**Problem statement:** Analyzing changes in evolving data fromadvanced habitation systems.

The goal of this task is to create a framework that monitors and provides change detection in a multimodal system. **Our objective** is two-fold (a) To provide the mining of patterns from older data as changes in data could reflect long term/ previous trends. (b) Mining for patterns over near data – as recent changes in data could indicate the recurrences of the previously known pattern or an upcoming event.

**Reported outcomes:** An exhaustive survey of related datasets that capture multi-modal environments/systems.

**Specific Aims:**

1. Searched for open source temporal datasets
   1. AMPds2 from Harvard University
2. Extract relevant data from these datasets

**Key Accomplishments:**

1. Getting familiar with reading large files using Python panda library
2. Trying to work with a dataset that seems to be similar to what was presented – by the TPOC

**Red Flags:**

1. After wrangling the datasets, we found the dataset proved useless

**Future Work:**

1. Work with the NASA bearing dataset (<http://data-acoustics.com/measurements/bearing-faults/bearing-4/>)
2. Work with the IGBT accelerated aging dataset (<https://c3.nasa.gov/dashlink/resources/134/>)

**Timeline (tentative timeline for the upcoming week)**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Future Specific aims** | **09/25** | **09/26** | **09/27** | **09/30** | **10/01** | **10/02** |
| Investigate the NASA dataset and extract relevant attributes |  |  |  |  |  |  |
| Using python panda library |  |  |  |  |  |  |
| Plot data and generate reports |  |  |  |  |  |  |

**References:**

[1] Makonin, Stephen. (2016). "AMPds2: The Almanac of Minutely Power dataset (Version 2)", <https://doi.org/10.7910/DVN/FIE0S4>, Harvard Dataverse, V2.

[2] Xiang Li, Wei Zhang, Qian Ding. (2018). Deep learning-based remaining useful life estimation of bearings using multi-scale feature extraction.

**Appendix A**

**Results**

A dataset that was found had the potential to be explored as a multi-modal system.

Due to difficulties to correlate its parts and lack of any perceptible change, it was abandoned.

Some plots were done to see how the data would behave.

4 parts of the dataset were taken into consideration: Electricity\_Billing.csv, Electricity\_WHE.csv, NaturalGas\_Billing.csv and Water\_Billing.csv.

Water\_Billing.csv had only 3 rows of data, making it useless.

Electricity\_Billing.csv had almost twice the number of rows than NaturalGas\_Billing.csv. Some strategies were tried in order to use any data from it, but no correlation can be said in such conditions.

A close up of a map

Description automatically generated

Fig 1 – Billed GJ, over the course of 2 years

A screenshot of text

Description automatically generated

Fig 2 – Average Daily GJ, over the course of 2 years

For the Electricity\_Billing.csv, first zero padding was tried in order to see any correlation.

A screenshot of a social media post

Description automatically generated

Fig 3 – kWh Usage, over the course of 2 years (padded with 0)

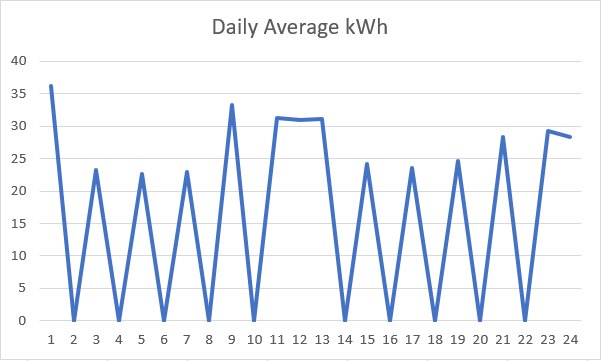


Fig 4 – Daily Average kWh, over the course of 2 years (padded with 0)

It did not work. The data was too noisy for the few existing samples.

Another tentative was made: NaturalGas\_Billing.csv was “compressed”, while Electricity\_Billing.csv remained with its original form.

A screenshot of a cell phone

Description automatically generated

Fig 5 – kWh Usage, over the course of 2 years

A screenshot of text

Description automatically generated

Fig 6 – Daily Average kWh, over the course of 2 years

A close up of a map

Description automatically generated

Fig 7 – Billed GJ, over the course of 2 years (“compressed”)

A screenshot of a cell phone screen with text

Description automatically generated

Fig 8 – Average Daily GJ, over the course of 2 years (“compressed”)

It does not have enough data to draw any conclusions.

A final attempt was made to see if a part alone could be used as change detection point.

A screenshot of a cell phone

Description automatically generated

Fig 9 – Line plots from random columns in Electricity\_WHE.csv

Apart from Q (middle top plot), all others are just noise or a line with positive slope. Even Q being a little more than just noise, it does not justify using it.