

Why Annotation Matters: Semi-Structuring MWO Text to Inform Fault Detection Methods

Anna Conte

Coline Bolland

Lynn Phan

Overview of Nestor

Nestor

- An NLP annotation software tagging system that performs structured data extraction from the MWO free text fields
- Allows user to **tag** individual words (**tokens**) from unstructured text
- Tokens can be aliased in order to handle abbreviations and misspellings
 - Usually, MWO text needs to be grammatically perfect for themes to be detected
 - Problem avoided: misspelled words would be grouped with correctly spelled words under the same theme if they are related
- Classifications: S (Solution), P (Problem), I (Item), SI (Solution Item), PI (Problem Item)

Word Annotation				
	Words	Classification	Tag	Note
1	action	S	action	
2	taken	U	taken	
3	hot	P	hot	
4	text	U	text	
5	rich	X	rich	
6	room	X	rm	
7	stat	I	stat	
8	operation	U	operation	
9	cold	P	cold	
10	replaced	S	replaced	
11	poc	X	poc	
12	checked	S	checked	
13	air	U	air	
14	set	U	set	
15	water	U	water	
16	unit	I	unit	
17	acu	I	acu	
18	adjusted	S	adjusted	
19	nan			
20	alarm	I	alarm	

Similar words from csv

- replaced
- replace
- relaced
- repaced
- rep
- repacked
- repalced
- replaces
- replaired
- related

Overview of Use Case

Use Case Summary

- **Purpose**
 - Help HVAC researchers determine sensor placements using data from historic MWOs
- **Maintenance Work Order Dataset**
 - Mixed use of laboratory and office space
 - Combination of scheduled maintenance, ad hoc maintenance, and occupant-submitted requests (e.g. temperature complaints)
- **Approach**
 - Focus: 1 building
 - Use KPI calculations (with Nestor tags) to identify assets that may cause the most problems or need more attention
 - → Placing sensors on these machines would help HVAC researchers collect the most data about faults

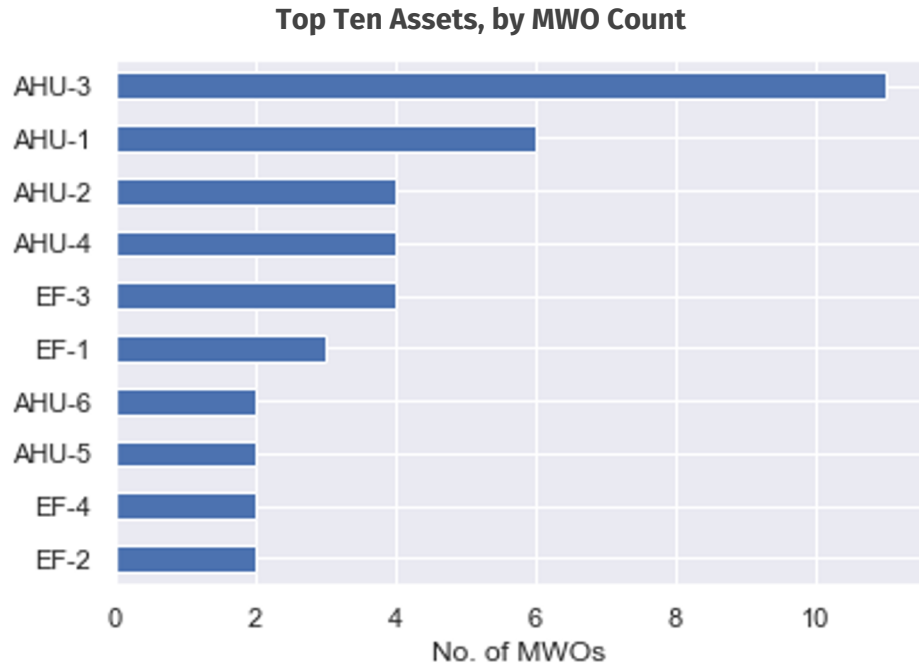
Overview of Data

- **405 rows** for this building
- Ten years (2009-2019)

- **Missing data:**
 - 87% missing “ASSETNUM”

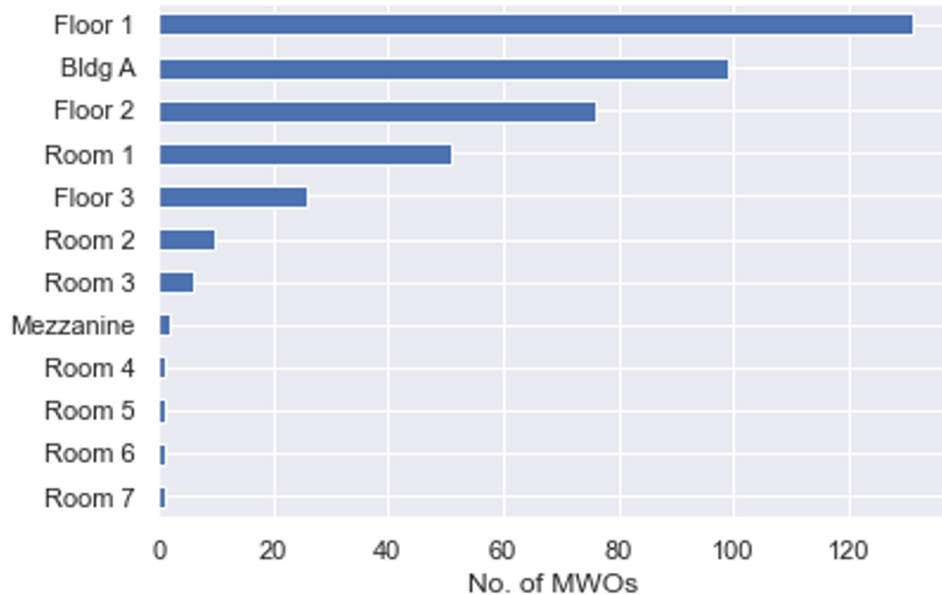
- Most frequent tag: “**set set point**” (64.6% among the tags) for the “SI Solution Item” tag
 - Large portion of the dataset is related to adjusting temperature setpoints, not necessarily mechanical issues

First Attempt to Find Worst Asset: Use MWO Counts



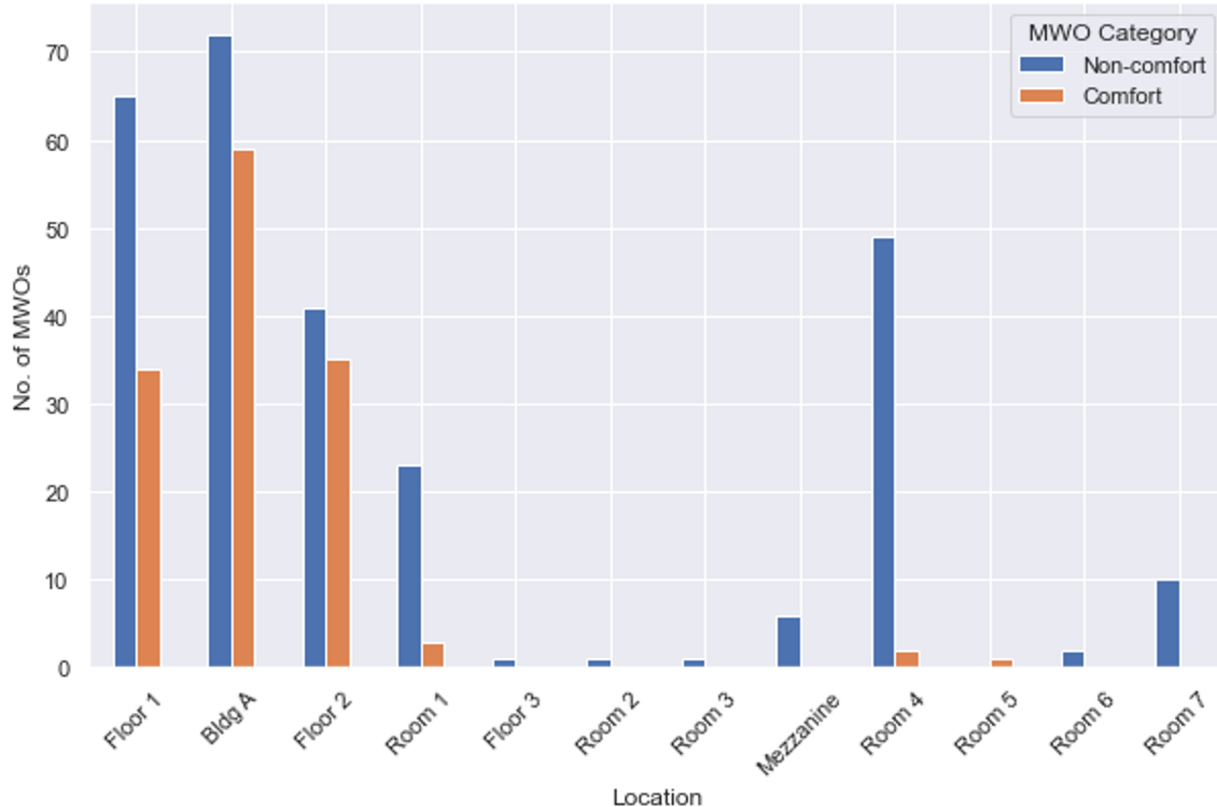
- Try to overcome missing Asset Number data with text extraction
- Using **text extraction to gather asset numbers from the free text fields**, we were able to **increase from 13% to 22% of MWOs with an asset number**, but this is still only a portion of the data.
- So, we need a different approach

Use MWO Counts: By Location



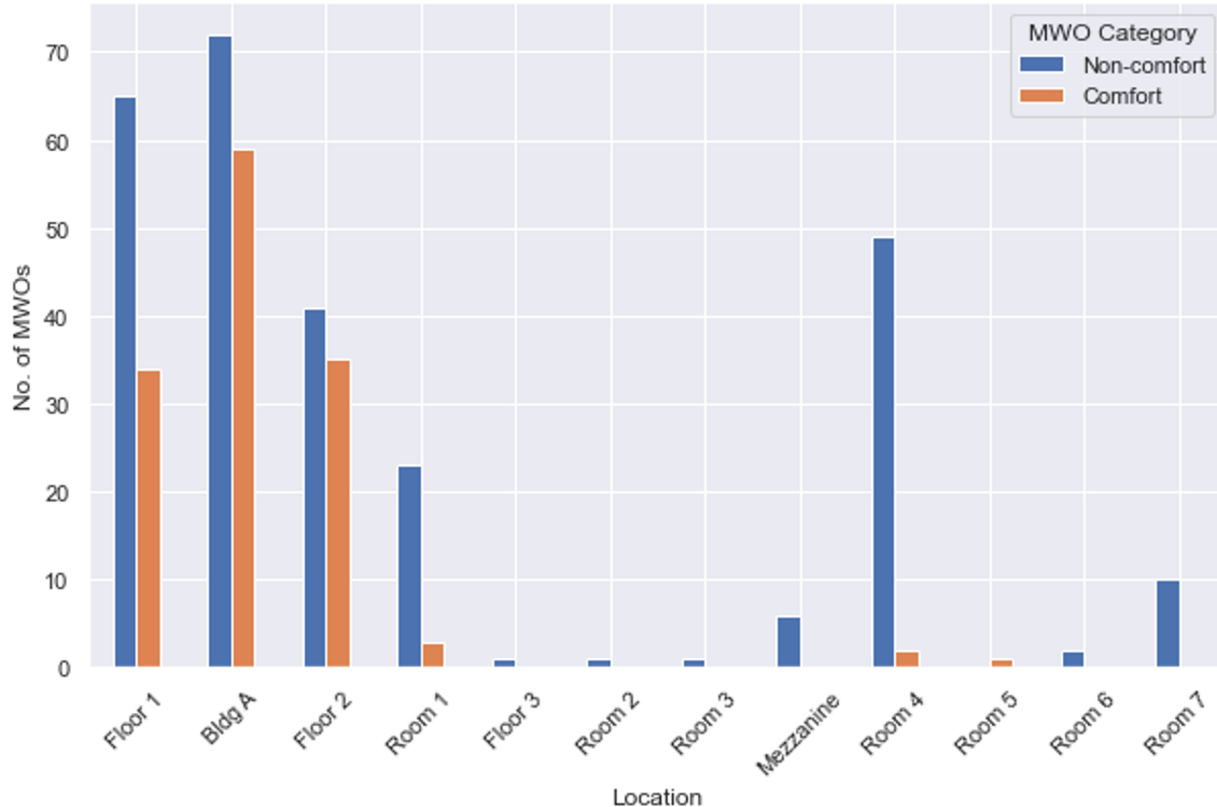
- **“LOCATION”** is a more complete data field
- About 100 MWOs are simply labeled “Building A”, but that still leaves about 75% of the dataset with useful information
- The first and second floors have the most MWOs. Assuming that these are the more occupied floors, they are likely to also have an inflated amount of MWOs due to comfort/temperature complaints.

Using Tags to Categorize MWOs



- If MWO includes “too_hot”, “too_cold”, and/or “request”, it is a “**comfort**” MWO
- Room 4 has large amount of complaints, but they are not comfort-related (mechanical space)

Using Tags to Categorize MWOs

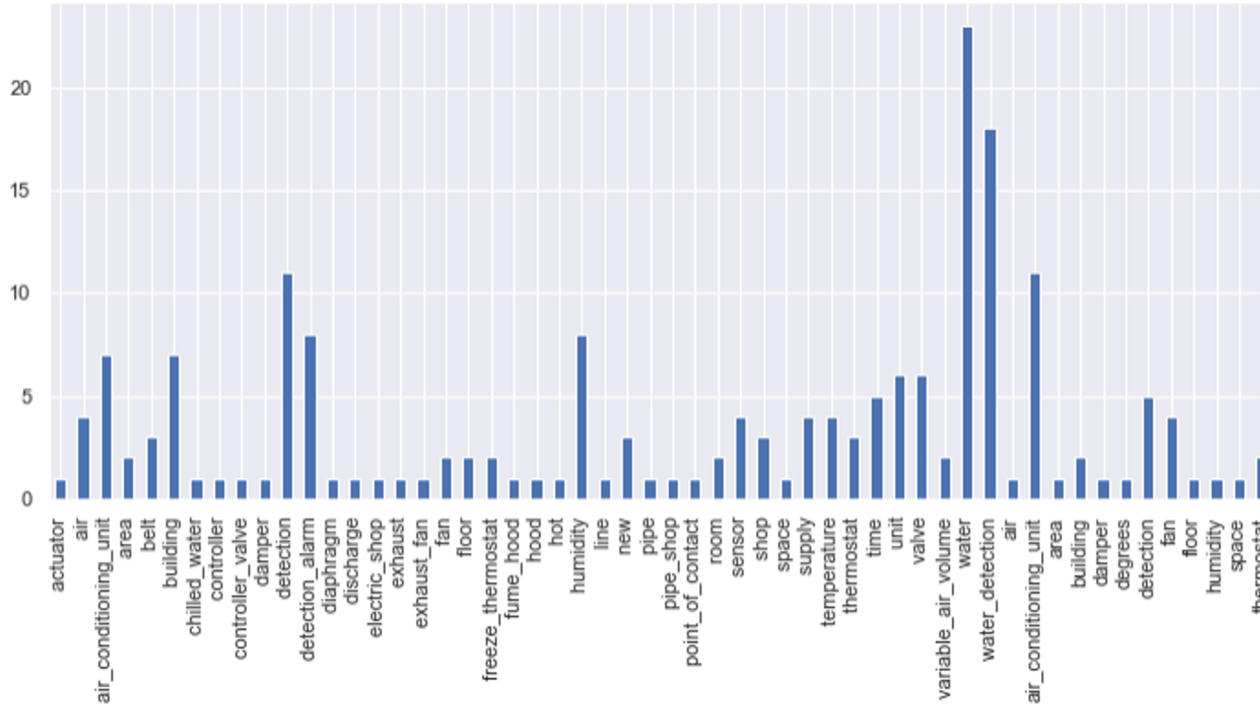


→ Other ways to apply categorization with tags:

- Planned vs. Unplanned Maintenance
 - “Monthly”, “Weekly”, “Daily”
- Specific tags
 - “Replace”, “Repair”

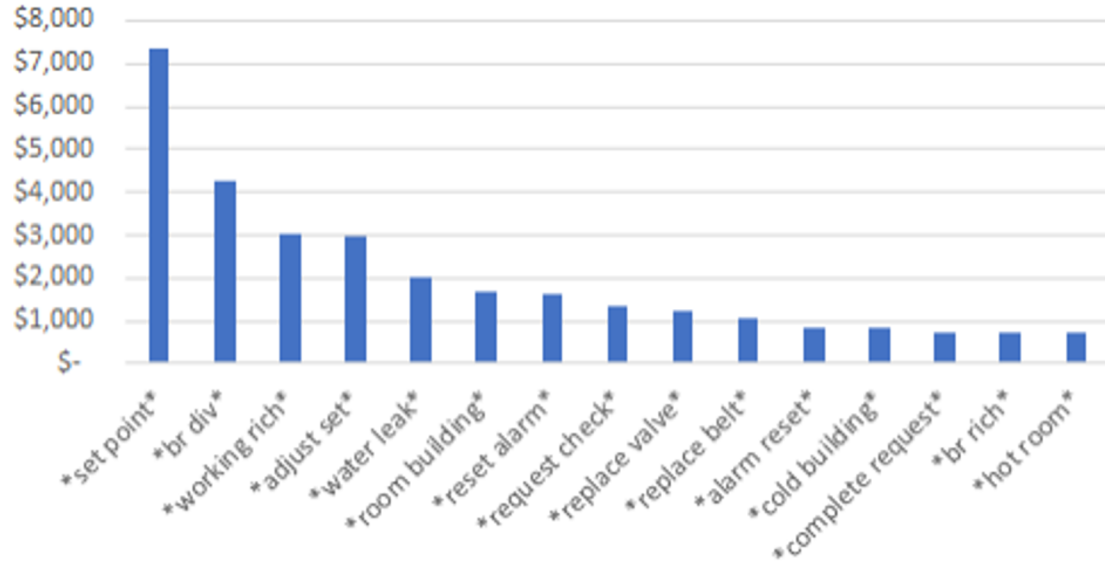
Room 4 (Mechanical Room)

Items



- Water
- Water_detection
- Air_conditioning_unit
- Detection
- Detection_alarm
- Humidity
- Valve

Worst “Topics” by Cost



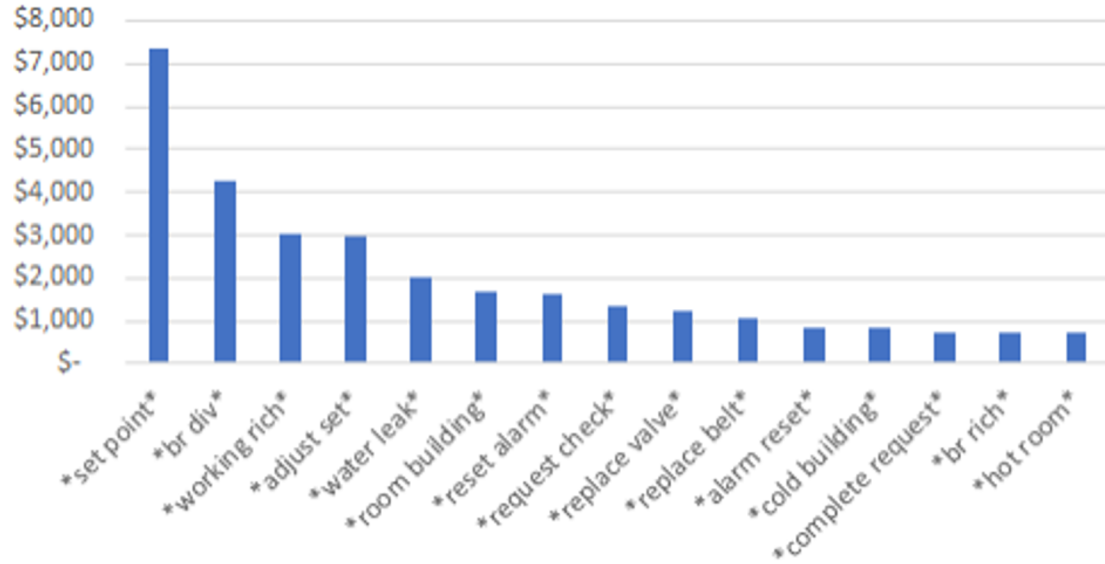
Building A: Top 15 Costliest Tags

Costliest (meaningful) tags from Nestor were:

- Set point
- Water leak
- Reset alarm
- Replace valve
- Replace belt
- Cold building
- Hot room

Assets relating to these tags may be the **most “problem-causing” (costliest to address)**

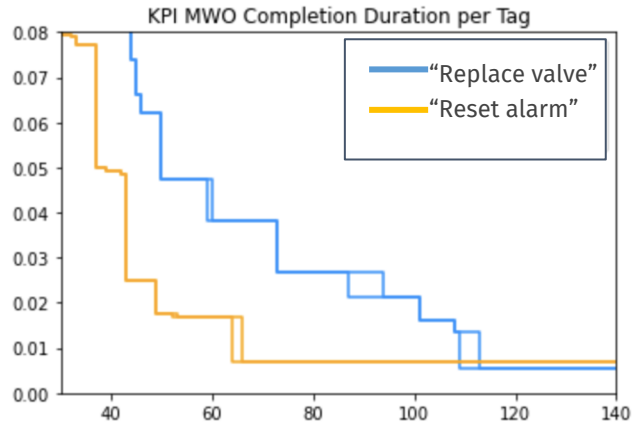
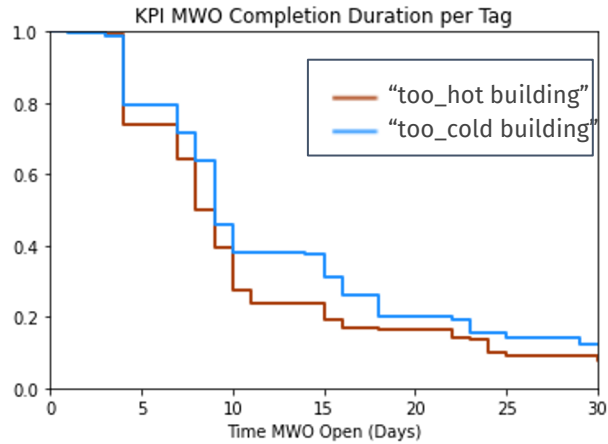
Worst “Topics” by Cost



Building A: Top 15 Costliest Tags

- **Nestor tags** allow opportunities for **data values** to be examined in terms of **concepts**:
- ◆ Cost
 - ◆ Average MWO duration
 - ◆ Average amount of time behind schedule

Examine “Topics” by *Time*



- Structured text allows to analyze topics relative to time
- Survival analysis to observe how long work orders for different problems remain open
- MWOs for complaints of “too hot” are typically addressed more quickly than complains of “too cold”
- Same logic between “Replace valve” and “Reset alarm”

Worst Assets

Based on high volume of MWOs for Floor 1, whichever unit in Room 4 (Mechanical Room) is serving the first floor is likely the best candidate for “worst asset”.

AHU-1

- Water_detection alarm
- Fan
- Freeze_thermostat
- Discharge
- Valve
- Chilled_water
- Thermostat alarm
- Alarm reset
- Drain
- Freeze
- Set_point

AHU-2

- Water_detection alarm
- Fan
- Controller_valve
- Alarm
- Air leak
- Cleaned

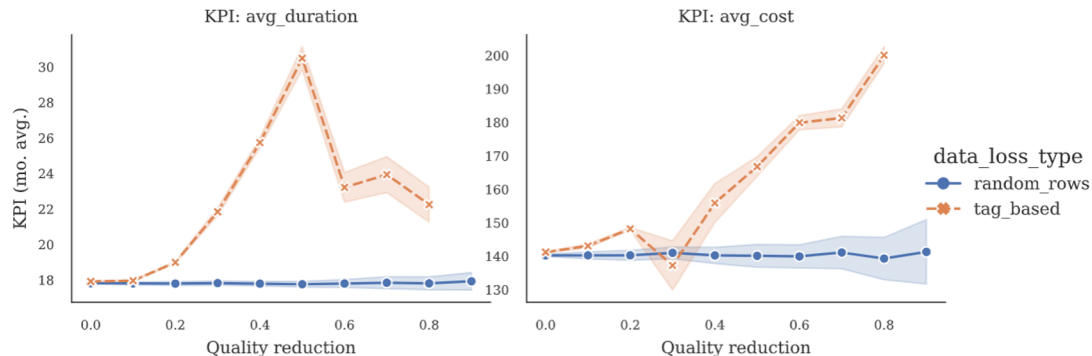
AHU-3

- Water_detection
- Damper
- Actuator
- Diaphragm
- Exhaust
- Hood
- Exhaust_fan
- Fume_hood
- Leak
- Set_point

Other Applications

- Similar analyses have been done for other facilities (manufacturing)
- Our upcoming paper (under review) examines effects of missing data
 - When data quality varies based on the context of a MWO, the resulting calculated KPIs can become very inaccurate
 - By structuring text data, we can understand these patterns

Quantifying Some Effects of Missing Data on KPI Calculations



Key Takeaways

- By semi-structuring the free-text fields, we can make greater sense of what is going on within the facility, despite having a less-than-ideal dataset
 - Highlight **types** of issues that arise in a facility
 - Understand data quality patterns within dataset
- Analyses involving semi-structuring of raw text can be applied to a variety of different contexts
 - Maintenance (HVAC, Manufacturing)
- Using Nestor tags offers more insight into MWO data than analyzing raw text and numerical data