CITY DIGITAL –

BUILDING DATA SERVICES

Professor: Dr. Shlomo Argamon

Students: Feixiang Xi, Sachet Misra, Yufei Yu

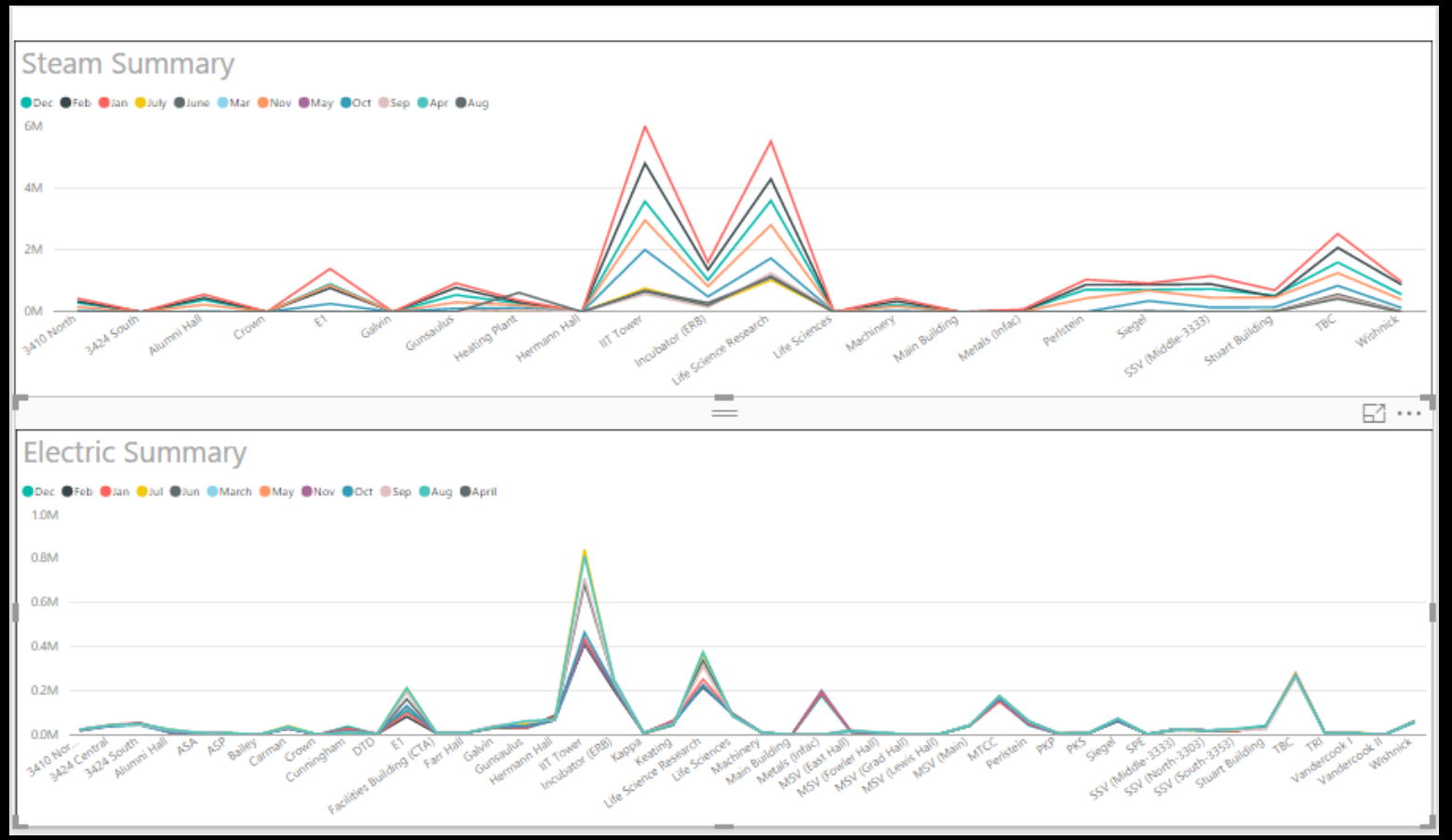
OBJECTIVES

- What can we do with the data?
- What data do we need?
- How can we use Machine Learning Algorithms?
- To find the right database structure for each of the dataset
- To find the errors in data

DATASETS

- Energy Dataset:
  - 8940 files extracted from an SQL database using an access database as an interface.
  - Update rate is every 15 minutes and involves 5 years of data
  - Three categories – steam, water, electricity
  - 175,200 rows

ENERGY DATA - PLOTTING



DATASETS

- Work Order Data:
  - This is a file extracted from an SQL database
  - involves all the work orders for the past one year
  - Has about 10 columns and 28,510 rows

DATASETS

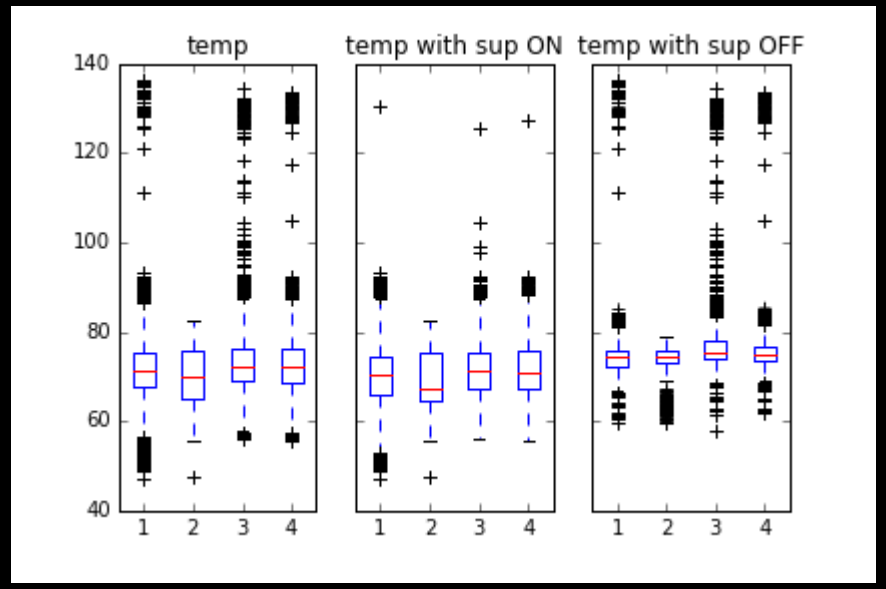
- Siemens Data –
  - Involves 5111 text files, stored from reports of Siemens
  - Targets machines like AHU(Air-Handling Unit), RTU (Roof Top Unit), CHW (Chilled water Units), 20th floor KWH(Kilo-Watt Hour), Exhaust Fans
  - Update rate – once per day
  - Each file has about 15 columns and 250,000 rows per dataset
  - <a total of about 250,000 x 22 = 5,500,000 rows>

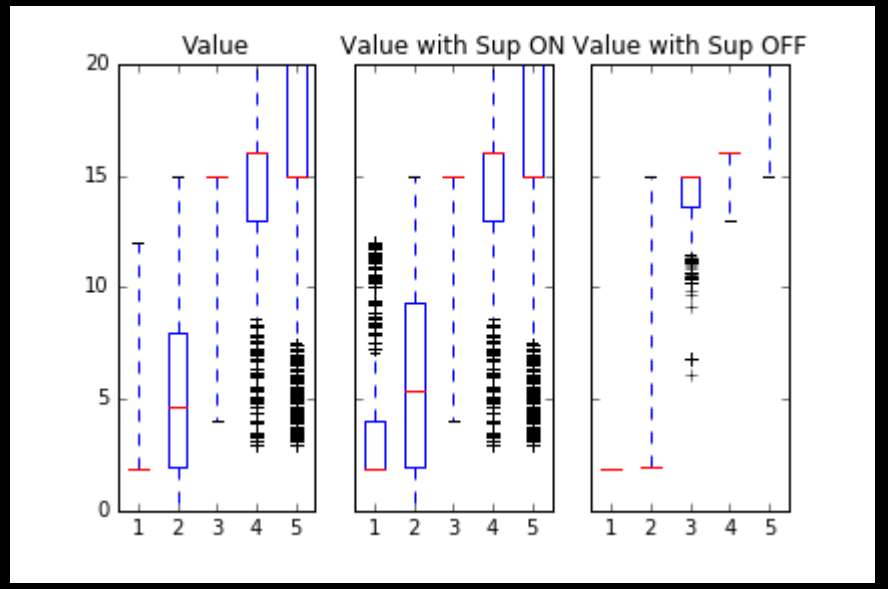
CLEANING

- When converting files,
  - Had about 1000 random entries per column out of 250,000 (rest were empty)
  - Used techniques like forward fill and backward fill
- Divided columns into three categories
  - Values, temperature, (not value or temp)
  - Box plots to identify outliers
  - Found relationships between values and temperature

ERROR CHECKING

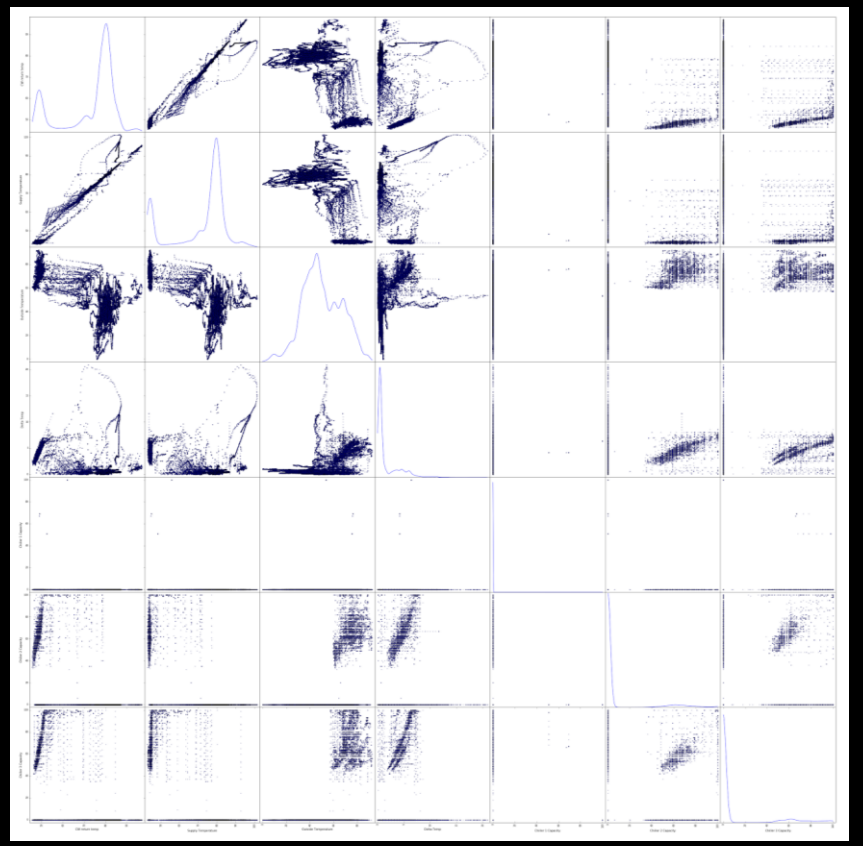
- Used box plots to check for outliers in every variable in the dataset
- AHU02 data – after column categorisation



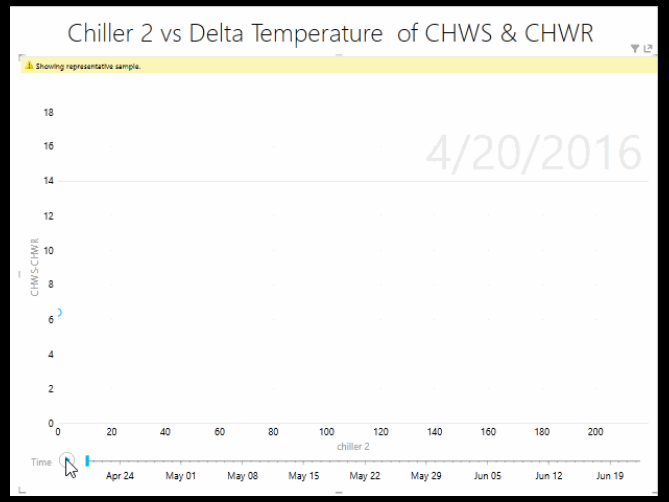


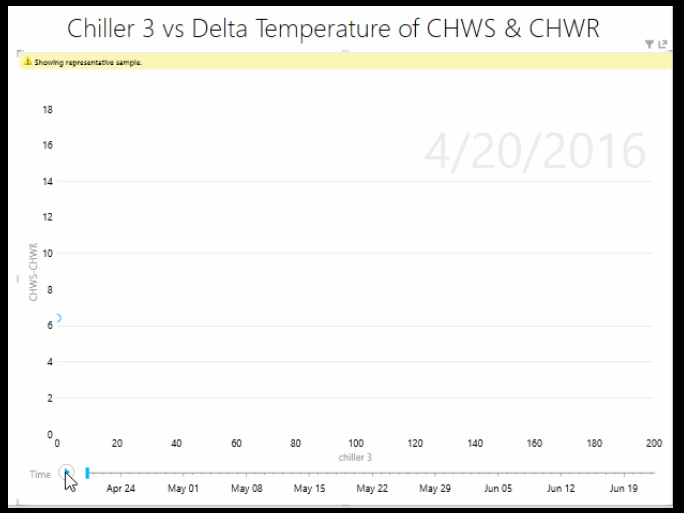
RELATIONSHIPS

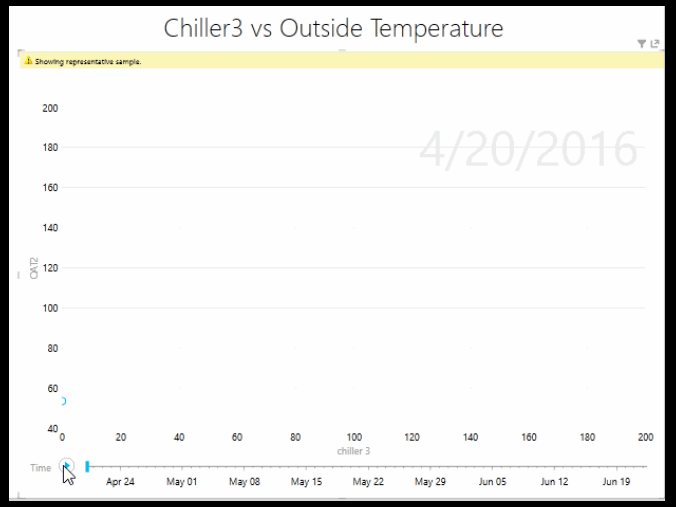
- Dependencies found in every dataset using bivariate analysis
- Shows CWH data

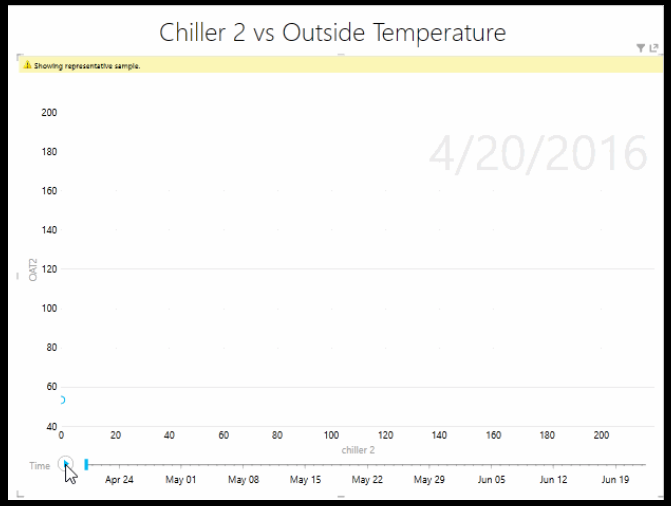


CHW DATASET VISUALISATION



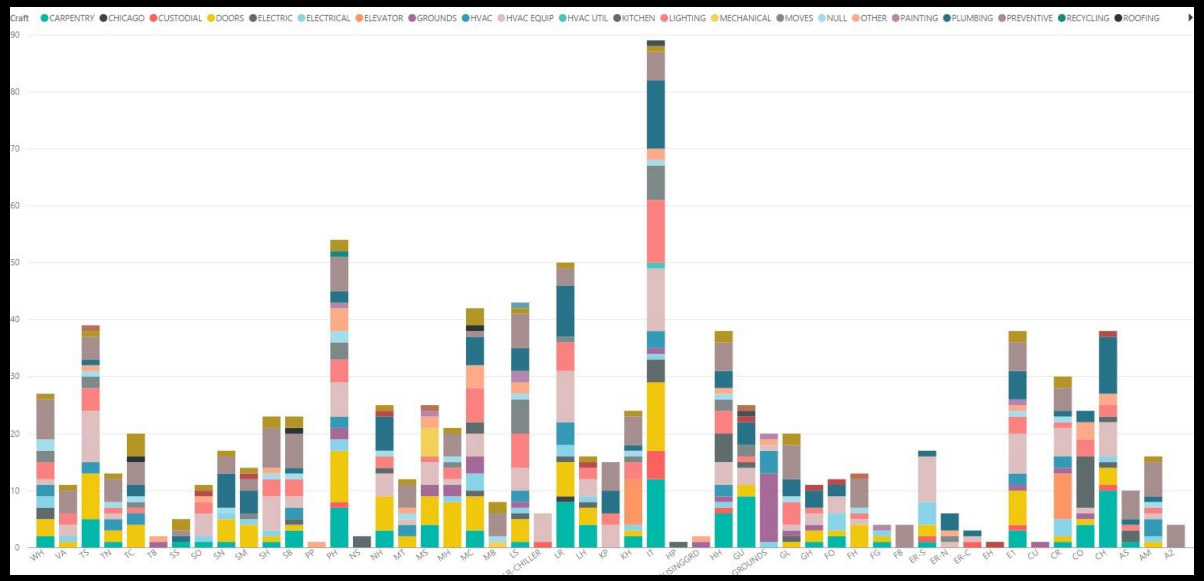






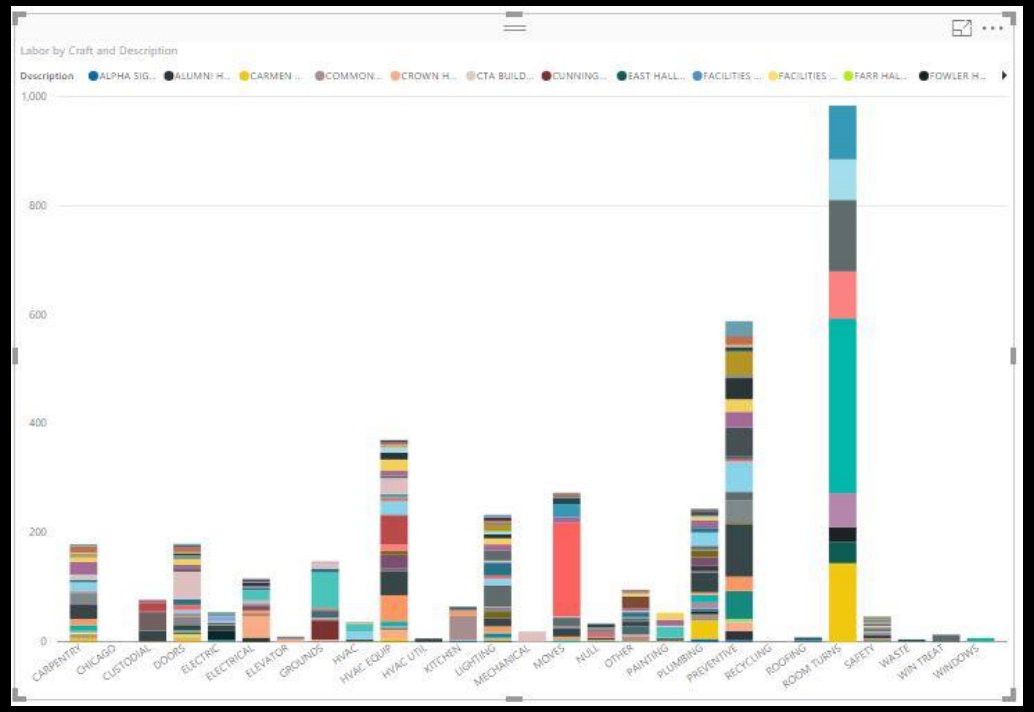
PRELIMINARY ANALYSIS

- To choose the right dataset we work on some preliminary analysis using PowerBI
- Work order data – Most variety and Most data – categorised in terms of buildings



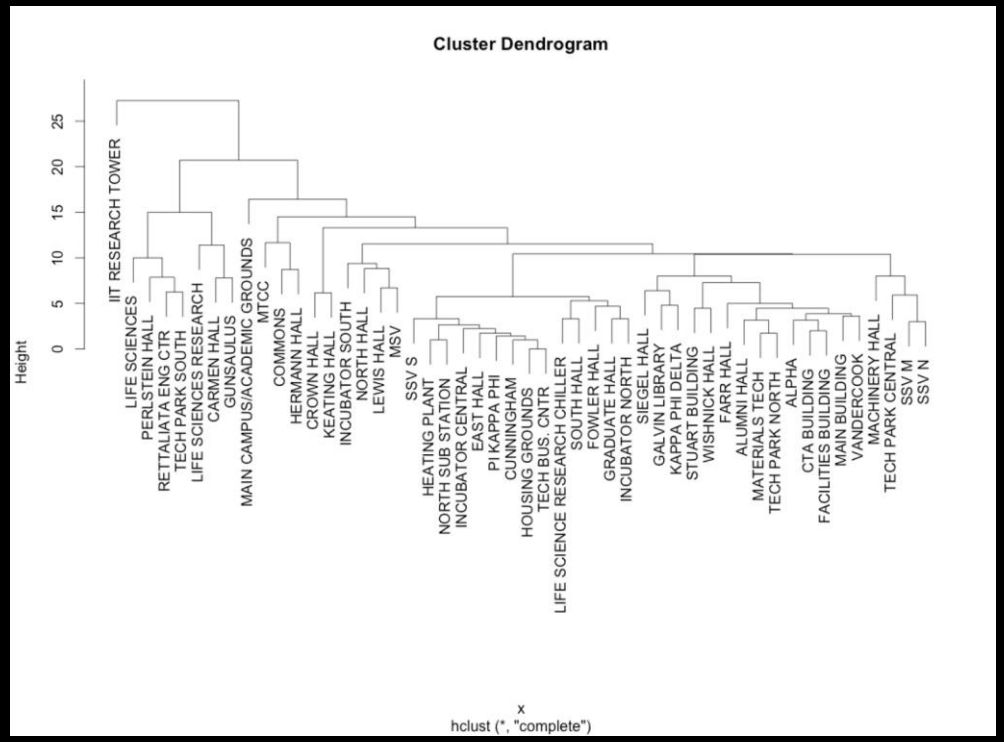
PRELIMINARY ANALYSIS

- Work order data – categorised based on different kinds of work



PRELIMINARY ANALYSIS

- Energy dataset – dendograms



DATA INTEGRATION

- Objective:
  - To be able to predict a work-order
  - Hence need to find a some trend with the various variables in the dataset
- Textual analysis to find the most ordered work order
- Here we have considered information from the18th floor, AHU-18

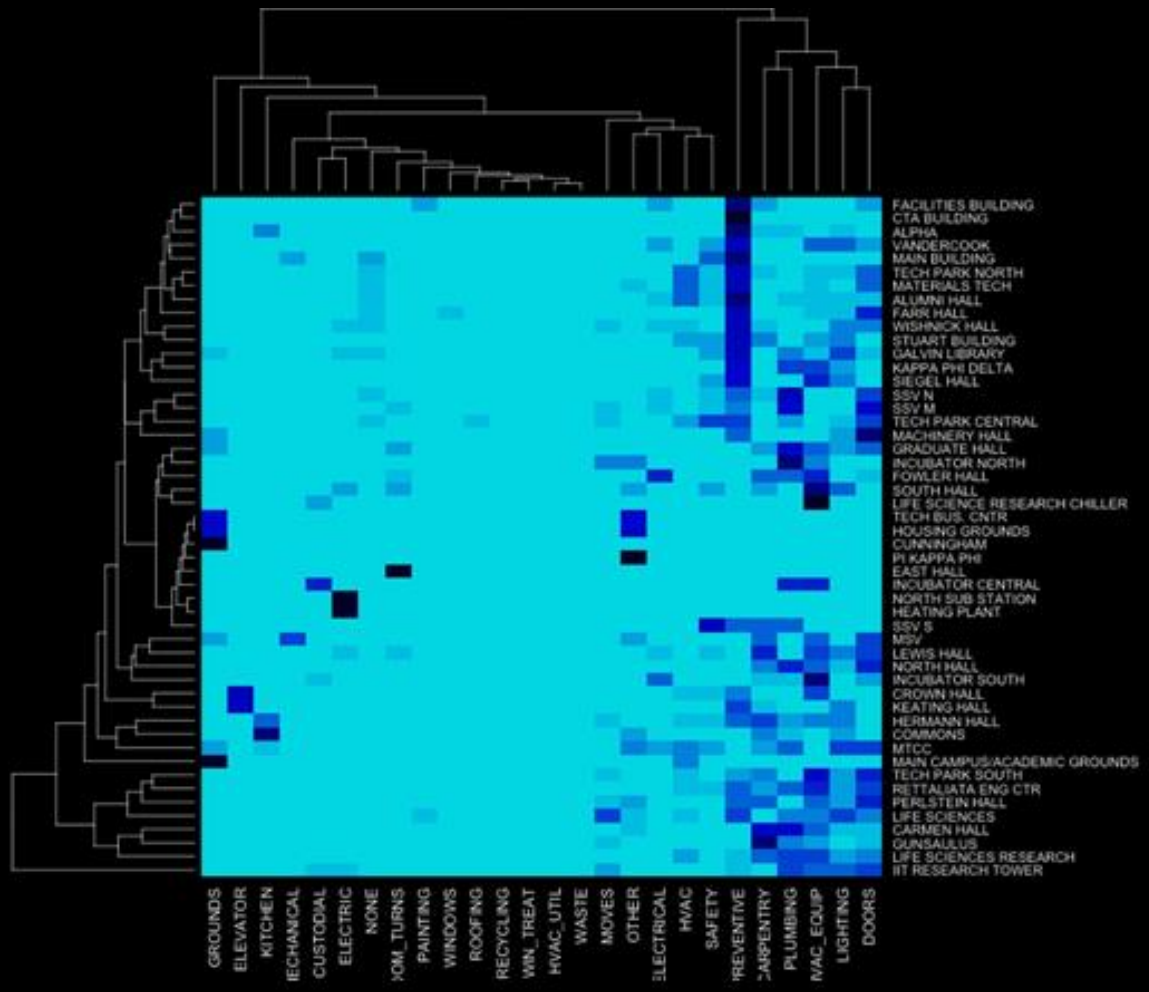
WORK ORDER DATA

- TEXT analysis to do a frequency analysis on work order description column
- Some words were provided by Accenture to help us understand the important ones
- Found 40 such keywords – (left column – from analysis; right column – from companies)

ROOM	548	FILTER	32
COLD	129	FAN	32
REPAIR	369	COIL	16
BASEMENT	64	TOO HOT	85
SUPPLY	30	TOO COLD	125

WORK ORDER DATA

- Heat map



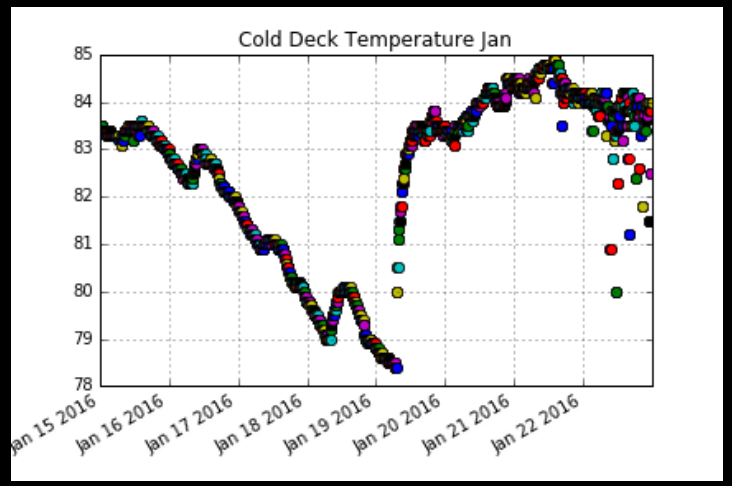
TEMPORAL ANALYSIS

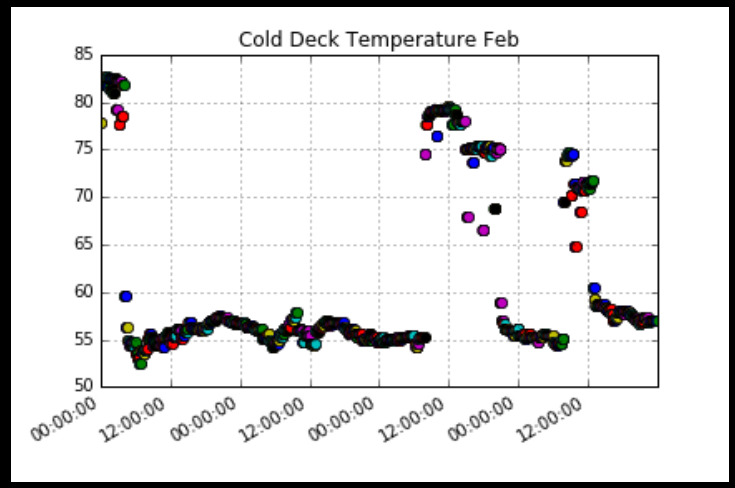
- Took the phrase "TOO COLD" from the dataset to do some temporal analysis

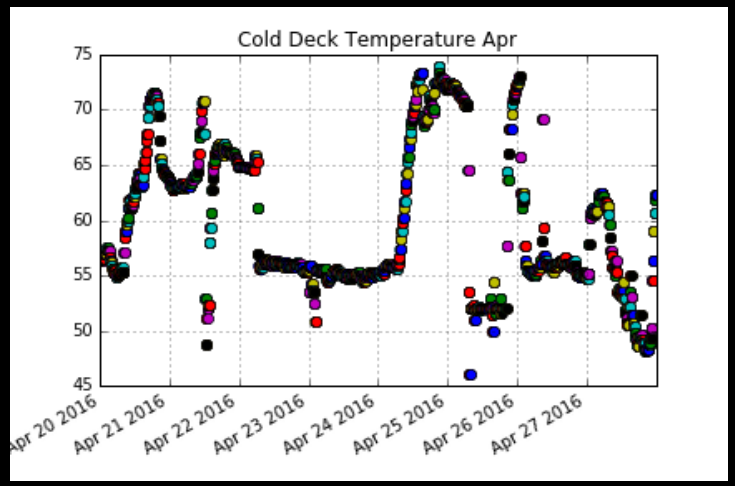
6647	WO128240	ROOM IS TOO COLD	TJABCZYN	IT	18.0	18E4-1	HVAC	2015-05-18 10:24:11	2015-05-29 14:55:14	0.50	2015-05-29	14:55:14
9372	WO130987	ROOM IS TOO COLD	TJABCZYN	IT	18.0	18E4-1	HVAC	2015-07-13 13:18:27	2015-07-14 14:45:19	1.50	2015-07-14	14:45:19
14011	WO135624	ROOM IS TOO COLD	TJABCZYN	IT	18.0	18E4-1	HVAC EQUIP	2015-10-05 09:50:52	2015-10-08 15:03:05	1.00	2015-10-08	15:03:05
19516	**WO141158**	**ROOM IS TOO COLD**	**TARIANOU**	**IT**	**18.0**	**18D3-1**	**HVAC EQUIP**	**2016-01-21 15:08:25**	**2016-01-22 16:06:45**	**1.00**	**2016-01-22**	**16:06:45**
20548	**WO142182**	**ROOM IS TOO COLD**	**TJABCZYN**	**IT**	**18.0**	**18D3-2**	**HVAC EQUIP**	**2016-02-10 10:48:55**	**2016-02-16 14:57:32**	**1.00**	**2016-02-16**	**14:57:32**
24388	**WO146029**	**ROOM IS TOO COLD**	**TJABCZYN**	**IT**	**18.0**	**n/a**	**HVAC EQUIP**	**2016-04-27 09:12:11**	**2016-04-27 15:09:15**	**0.50**	**2016-04-27**	**15:09:15**

TEMPORAL ANALYSIS

- Cold deck temperature

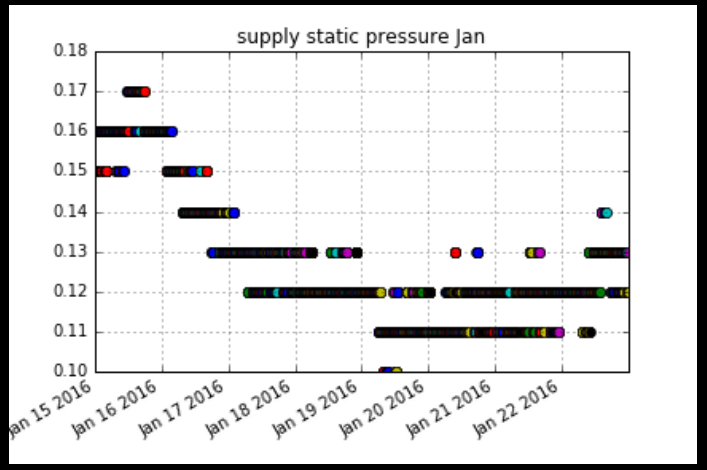


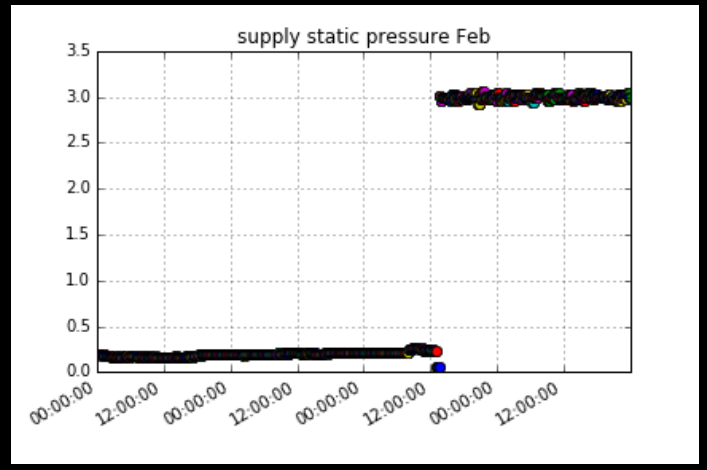


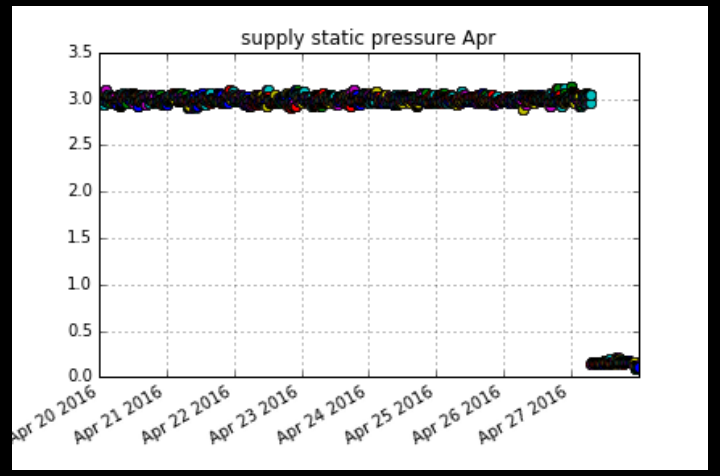


TEMPORAL ANALYSIS

- Supply static pressure

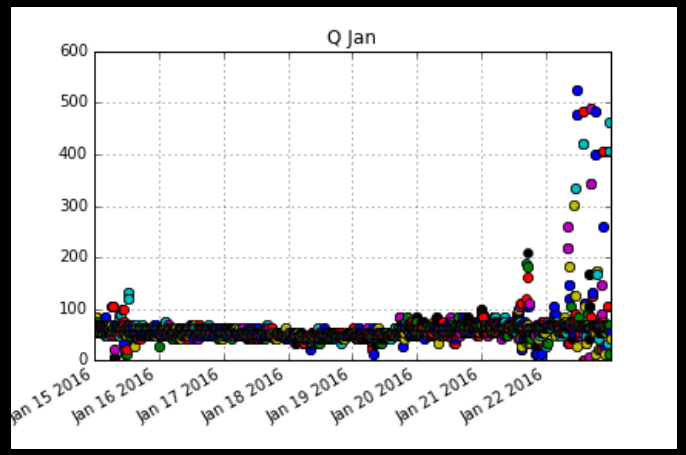


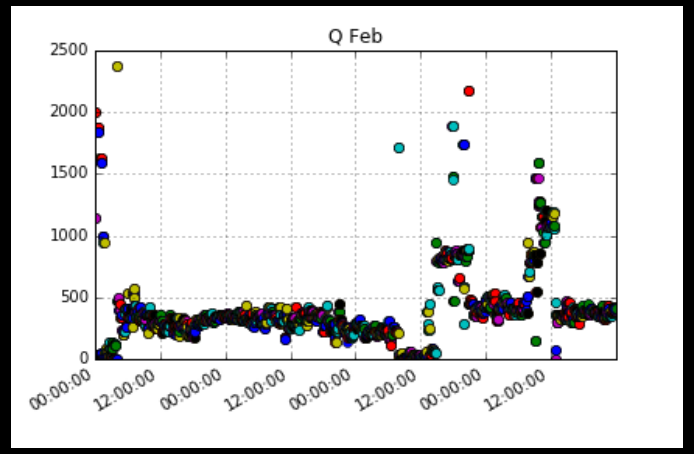


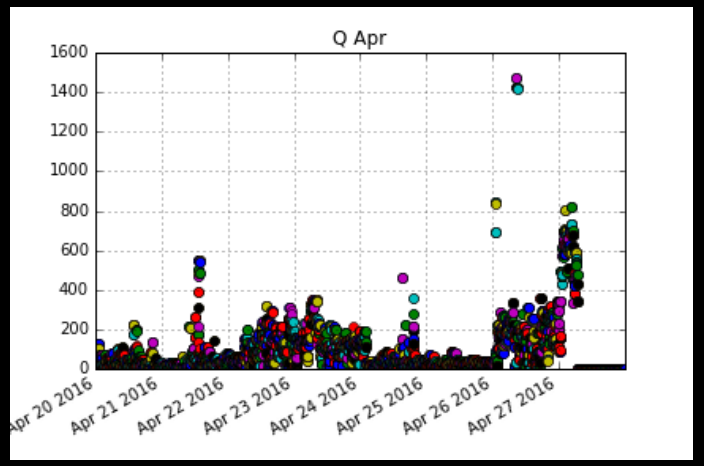


TEMPORAL ANALYSIS

- Q – efficiency of AHU
- Probable spike in efficiency before the call - Conclusion







POST MORTEM

- Database suggestions (AHU)

R	1
SAF/S	14
CCO/CCV	18
CDAT	1
CDT	6
DAT	10
DSP_S17	1
DX1_S17	1
DX2_S17	1
EFCAWA/E	12
EFGEN	2
EFNEC	1
FBD_S17	1
FBO	1
HCO	1
HDT	6
HDV	1
MAM_S17	1
MAT	16
MND	4
PHT	13
PHV/PH1V	15
PH2V	1
PVM	1
RAT	12
RH1T/R1T_S10	8
RH1V_S3/RHV_S6/R1V_S10	14
RH2T/R2T_S10	8
S15RAT	1
HVO/HV	2
p	1
KH_S17	1
KL_S17	1
MAD	16
RH2V_S3/R2V_S10	8
RVD	1
SAF	3
SDT	1
SFVFD:RUN.STOP_S15	1
SFVFD:SPEED/SVD/SFSPD	9
SSP_18	1
TMD	1
E13RVD_S14	1
S15E15	1

POST MORTEM

- Database suggestions
  - Work Order:
    - Work order number
    - Description
    - Building
    - Floor
    - Room
    - Craft
    - Enter Date
    - Close Date
    - Hours SUM
    - Requestor

WHAT MORE CAN WE DO?

- Performance comparison between various (similar) devices (AHU1 and AHU2)
  - Need of common set of column names according to a naming scheme required
- Pattern recognition of work to predict which might need repair
  - Need of data to be in more detail –
    - when the work-order was finished
    - What was repaired
  - AHU,CHW, etc. data needs to be in sync with the work order data
- An overall model to regulate temperature and other factors more efficiently

INTERESTING FACTS OF THE PROJECT

- 6 weeks spent in data collection and cleaning
  - Includes preliminary analysis
- 130 graphs drawn
- 6 million unique records across various datasets
- 3 database sources
- 3 students
- 5 guides for the project
- 4 programming languages used

SOFTWARES USED

- R (ggplot2, RODBC,)
- PowerBI - for graphs
- Python (matplotlib, pandas, numpy, collections, nltk, string, itertools)
- Microsoft (Excel (VB-Macros, graphs),Microsoft SQL Server Management System, Azure, Access DB)
- Google (Drive, Documents)
- Basecamp2
- SQL

“

DATA IS NOT INFORMATION, INFORMATION IS NOT KNOWLEDGE, KNOWLEDGE IS NOT UNDERSTANDING, UNDERSTANDING IS NOT WISDOM.

- CLIFFORD STOLL

”

Thank You