# Status Report:
## UBCO MDS Capstone - Urban Data Labs

WEEK 3

Connor Lee, Claudia Nikel, Eva Nguyen, Alex Tamm

# Outline

- Overall progress
- Progress made during previous week
  - Individual logs
  - Team logs
  - Difficulties and roadblocks
- Updated project approach
- Plan for next cycle

# Overall Progress

# Overall Progress

## Accomplishments

- Developed workflow processes that has worked well for the team
- Have a better understanding of data dictionaries for ML methodology
- Have a better understanding of data format for pipeline
- Have a better understanding of UDL's needs and project objectives
- Identified inconsistent and problematic data
- Began identifying potential feature selections (metadata and data fields)
- Began transforming data for ML task

# Overall Progress

## Changes Made

- Creating .py scripts rather than building in Databricks/PySpark
- Considering moving to semi-supervised clustering methods

# Previous Week’s Progress

# Progress - Individual Work Logs

| | |
|---|---|
| **Connor** | Researched Gower's distance + coded for identifying inconsistent data and when server is down + coded data transformations for ML task |
| **Claudia** | Coded for missing values + looked for irregularities in the data + created units threshold list + organized labeling for training data |
| **Alex** | Coded for identifying missing/invalid data + examined skyspark metadata for all buildings + researched data flow connections |
| **Eva** | Researched feature selection techniques + joined metadata to SkySpark data + coded for fixing data inconsistencies |

# Progress - Team Work Logs

## Accomplishments

- Made improvements to our sprint workflow by creating smaller tasks in Jira and updating Jira more often
- Created a tool to identify inconsistencies in the data
- Researched what other tools can be used for data flow connections
- Identified NRCan secondary end-use classifications

# Difficulties/Roadblocks

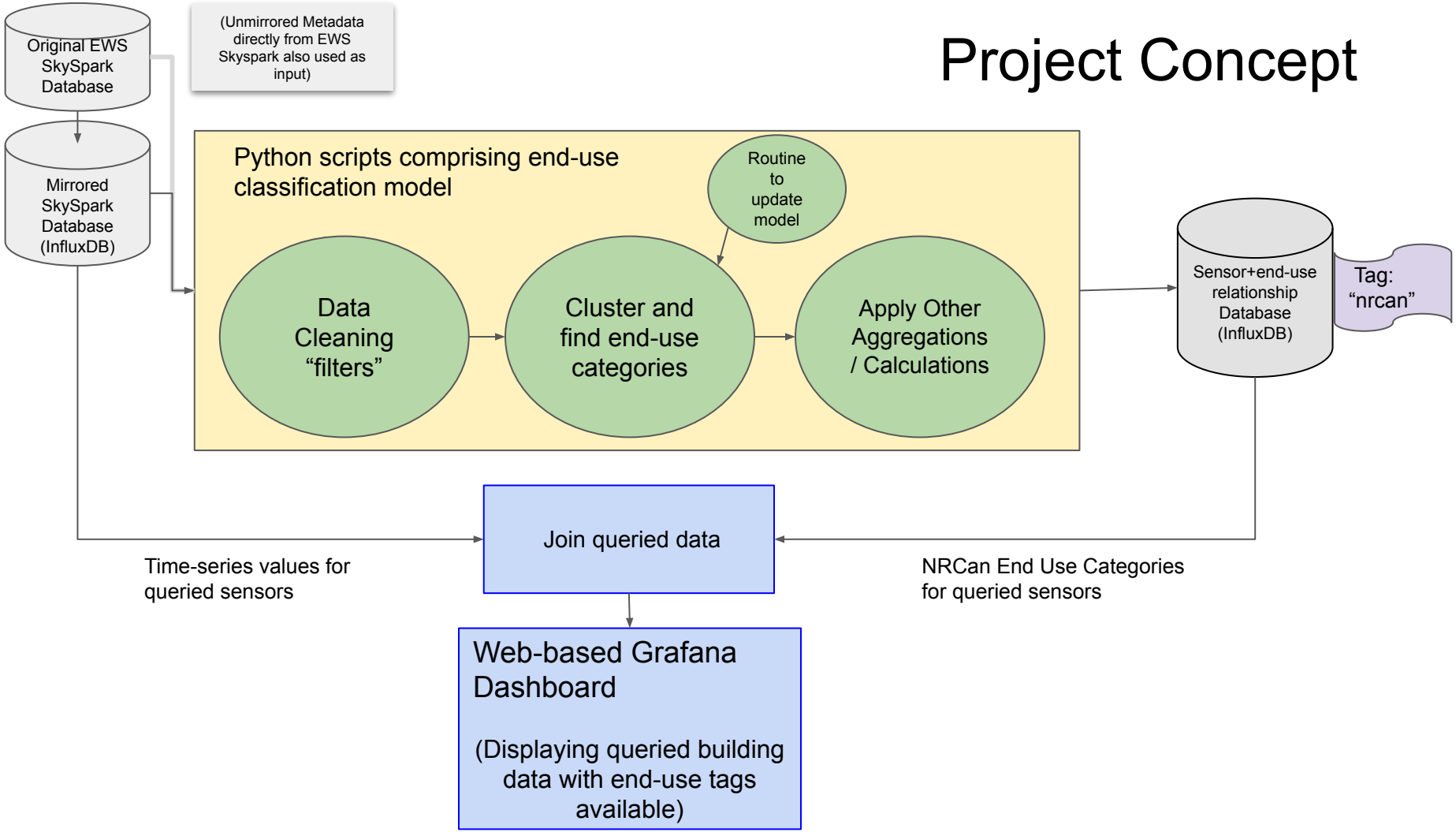
## Difficulty

- Our project approach changes slightly with every client meeting. The client provides new requests/information they would like added to the objectives.

## Roadblock

- Can not directly visualize data from UDL's influxDB in common visualization dashboard solutions because it is stored as strings. UDL either needs to improve design of database or we need a more complex pipeline for visualizing data.

# Project Concept



# Updated Project Approach

# Updated Project Approach

| NRCan Classification | Description |
|---|---|
| Space Heating | Use of mechanical equipment to heat all or part of a building |
| Water Heating | Use of energy to heat water for hot running water |
| Auxiliary Equipment | Stand-alone equipment powered directly from an electrical outlet |
| Auxiliary Motors | Refers to devices used to transform electric power into mechnical energy |
| Lighting | Non-street related lighting |
| Space Cooling | Conditioning of room's air for human comfort by a refrigeration unit |
| Street Lighting | Street-related lighting |

- From the client meeting, the following requests were made:
  - Primary Visualization = Piechart of energy consumption by NRCan End-Use
    - Only energy consumption has an end-use tag (which means most sensor readings are basically predictor variables)
  - Secondary (if there is time) Visualization is less defined but something to allow the user to drill down into sensors and view their readings (involves having appropriate grouping/Haystack-tagging for the sensors)

# Tasks for Next Cycle

# Tasks for the Next Weekly Cycle

1. Create training and testing data
2. Feature engineering
3. Implement feature selection techniques
4. Research semi-supervised models

# Questions