Midterm Report:
UBCO MDS Capstone - Urban Data Lab (UDL)

WEEK 5-June 2, 2020

Connor Lee, Claudia Nikel, Eva Nguyen, Alex Tamm

[TEMP]

Midway through the course, teams will be responsible for a formal presentation highlighting their research questions, progress, techniques and analysis plan (how the team is approaching the problem) in addition to roadblocks and difficulties encountered.

You will be highlighting the:

- research questions
- Progress
- techniques and analysis plan (how the team is approaching the problem)
- roadblocks and difficulties encountered.
- present and discuss interim results and/or deliverables (if you have something even if it is not complete, you can speak to it here).

The audience should be able to gain an understanding of what the problem is you are working on, what your approach is as well as understanding and appreciating the progress made with the presentation of interim work (i.e. show what you've produced so far/preliminary results, etc).

Outline

- Project Background
- Overview of the Project
- Problem Approach
  - What the data looks like
  - Our approach
- Our Progress
  - Project schedule
  - What we have completed
  - What we are currently working on
  - What needs to be completed
- Difficulties & Roadblocks
- Questions

Project Background

Urban Data Lab (UDL) Capstone Client Overview

Contacts are:

- Jiachen Wei (MDS alum)
- Mike Kennedy, Ph.D

Client Overview

- Formed in September 2019 to advance data analytics capabilities on UBC Vancouver campus to address campus-wide sustainability challenges

Data Overview

- UDL mirrored and stored live-streaming building energy datasets in InfluxDB and made it available to students and researchers
- The SKYSPARK database provides data recorded by the meters and smart devices of many UBC buildings

Overview of Project

Overview

Research Question

- Based on a building's sensor data, how can the data be grouped automatically into NRCan secondary end-use classifications?

Proposed Solution and Deliverables

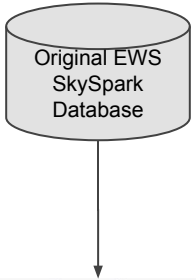
- A Python program that:
  - Queries and cleans the data required for classifying instrumentation by end-use for the Pharmacy building
  - Classifies instruments by end-use (NRCAN guidelines)
  - Time permitting, scales the model so that it can be used on buildings other than the Pharmacy building
- Web-based Grafana dashboard
- Final report & presentation to UDL

Problem Approach

Problem Definition

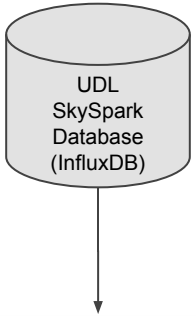
Classify all Pharmacy Building sensors that record energy consumption into appropriate end-uses.

What does the data look like?



id	ahu	ahuMode	air	alarm	avg	bacnetConnRef	bacnetCur	bacnetHis	bacnetObjectId	bac
Pharmacy Heating Plant HX-2 P-HX2A HX2_PHX2A_VFD_PWR(kW...						PHARMA PHARM_HX2_FCU_B05 (701100)		TL56		✓
Pharmacy Elec Submeters LEED-2N1PC3 2N1PC3_CurrentC						PHARMA PHARM_LEED_METER_MOD2 (702100)		TL59		✓
Pharmacy Rm Corr FC-513 FCU_513_S			✓			PHARMA PHARM_FCU_513 (700742)		TL1		✓
Pharmacy Rm 1420 FC-111 FCU_111_S			✓			PHARMA PHARM_FCU_111/FF_102/EF_106 (701206)		TL1		✓
Pharmacy Elec Submeters LEED-6ETLE1 6ETLE1_CurrentB						PHARMA PHARM_LEED_METER_MOD1 (702000)		TL208		✓
Pharmacy Rm B503 EAV-BS048 BS048_AVG_SPACE_TEMP_AV			✓			PHARMA PHARM_MACRO_SERVER_FH_TEMP_RM (702900)		TL182		✓
Pharmacy AHU-15 AHU15_FIRE_MODE						PHARMA PHARM_AHU14_15,EF3,SB1~2 (700900)		TL122		✓
Pharmacy Rm 6107 FC-601 FCU_601_SCHED						PHARMA PHARM_FCU_601 (700514)		TL9		✓
Pharmacy Unsorted Points CO2_RM_B212						PHARMA PHARM_MISC_BSMNT_AL (700758)		TL3		✓
Pharmacy CRAH System CRAH-2 CRAH2_LO_RT_AL_BV				✓		PHARMA PHARM_HTREJ,HWS,HX3,CRAH (700200)		TL178		✓
Pharmacy Rm 4616 RAD-4-09 RZ4_09_HWRT						PHARMA PHARM_RZ4_04&09 (701340)		TL11		✓
Pharmacy Heating Plant BLR-3 BLR_PB3_S						PHARMA PHARM_BLR1~4,CT,HX1_7,DHW (700800)		TL47		✓
Pharmacy EAV-BE001 BE001_SASH_OPEN_PERCENT						PHARMA PHARM_MACRO_SERVER_FH_TEMP_RM (702900)		TL360		✓
Pharmacy LEF-2 EF-2C EF2_F3_VFD_INST_PWR(kW)						PHARMA PHARM_SB3~6,EF2,SMOKE_EF14~16 (701600)		TL144		✓

elec	elecMeterLoad	enable	energy	entering	enum	equipName	equipRef	esc	exhaust	fan
✓			✓				Pharmacy Heating Plant HX-2 P-HX2A	✓		
✓							Pharmacy Elec Submeters LEED-2N1PC3	✓		
							Pharmacy Rm Corr FC-513	✓		✓
							Pharmacy Rm 1420 FC-111	✓		✓
✓							Pharmacy Elec Submeters LEED-6ETLE1	✓		
							Pharmacy Rm B503 EAV-BS048	✓		
					OFF,ON		Pharmacy AHU-15	✓		
					OFF,ON		Pharmacy Rm 6107 FC-601	✓		
							Pharmacy Unsorted Points	✓		
					OFF,ON		Pharmacy CRAH System CRAH-2	✓		
				✓			Pharmacy Rm 4616 RAD-4-09	✓		
							Pharmacy Heating Plant BLR-3	✓		
							Pharmacy EAV-BE001	✓		
✓							Pharmacy LEF-2 EF-2C	✓		
							Pharmacy LEF-3	✓	✓	
							Pharmacy Unsorted Points	✓	✓	✓



	time	equipRef	groupRef	navName	siteRef	typeRef	unit	value
0	2019-12-13T22:24:37Z	AHU-02	CIRS Air Systems	Discharge Air Temp	CIRS	CIRS_AHU2_SUPPLY_AIR_T	°C	16.707474
1	2019-12-13T22:39:37Z	AHU-02	CIRS Air Systems	Discharge Air Temp	CIRS	CIRS_AHU2_SUPPLY_AIR_T	°C	16.105682
...	...	...	...	...	...	...	...	...
580	2019-12-19T23:39:59Z	AHU-02	CIRS Air Systems	Discharge Air Temp	CIRS	CIRS_AHU2_SUPPLY_AIR_T	°C	23.058758
581	2019-12-19T23:54:59Z	AHU-02	CIRS Air Systems	Discharge Air Temp	CIRS	CIRS_AHU2_SUPPLY_AIR_T	°C	23.049675

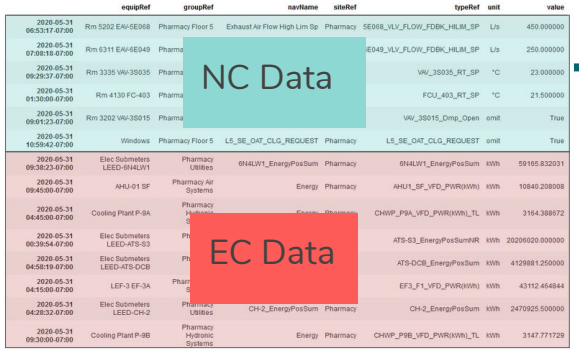
Energy Consumption (EC) Sensors vs Non-Energy Consumption (NC) Sensors

	equipRef	groupRef	navName	siteRef	typeRef	unit	value
2020-05-31 06:53:17-07:00	Rm 5202 EAV-5E068	Pharmacy Floor 5	Exhaust Air Flow High Lim Sp	Pharmacy	5E068_VLV_FLOW_FDBK_HILIM_SP	L/s	450.000000
2020-05-31 07:08:18-07:00	Rm 6311 EAV-6E049	Pharmacy Floor 6	Exhaust Air Flow High Lim Sp	Pharmacy	6E049_VLV_FLOW_FDBK_HILIM_SP	L/s	250.000000
2020-05-31 09:29:37-07:00	Rm 3335 VAV-3S035	Pharmacy Floor 3	Zone Temp Effective Sp	Pharmacy	VAV_3S035_RT_SP	°C	23.000000
2020-05-31 01:30:00-07:00	Rm 4130 FC-403	Pharmacy Floor 4	Zone Temp Effective Sp	Pharmacy	FCU_403_RT_SP	°C	21.500000
2020-05-31 09:01:23-07:00	Rm 3202 VAV-3S015	Pharmacy Floor 3	Discharge Air Damper Open Cmd	Pharmacy	VAV_3S015_Dmp_Open	omit	True
2020-05-31 10:59:42-07:00	Windows	Pharmacy Floor 5	L5_SE_OAT_CLG_REQUEST	Pharmacy	L5_SE_OAT_CLG_REQUEST	omit	True
2020-05-31 09:38:23-07:00	Elec Submeters LEED-6N4LW1	Pharmacy Utilities	6N4LW1_EnergyPosSum	Pharmacy	6N4LW1_EnergyPosSum	kWh	59165.832031
2020-05-31 09:45:00-07:00	AHU-01 SF	Pharmacy Air Systems	Energy	Pharmacy	AHU1_SF_VFD_PWR(kWh)	kWh	10840.208008
2020-05-31 04:45:00-07:00	Cooling Plant P-9A	Pharmacy Hydronic Systems	Energy	Pharmacy	CHWP_P9A_VFD_PWR(kWh)_TL	kWh	3164.388672
2020-05-31 00:39:54-07:00	Elec Submeters LEED-ATS-S3	Pharmacy Utilities	ATS-S3_EnergyPosSumNR	Pharmacy	ATS-S3_EnergyPosSumNR	kWh	20206020.000000
2020-05-31 04:58:19-07:00	Elec Submeters LEED-ATS-DCB	Pharmacy Utilities	ATS-DCB_EnergyPosSum	Pharmacy	ATS-DCB_EnergyPosSum	kWh	4129881.250000
2020-05-31 04:15:00-07:00	LEF-3 EF-3A	Pharmacy Air Systems	Energy	Pharmacy	EF3_F1_VFD_PWR(kWh)	kWh	43112.464844
2020-05-31 04:28:32-07:00	Elec Submeters LEED-CH-2	Pharmacy Utilities	CH-2_EnergyPosSum	Pharmacy	CH-2_EnergyPosSum	kWh	2470925.500000
2020-05-31 09:30:00-07:00	Cooling Plant P-9B	Pharmacy Hydronic Systems	Energy	Pharmacy	CHWP_P9B_VFD_PWR(kWh)_TL	kWh	3147.771729

NC Data

EC Data

High Level Overview

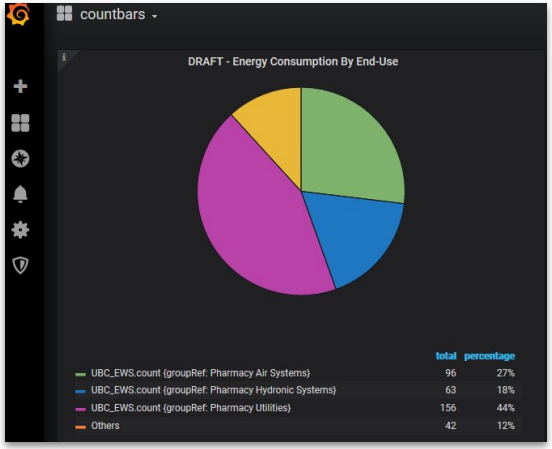


Classify EC Data into End-Use Types

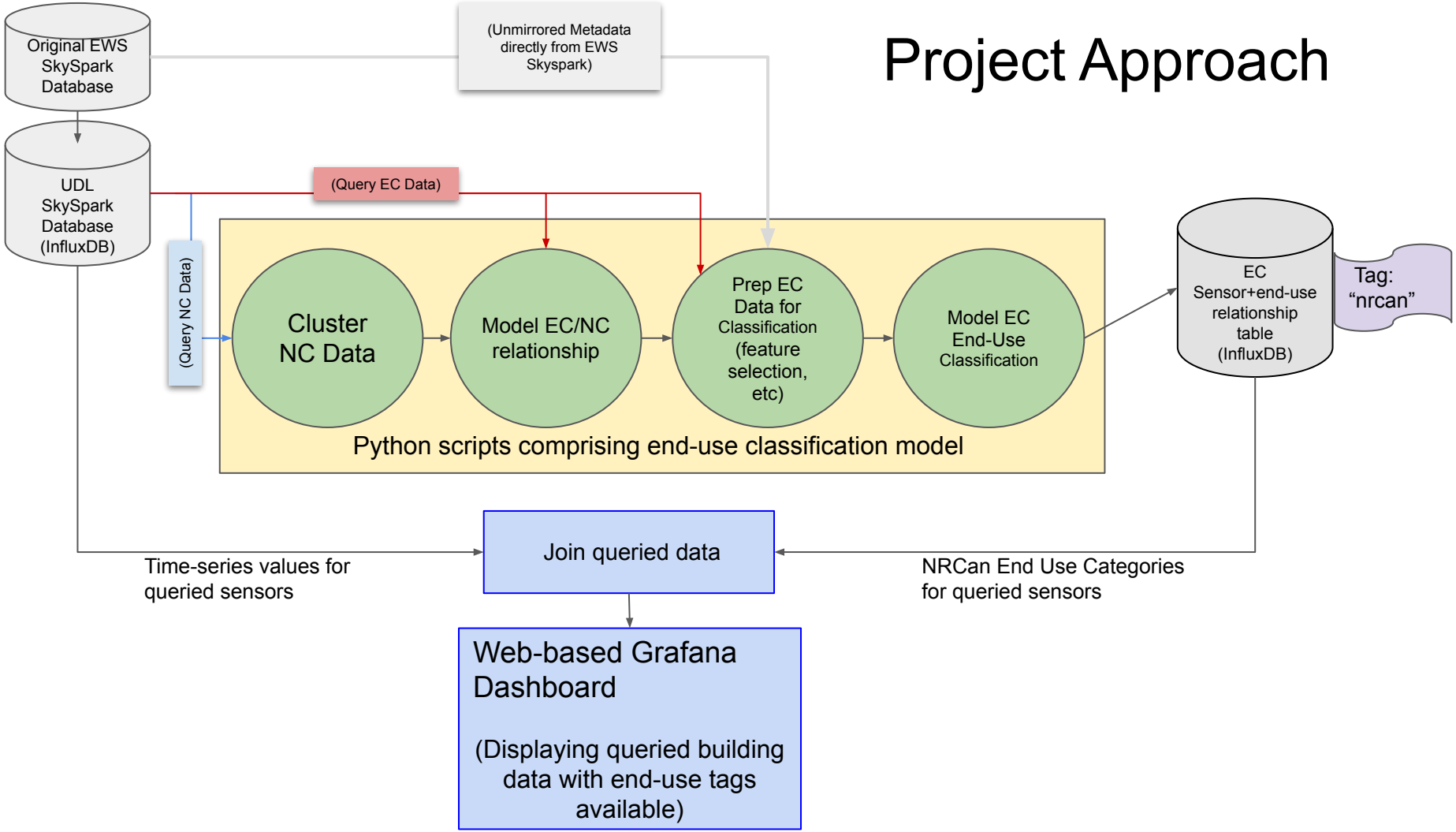
(Use NC Data as some of the predictors to help with this)

End-Use
Space Heating
Water Heating
Auxiliary Equipment
Auxiliary Motors
Lighting
Space Cooling
Street Lighting

Visualize Energy Consumption By End-Use

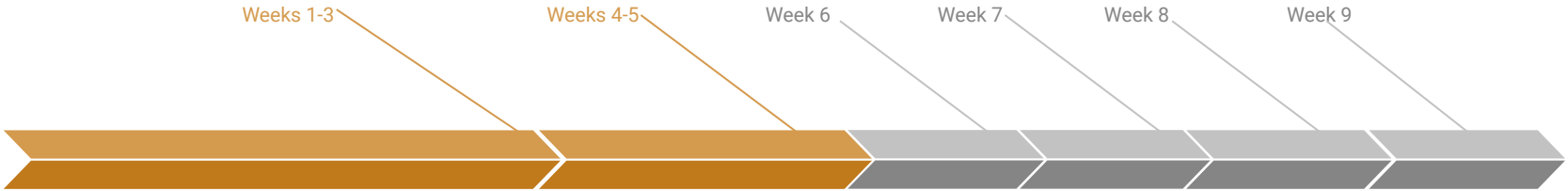


Project Approach



Progress

Project Schedule



Investigation & Data Prep

- Identify project objectives and key data features
- Understand data dictionaries
- Transform data for machine learning tasks

Feature Selection/Engineering

- Research feature selection techniques
- Merge data & metadata
- Make categorical data into smaller fields
- Aggregate different values
- Identify relevant continuous & categorical features
- Create testing and training data

Initial Modelling

- Create 3 models for each step in our project
- Run test through main.py with test dataset to get a result

Model Tuning

- Adjust parameters of model

Finalize Model & Visualization

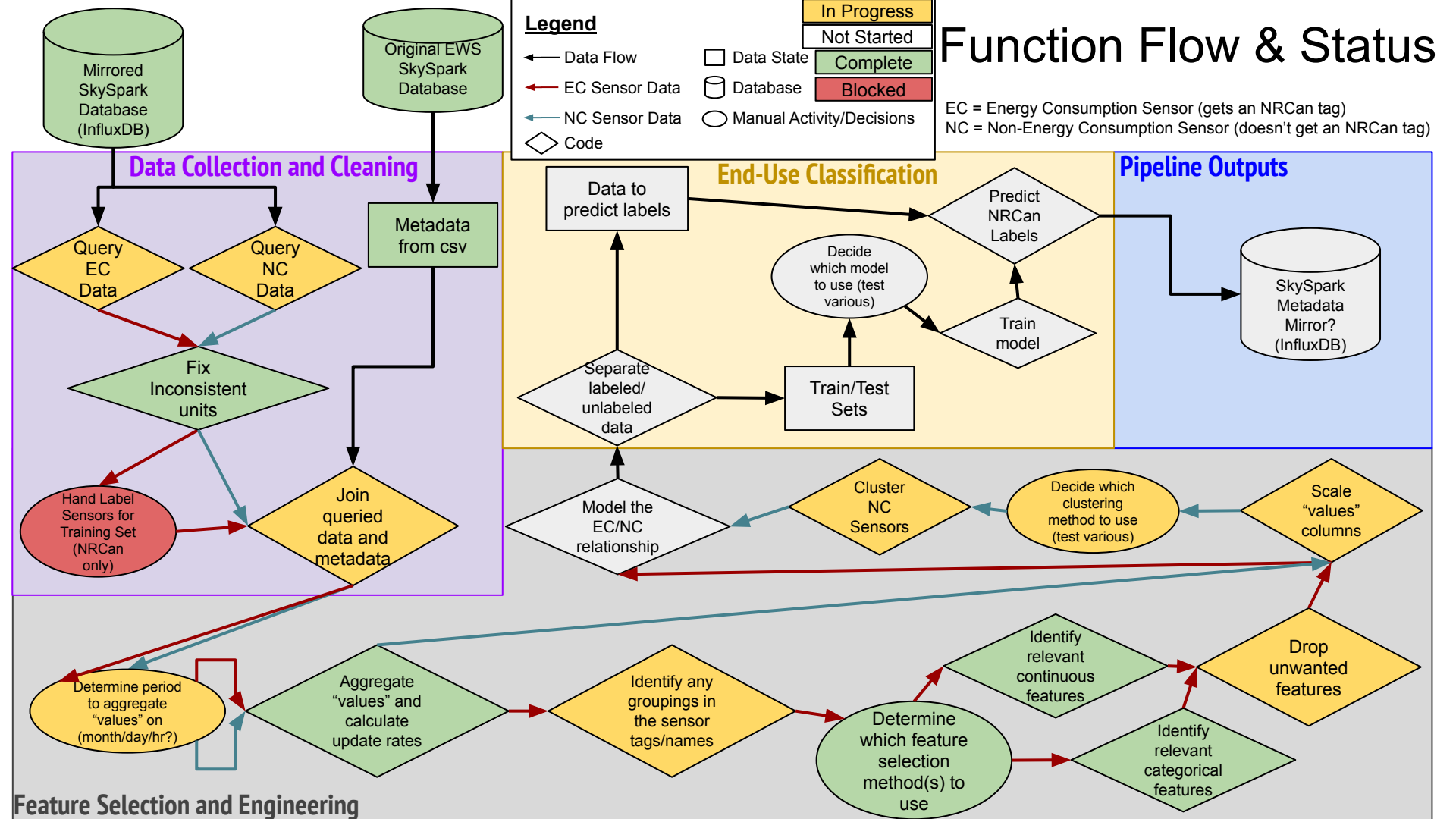
- Validate & evaluate model
- Create visualization of results

Wrap-up

- Presentation
- Final report
- Package final code

What are we currently working on?

01	**Finalize test data set**	• Confirm NRCan tags with UBC Energy and Water Services (EWS) • Investigate and choose date time range for dataset
02	**Populate the main function**	• Combine individual functions into the main.py • Add docstring to code • Confirm correct inputs & outputs for each step
03	**Develop & test various clustering models for NC sensors**	• Decide which clustering model to implement • Write code to join labels from clustering process with the raw NC sensor data
04	**Model the EC & NC sensor relationship**	• Cluster NC sensors & fit linear model to each EC sensor → use coefficients as predictors for EC sensors • Decide if we want to use LM, LASSO, or RELAXO
05	**Develop & test various classification models for EC sensors**	• Research and test supervised models (Random Forest, Neural Net, Bagging, Boosting, etc...)


Function Flow & Status
EC = Energy Consumption Sensor (gets an NRCan tag)
NC = Non-Energy Consumption Sensor (doesn't get an NRCan tag)
Legend
Data Flow
EC Sensor Data
NC Sensor Data
Code
Data State
Database
Manual Activity/Decisions
In Progress
Not Started
Complete
Blocked
Mirrored SkySpark Database (InfluxDB)
Original EWS SkySpark Database
Data Collection and Cleaning
Query EC Data
Query NC Data
Metadata from csv
Fix Inconsistent units
Hand Label Sensors for Training Set (NRCan only)
Join queried data and metadata
End-Use Classification
Data to predict labels
Predict NRCan Labels
Decide which model to use (test various)
Train model
Separate labeled/ unlabeled data
Train/Test Sets
Pipeline Outputs
SkySpark Metadata Mirror? (InfluxDB)
Model the EC/NC relationship
Cluster NC Sensors
Decide which clustering method to use (test various)
Scale "values" columns
Determine period to aggregate "values" on (month/day/hr?)
Aggregate "values" and calculate update rates
Identify any groupings in the sensor tags/names
Determine which feature selection method(s) to use
Identify relevant continuous features
Identify relevant categorical features
Drop unwanted features
Feature Selection and Engineering

What needs to be completed?

01	Compare models for effectiveness & recommend	• Research ways to evaluate models & choose performance metrics for comparison • Develop code to grid search for optimal model
02	Create Grafana visualization	• Create pie chart of Energy Consumption based on NRCan labels (if database is changed) • Create a visualization showing the counts for each NRCan sensor (if database isn't changed)
03	Write final report	• Write report with our findings • Create final flow charts of input and output

Difficulties & Roadblocks

Difficulties

- Misunderstanding about some details of goal → only a small set of the sensors need to be classified and given an NRCan tag, not ALL of the sensors. Required change of project approach.
- Developing training and testing datasets → lack of metadata clarity causing uncertainty on EC sensor end use labeling (for labelled training/test set)
- Remake our schedule→ off by 1 week because tasks were taking us longer than expected
- Identified issue with structure of UDL's influxDB. Have been working with UDL on solution.

Roadblocks

- Waiting for response from UDL for data labelling in order to create a training set
- Timing of database changes may be an issue

Questions