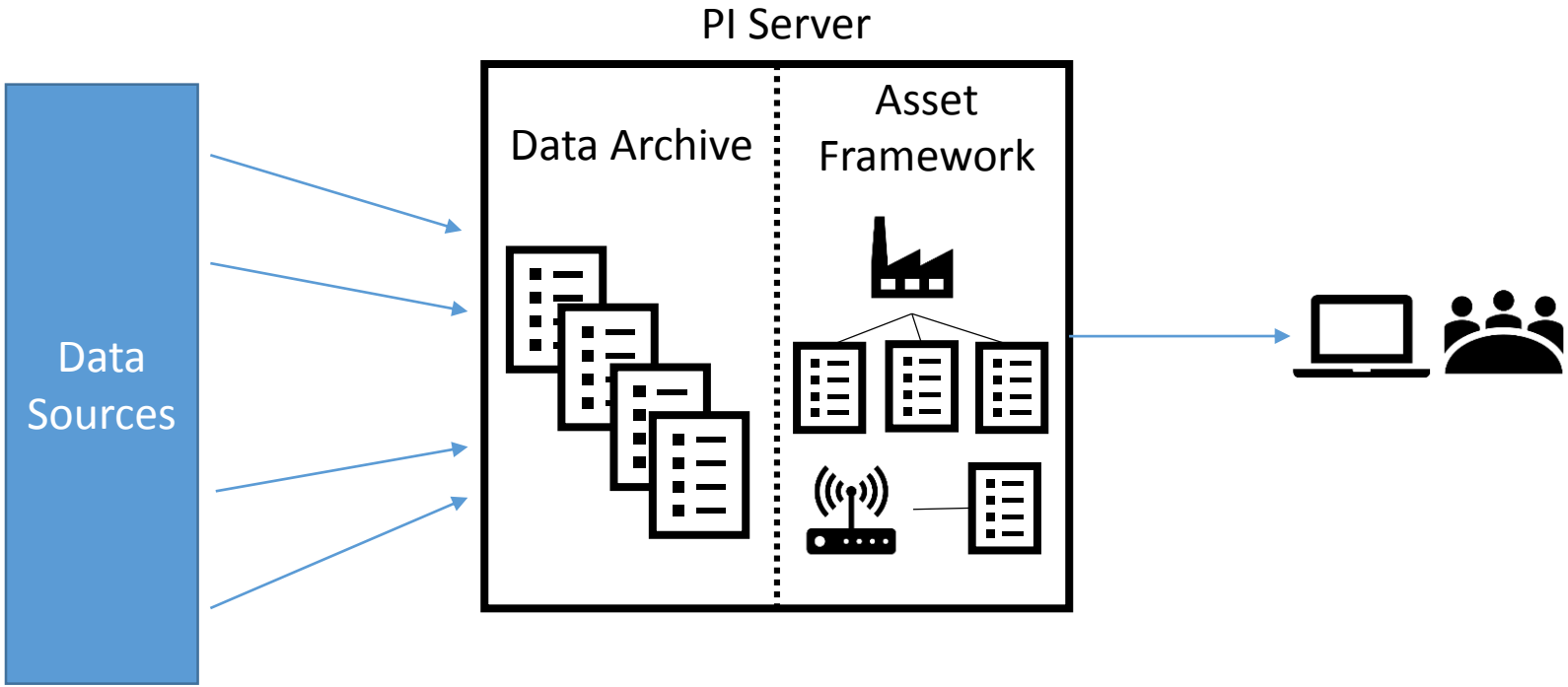
Setting ourselves up for success



- How are the data streams structured?
- How do the data behave?
- What information is relevant for the problem?

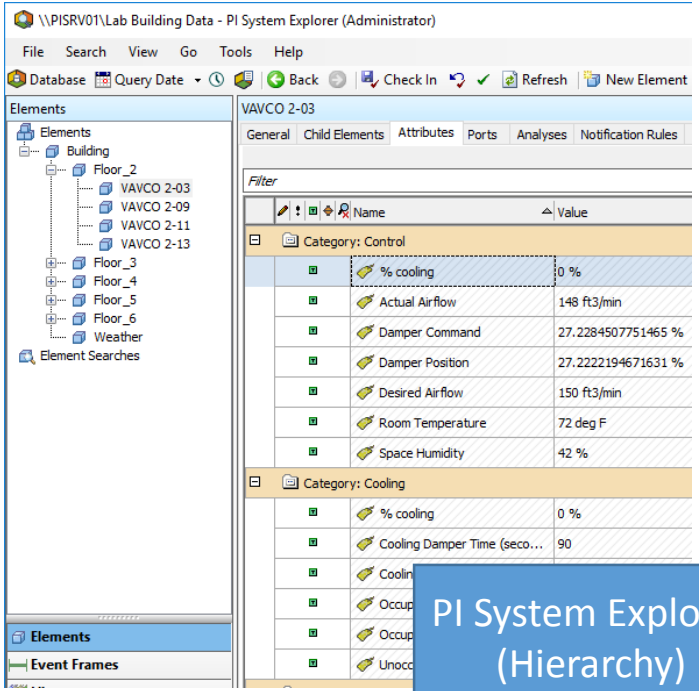
©2018 OSIsoft, LLC

Understanding the Asset Framework

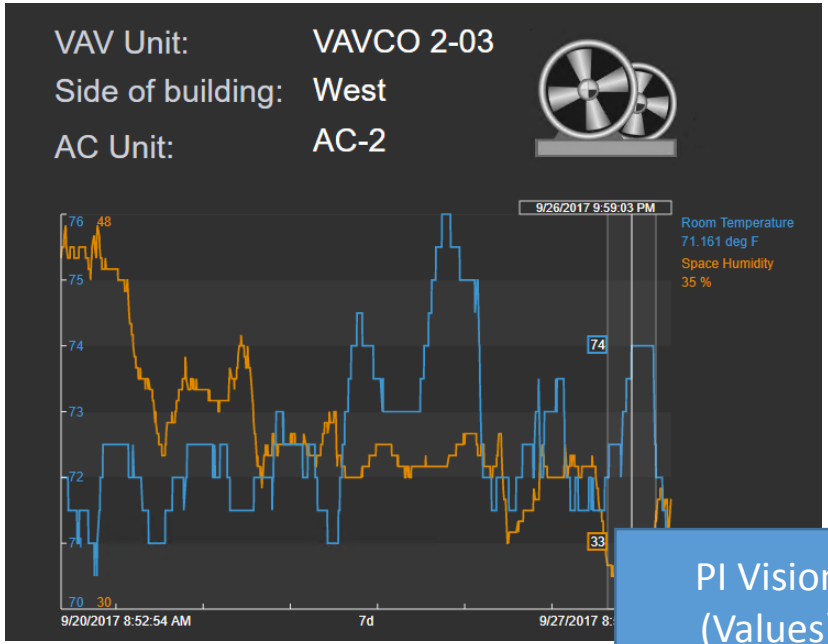


©2018 OSIsoft, LLC

Explore hierarchy and trends



PI System Explorer (Hierarchy)

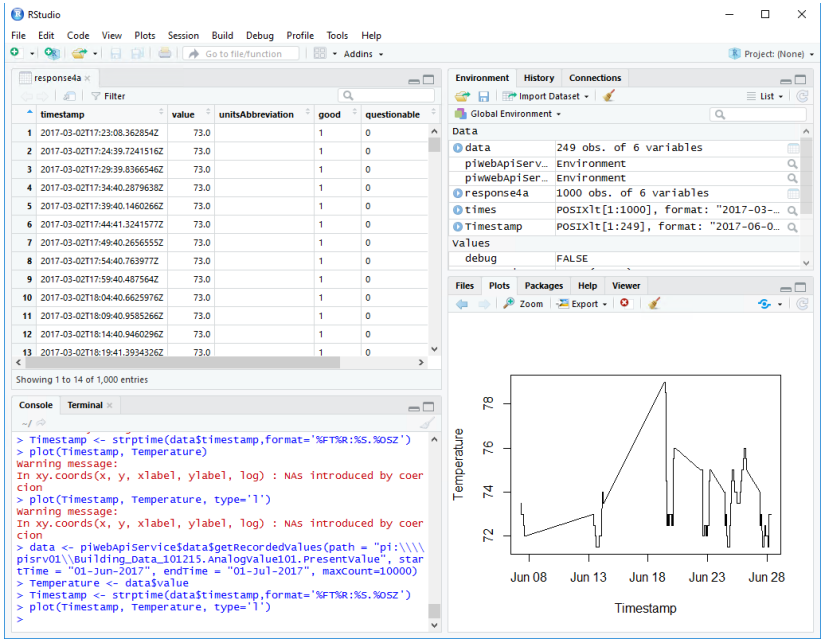


PI Vision (Values)

©2018 OSIsoft, LLC

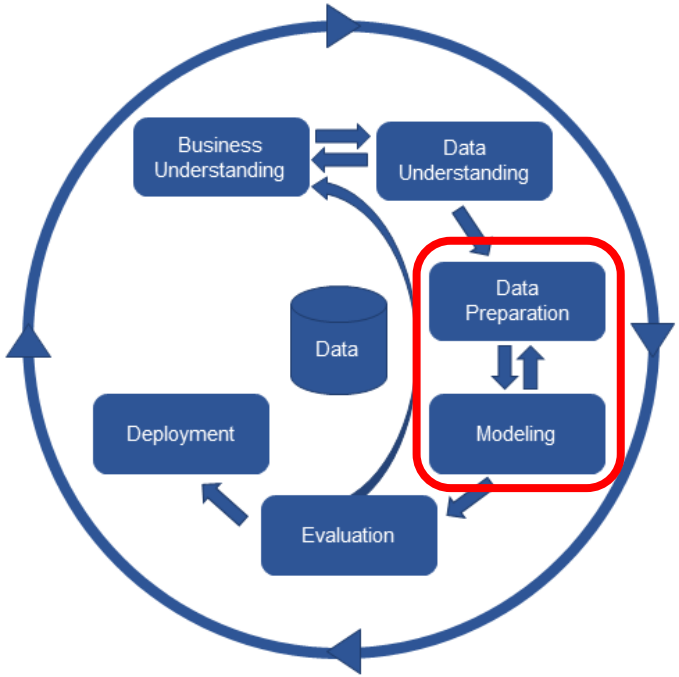
Leveraging data science tools

- Data science tools are great for data exploration
- R and Python libraries that use PI Web API are available via PI Developers Club
  - https://github.com/osimloeff/PI-Web-API-Client-R
  - https://github.com/osimloeff/PI-Web-API-Client-Python



©2018 OSIsoft, LLC

Transforming data to information



- How should I aggregate time-series data?
- Which features are relevant for model prediction?
- How can I make the data available for modeling?

©2018 OSIsoft, LLC

Time series data are complex!

VAVCO-1

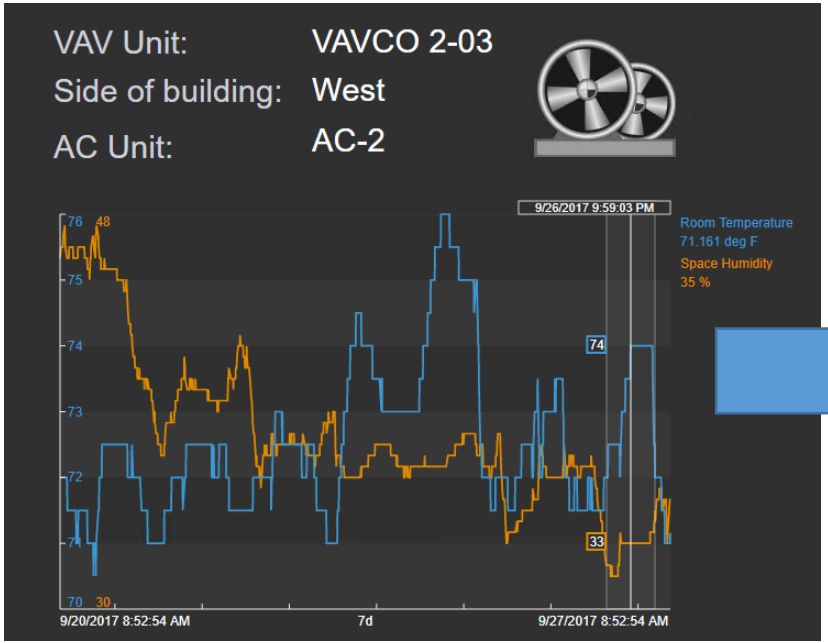
Temperature
Air flow
Humidity

VAVCO-2

Temperature
Air flow
Humidity
CO_2

©2018 OSIsoft, LLC

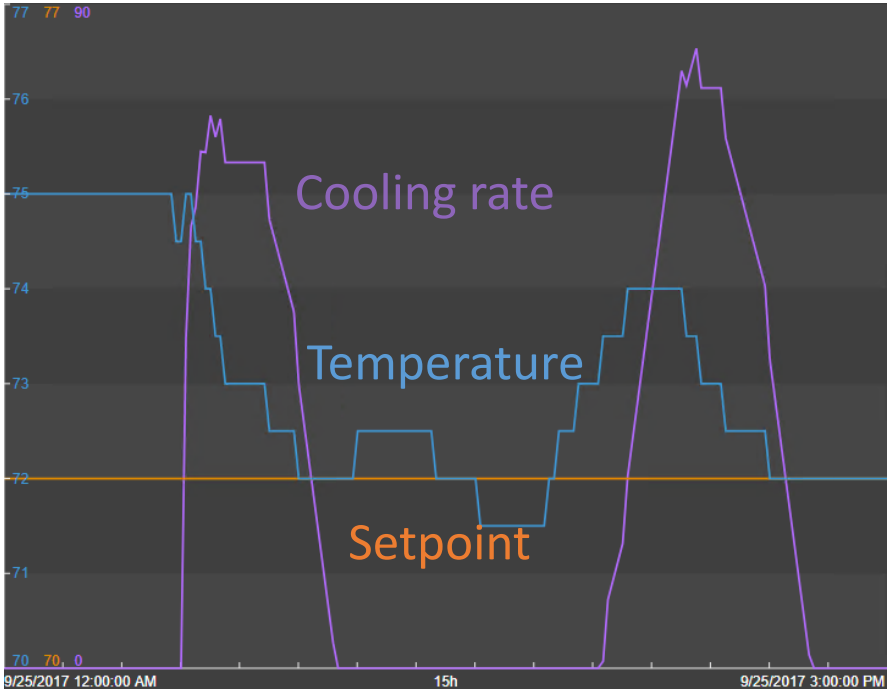
Need to shape and export data



% Cooling at VAV Start	Actual Airflow at VAV Start	Damper Position at VAV Start	Element Name	C
19.917	0	50	VAVCO 6-11	4
19.85	0	50	VAVCO 6-11	5
19.883	0	50	VAVCO 3-10	5
27.608	0	50	VAVCO 3-09	5
19.9	0	50	VAVCO 5-12	5
19.9	40	52.667	VAVCO 5-10	5
19.85	66	44	VAVCO 6-07	5
19.9	96	35.333	VAVCO 6-11	4
[illegible]5	343	24.556	VAVCO 4-03	5
[illegible]	94	41.444	VAVCO 6-11	5
[illegible]	79	33.333	VAVCO 3-09	5
[illegible].85	87	44.556	VAVCO 6-07	5
19.9	0	50	VAVCO 5-12	5
19.85	38	52	VAVCO 5-10	5
19.85	0	50	VAVCO 3-09	5

©2018 OSIsoft, LLC

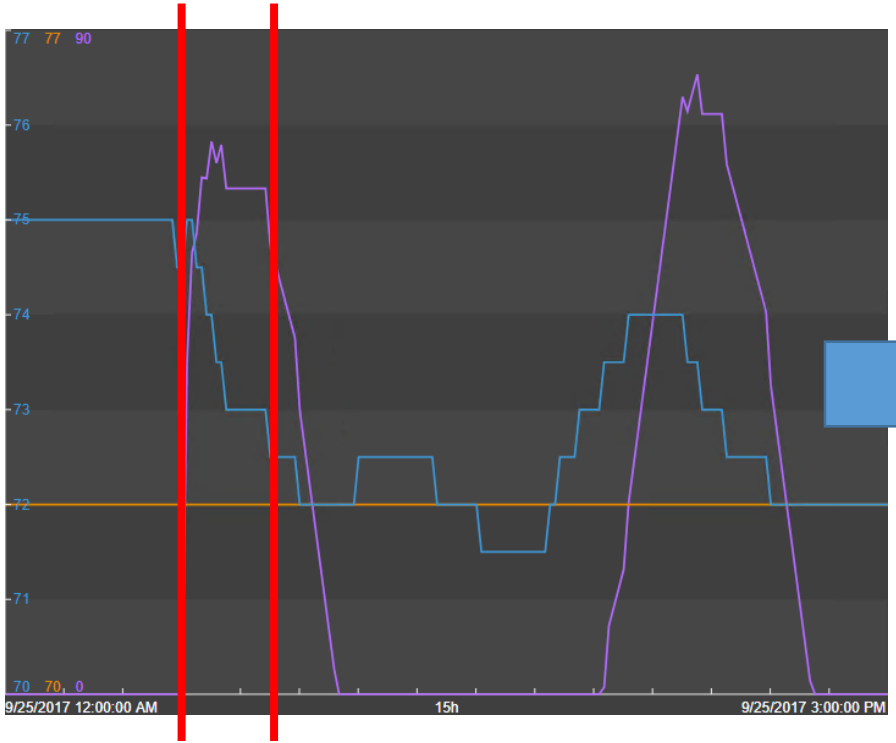
Labeling the data – easy, right?



- Separate first cooling period of day from others
- When is a cooling period finished?
- Typical process data issues (data alignment, gaps, etc.)

©2018 OSIsoft, LLC

Event Frames help aggregate data



VAVCO startup - VAVCO 5-10 - 2017-09-25 03:05:30.938

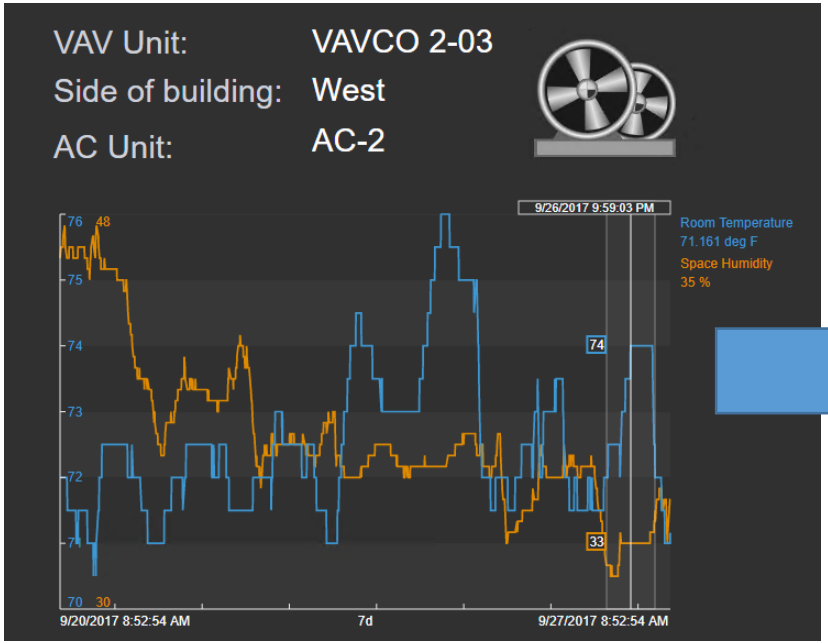
General | Child Event Frames | Referenced Elements | Attributes

Filter

Name	Value
% Cooling at VAV Start	66.8999328613281 %
Actual Airflow at VAV Start	51 ft3/min
Damper Position at VAV Start	54.8888893127441 %
Element Name	VAVCO 5-10
Outside Air Temperature at VAV S...	60.4192047119141 °F
Outside Relative Humidity at VAV ...	60.2470664978027 %
Room Temperature at VAV Start	76.5 °F
Room Temperature when setpoint ...	72.5 °F
Setpoint at VAV Start	72 °F
Setpoint Offset at end time	0.5
Setpoint Offset at start time	4.5
Setpoint reached	True
Setpoint when setpoint reached	72 °F
Space Humidity at VAV Start	36.5 %

©2018 OSIsoft, LLC

Data ready to go into model

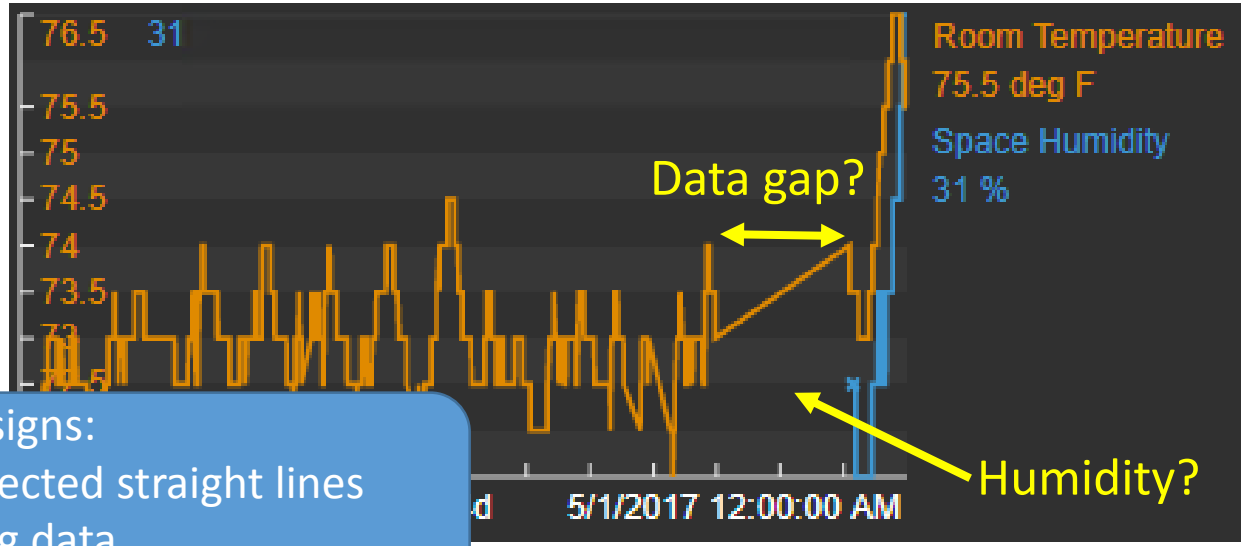


% Cooling at VAV Start	Actual Airflow at VAV Start	Damper Position at VAV Start	Element Name	C
19.917	0	50	VAVCO 6-11	4
19.85	0	50	VAVCO 6-11	5
19.883	0	50	VAVCO 3-10	5
27.608	0	50	VAVCO 3-09	5
19.9	0	50	VAVCO 5-12	5
19.9	40	52.667	VAVCO 5-10	5
19.85	66	44	VAVCO 6-07	5
19.9	96	35.333	VAVCO 6-11	4
[illegible]	343	24.556	VAVCO 4-03	5
[illegible]	94	41.444	VAVCO 6-11	5
[illegible]	79	33.333	VAVCO 3-09	5
[illegible]	87	44.556	VAVCO 6-07	5
19.9	0	50	VAVCO 5-12	5
19.85	[illegible]	[illegible]	[illegible]10	5
19.85	[illegible]	[illegible]	[illegible]9	5

- PI Integrator for Business Analytics
- PI OLEDB Enterprise
- Custom AF SDK

©2018 OSIsoft, LLC

That looks funny…



Warning signs:

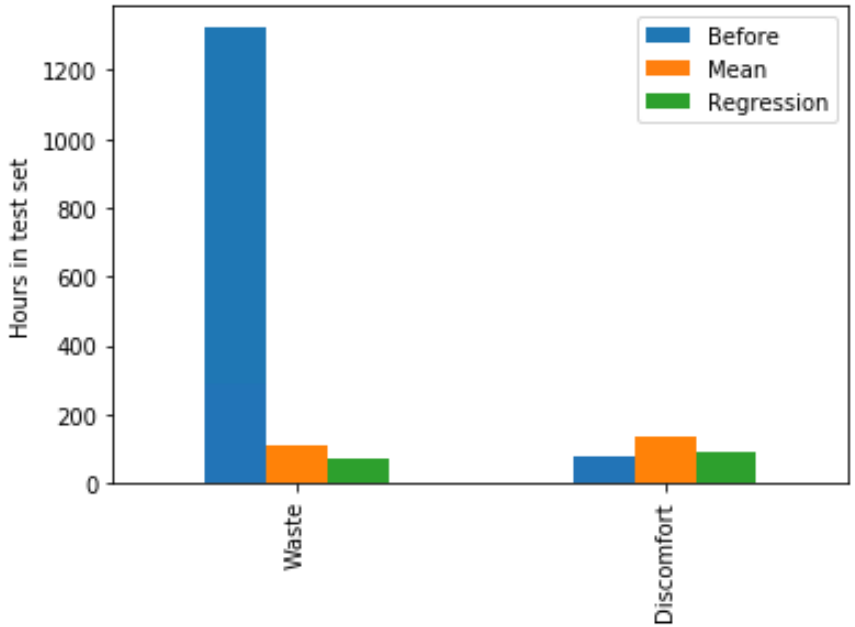
- Unexpected straight lines
- Missing data
- System digital states

©2018 OSIsoft, LLC

Potential energy savings discovered

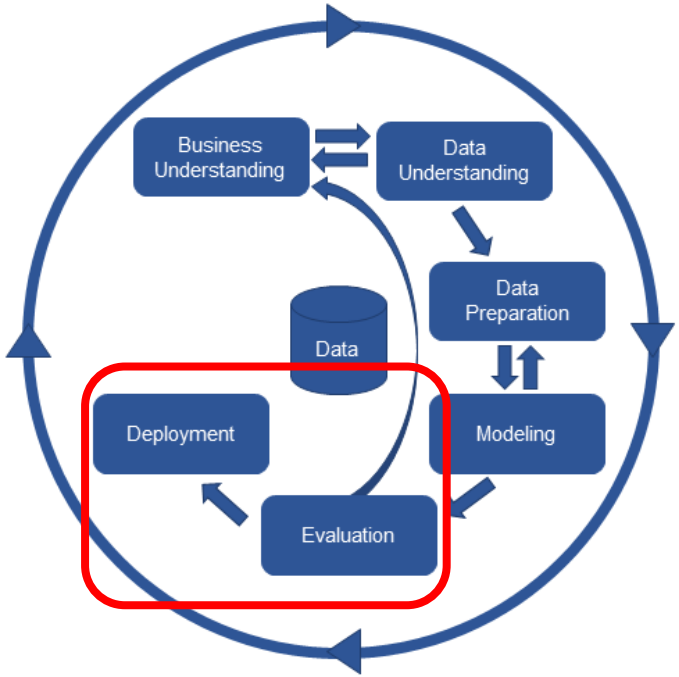
- Identified important factors for predicting cooling time
- Linear regression fits the data

$$t_{cool} = b + m_1 x_1 + \cdots + m_k x_k$$



©2018 OSIsoft, LLC

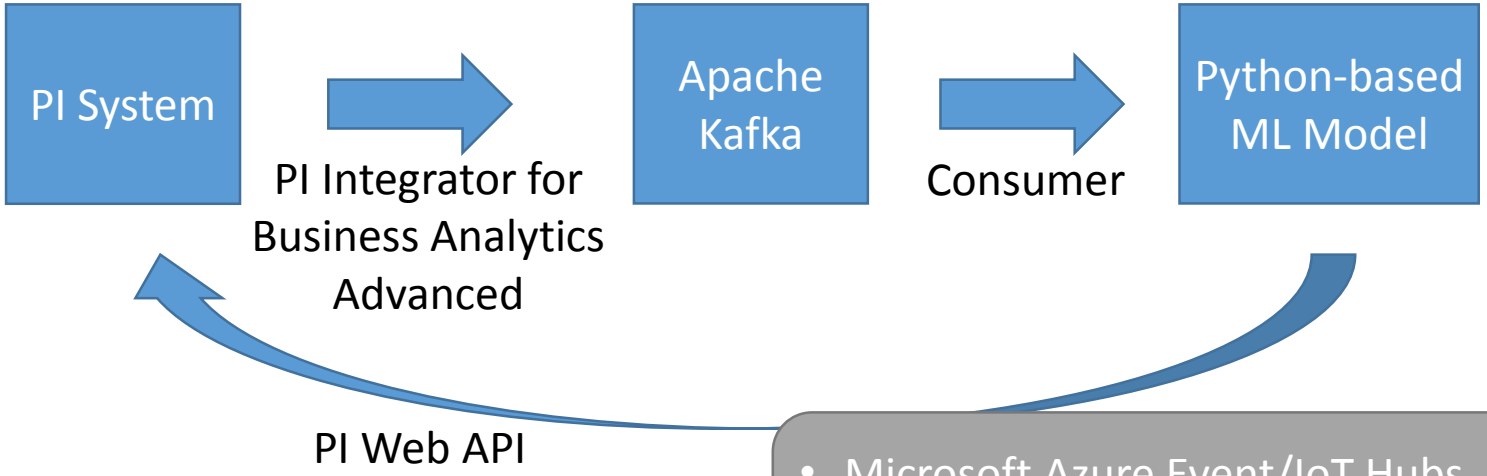
Putting the model to work



- How can I operationalize a model after it has been developed?
- What options are available for recording model predictions?

©2018 OSIsoft, LLC

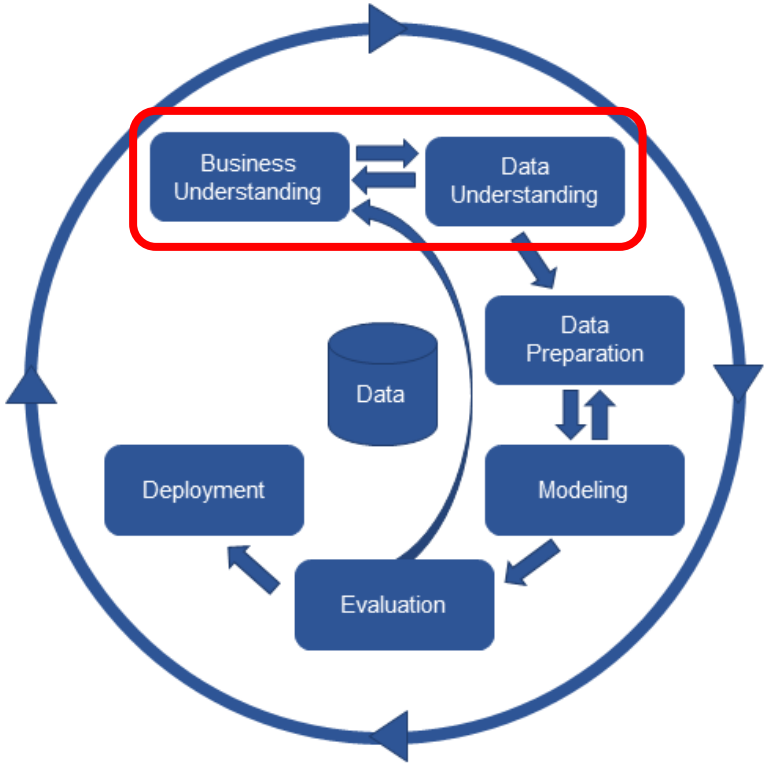
Data flow implemented in the lab



- Microsoft Azure Event/IoT Hubs
- SAP HANA Smart Data Streaming
- Asset Analytics - MATLAB Integration

©2018 OSIsoft, LLC

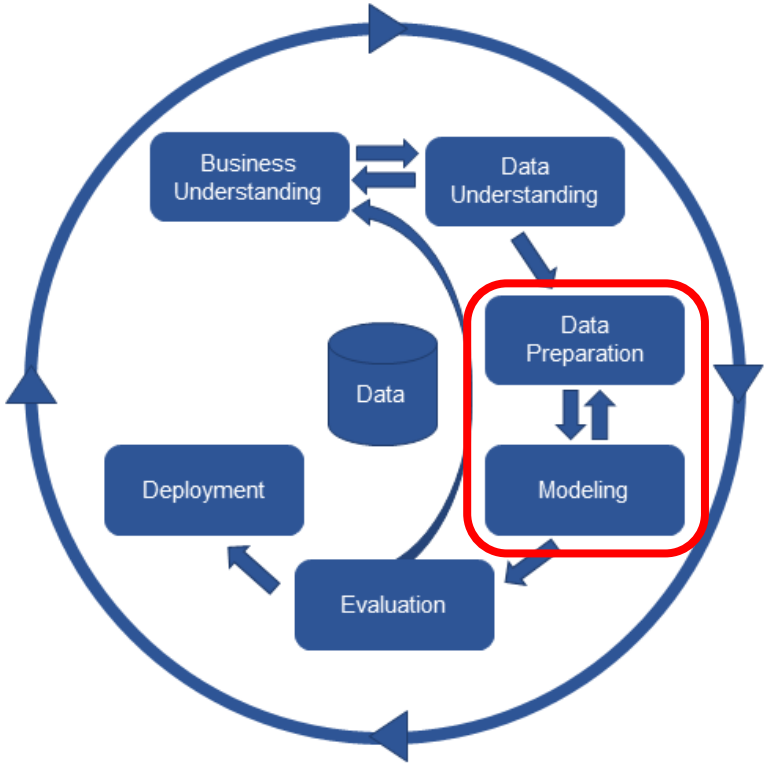
Different tools for different stages



- Asset Framework
- PI Vision/PI ProcessBook
- PI DataLink (MS Excel)
- Python/R libraries

©2018 OSIsoft, LLC

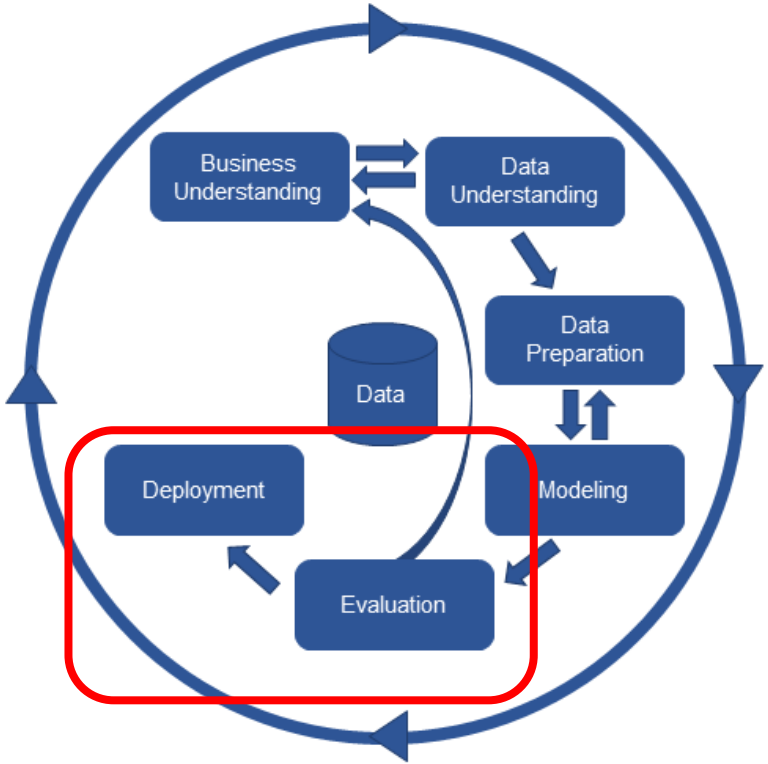
Different tools for different stages



- PI Integrator for Business Analytics
- PI OLEDB Enterprise

©2018 OSIsoft, LLC

Different tools for different stages



- PI Integrators
- Asset Analytics with MATLAB Integration

©2018 OSIsoft, LLC

Keys to success

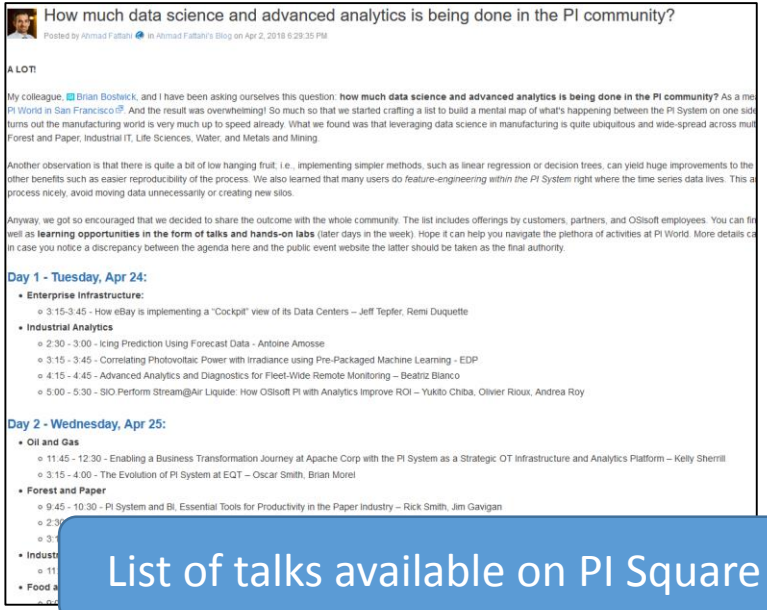
- Communication is king
- Process data has unique challenges
- PI System has tools to enable data science

- **Your knowledge of data science is a major differentiator. Leverage it!**

©2018 OSIsoft, LLC

Keep on learning!

- Labs and online courses
- PI World presentations
- Talk to other users, partners, and us



How much data science and advanced analytics is being done in the PI community?

Posted by Ahmad Fattahi in Ahmad Fattahi's Blog on Apr 2, 2018 6:29:35 PM

A LOT!

My colleague, Brian Bostwick, and I have been asking ourselves this question: **how much data science and advanced analytics is being done in the PI community?** As a me
PI World in San Francisco. And the result was overwhelming! So much so that we started crafting a list to build a mental map of what's happening between the PI System on one side
turns out the manufacturing world is very much up to speed already. What we found was that leveraging data science in manufacturing is quite ubiquitous and wide-spread across mult
Forest and Paper, Industrial IT, Life Sciences, Water, and Metals and Mining.

Another observation is that there is quite a bit of low hanging fruit, i.e., implementing simpler methods, such as linear regression or decision trees, can yield huge improvements to the
other benefits such as easier reproducibility of the process. We also learned that many users do *feature-engineering within the PI System* right where the time series data lives. This a
process nicely, avoid moving data unnecessarily or creating new silos.

Anyway, we got so encouraged that we decided to share the outcome with the whole community. The list includes offerings by customers, partners, and OSIsoft employees. You can fir
well as **learning opportunities in the form of talks and hands-on labs** (later days in the week). Hope it can help you navigate the plethora of activities at PI World. More details ca
in case you notice a discrepancy between the agenda here and the public event website the latter should be taken as the final authority.

Day 1 - Tuesday, Apr 24:

- **Enterprise Infrastructure:**
  - 3:15-3:45 - How eBay is implementing a "Cockpit" view of its Data Centers – Jeff Tepfer, Remi Duquette
- **Industrial Analytics**
  - 2:30 - 3:00 - Icing Prediction Using Forecast Data - Antoine Amosse
  - 3:15 - 3:45 - Correlating Photovoltaic Power with Irradiance using Pre-Packaged Machine Learning - EDP
  - 4:15 - 4:45 - Advanced Analytics and Diagnostics for Fleet-Wide Remote Monitoring – Beatriz Blanco
  - 5:00 - 5:30 - SIO Perform Stream@Air Liquide: How OSIsoft PI with Analytics Improve ROI – Yukito Chiba, Olivier Rioux, Andrea Roy

Day 2 - Wednesday, Apr 25:

- **Oil and Gas**
  - 11:45 - 12:30 - Enabling a Business Transformation Journey at Apache Corp with the PI System as a Strategic OT Infrastructure and Analytics Platform – Kelly Sherrill
  - 3:15 - 4:00 - The Evolution of PI System at EQT – Oscar Smith, Brian Morel
- **Forest and Paper**
  - 9:45 - 10:30 - PI System and BI, Essential Tools for Productivity in the Paper Industry – Rick Smith, Jim Gavigan

List of talks available on PI Square

bit.ly/DSPIWorld18

©2018 OSIsoft, LLC

Ahmad Fattahi
afattahi@osisoft.com
Manager, Data Science Enablement
OSIsoft

Dallas Swift
tswift@osisoft.com
Data Scientist
OSIsoft

 ©2018 OSIsoft, LLC

Questions

Please wait for the **microphone** before asking your questions

State your **name & company**

Please remember to…

Complete the Online Survey for this session

©2018 OSIsoft, LLC

Merci

谢谢

Danke

Спасибо

Gracias

Thank You

감사합니다

ありがとう

Grazie

Obrigado

©2018 OSIsoft, LLC