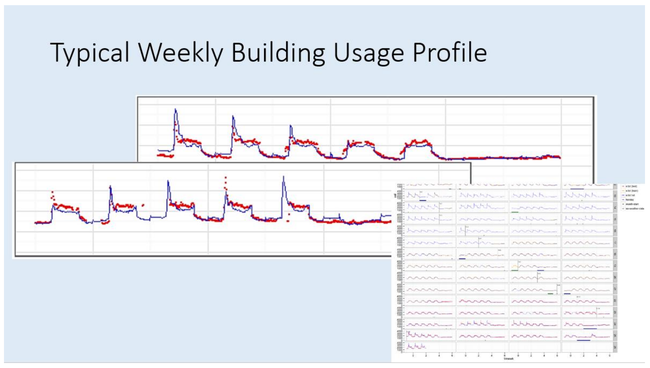
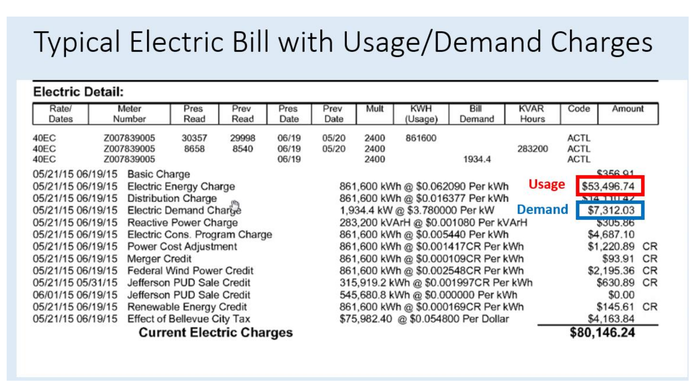
Machine Learning, Analytics & Data Science Conference

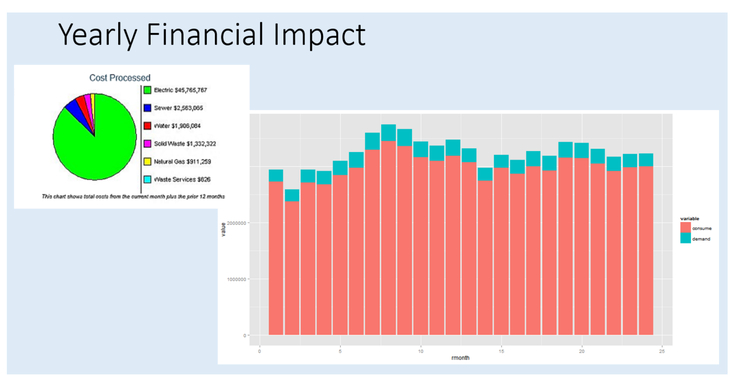
Dec 7-8 Redmond 2015

Predicting Energy Usage for 88 Acres (Microsoft Puget Sound Campus)

Business Problem: Predicting and reducing peak energy use in campus buildings



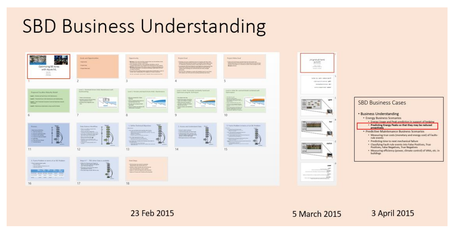


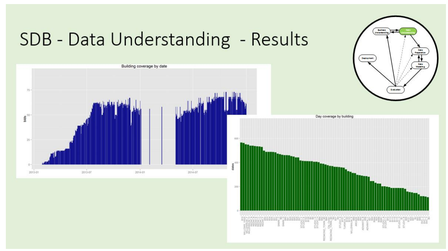


Development: Iterative building energy usage modelling

Business Understanding

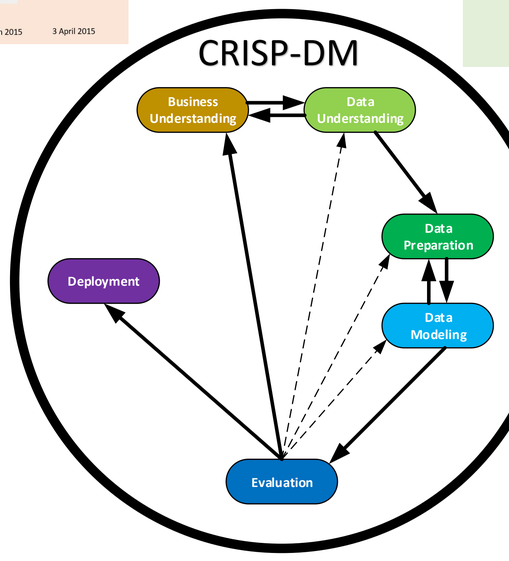
- Went through multiple phases
- Decided to concentrate on Energy first
- Concrete problem is reducing peak energy consumption
- See above for more details





Data Understanding

- 120+ Buildings, all with different HVAC systems, occupancy, age, technology, etc.
- Around 500k streams, 5000 per building, 1.5 TB of data over two years
- Data covered from 1 Jan 2013 to 31 Dec 2014
- Much missing data, uneven coverage over the period (see gaps and uneven blue barchart)
- Some data was clearly implausible
- Best building had an 80% coverage, B19
- City Center i(50% coverage) is the largest building

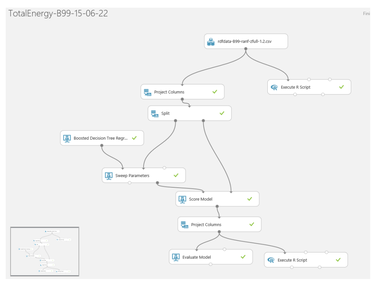


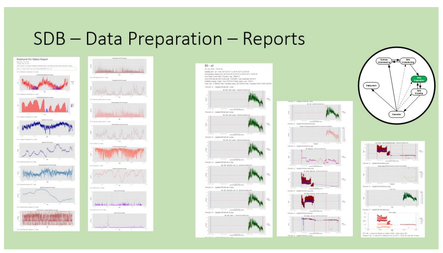
Data Preparation

- Weather data gathered and consolidated
  - Weather data in 3-5 minute intervals
  - Averaged and gathered into 15 minute chunks
  - Eliminated NAs Gaps covered by
    - Averaging weather stations
    - Interpolating few remaining gaps
- Building data gathered and consolidated
  - Energy data in variable
  - Averaged and gathered into 15 minute chunks
  - Created heating degrees and cooling degree variables
- Added time data
  - Marker for holidays
  - Conversion of time into decimal time
- End result is data for AML
- Data Understanding increased enormously, many surprises
- Spline fitting turned out not to work so well, abandoned
- Hard to find weather stations with complete coverage
- Looking for at least 6 years of coverag

Deployment

- Models deployed in Azure ML
- Used built-in modeling functions
- Pilot Deployment in RE&F Operations Dashboard
- Staging and Continuous Training currently in development
- See architecture diagram below for more details





CRISP-DM

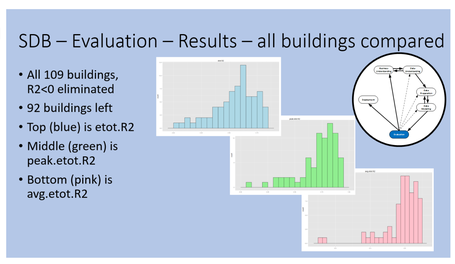
- Cross Industry Standard Process for Data Mining
- 5 Phases (See above and examples here)
- Highly iterative
- Best know methodology for Data Mining/Data Science

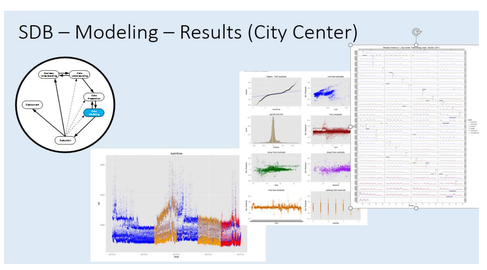
SDB = Software Defined Building

- An umbrella concept for smart building technology
- Includes Energy Management
- IoB (Internet of Buildings is also sometimes used)

Evaluation

- All 109 buildings, R2<0 eliminated
- 92 buildings left
- Top (blue) is etot.R2
- Middle (green) is peak.etot.R2
- Bottom (pink) is avg.etot.R2

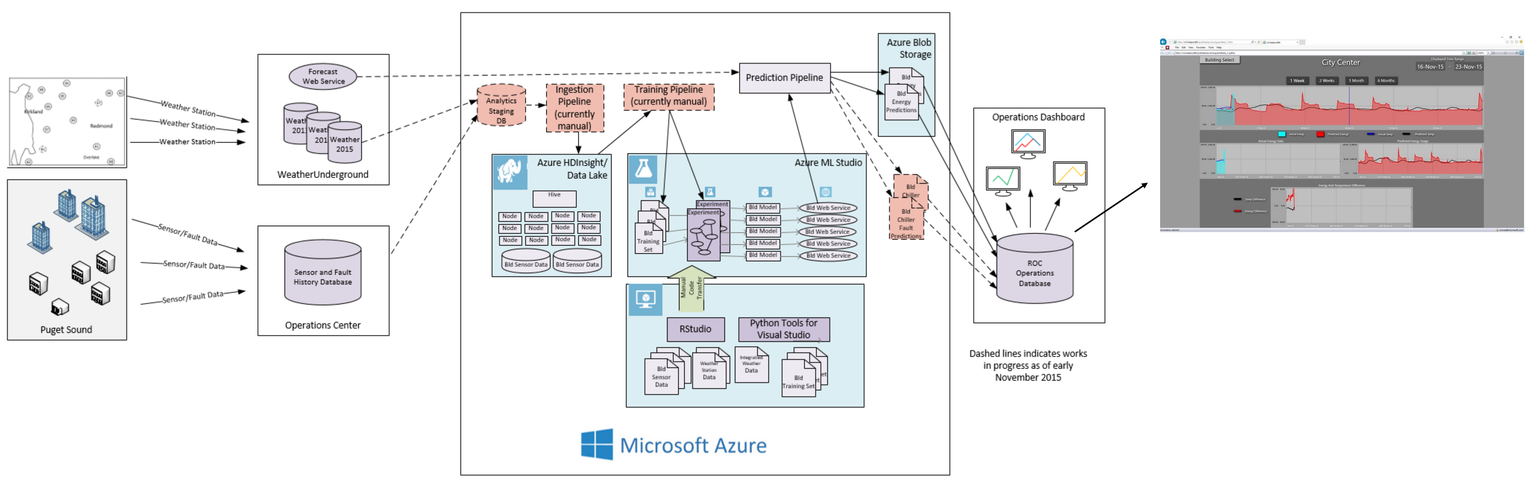




Modelling

- Considered different architectures
- Went with Standard R and AML in the end
- Many choices available today:
  - Classic regression and optimization
  - Ensemble methods (Boosted Trees, Random Forests, etc.)
  - Other ML methods (NN, DNN, SVM, etc.)
  - Time Series forecasting (ARIMA)
- Ensemble methods worked best

Deployment: Solution integration into 88 Acres Operations Dashboard



Next Steps

The following projects are now running

- Evaluation of Pilot results and model refinement
- Integration with 88 Acres data ingestion and historian
- Implementation of continuously updating of models
- Chiller Fault Prediction from alarm data
- Advanced Energy Model for Command-and-Control of energy consumption

Microsoft

Poster Session: # 157

Girish Nathan – Sr. Data Scientist - IMML ADS - girish.nathan@microsoft.com

Mike Wise – MCS Architect – Applied Incubation – mike.wise@microsoft.com