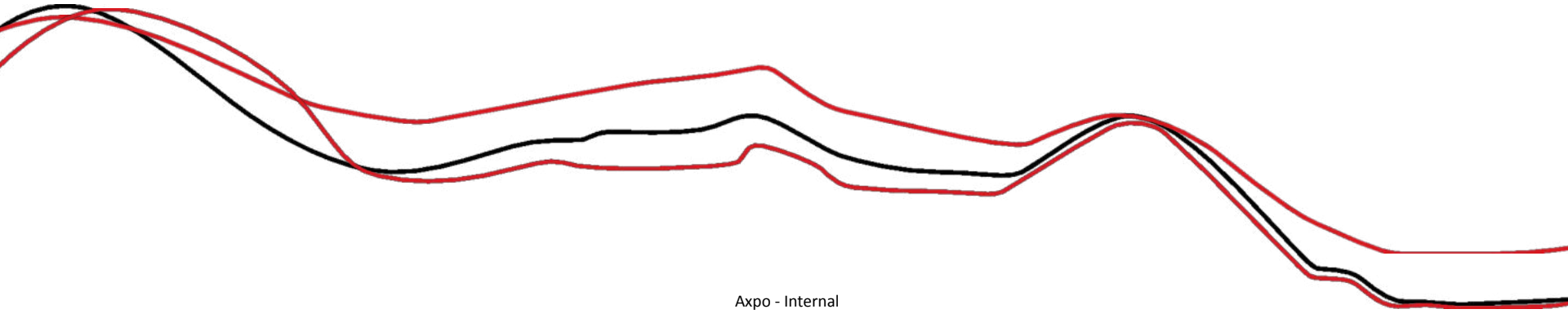
Challenge 7
Load Profile Clustering

ENERGY DATA HACKDAYS



ENERGY DATA HACKDAYS

Challenge

The primary goal of this challenge was to develop an innovative solution that leverages **public smart meter data** to generate **clusters of load profiles**, highlighting patterns into energy consumption.



Analytics process

Starting Point

id	Time	Demand (kWh)
c801982…	2021-06-19 12:00:00.000	5.84
c801982…	2021-07-17 01:45:00.000	5.77
dae3208…	2021-06-19 08:45:00.000	5.73
dae3208…	2021-06-19 13:15:00.000	5.694

Grid electricity consumption every 15 minutes in 9000 individual households in LU

ENERGY DATA HACKDAYS

Analytics process

Starting Point

id	Time	Demand (kWh)
c801982…	2021-06-19 12:00:00.000	5.84
c801982…	2021-07-17 01:45:00.000	5.77
dae3208…	2021-06-19 08:45:00.000	5.73
dae3208…	2021-06-19 13:15:00.000	5.694

Grid electricity consumption every 15 minutes in 9000 individual households in LU

reduce to more easily computable 1775 IDs

Cleaning

filter infrequent consumers

ENERGY DATA HACKDAYS

Analytics process

Starting Point

id	Time	Demand (kWh)
c801982…	2021-06-19 12:00:00.000	5.84
c801982…	2021-07-17 01:45:00.000	5.77
dae3208…	2021-06-19 08:45:00.000	5.73
dae3208…	2021-06-19 13:15:00.000	5.694

Grid electricity consumption every 15 minutes in 9000 individual households in LU

reduce to more easily computable 1775 IDs

Cleaning

filter infrequent consumers

Supplemental Data



Features List

Data Transformations

```
mean, std, min, max

.diff()

.groupby(["id","year-month"])
```

ENERGY DATA HACKDAYS

Analytics process

Starting Point

id	Time	Demand (kWh)
c801982…	2021-06-19 12:00:00.000	5.84
c801982…	2021-07-17 01:45:00.000	5.77
dae3208…	2021-06-19 08:45:00.000	5.73
dae3208…	2021-06-19 13:15:00.000	5.694

Grid electricity consumption every 15 minutes in 9000 individual households in LU

reduce to more easily computable 1775 IDs

Cleaning

filter infrequent consumers

Features List

Supplemental Data



Clustering

silhouette or elbow method to find optimal amount

selecting features

Data Transformations

```
mean, std, min, max

.diff()

.groupby(["id","year-month"])
```

ENERGY DATA HACKDAYS

Analytics process

Starting Point

id	Time	Demand (kWh)
c801982…	2021-06-19 12:00:00.000	5.84
c801982…	2021-07-17 01:45:00.000	5.77
dae3208…	2021-06-19 08:45:00.000	5.73
dae3208…	2021-06-19 13:15:00.000	5.694

Grid electricity consumption every 15 minutes in 9000 individual households in LU

reduce to more easily computable 1775 IDs

Cleaning

filter infrequent consumers

Supplemental Data



Features List

Data Transformations

```
mean, std, min, max

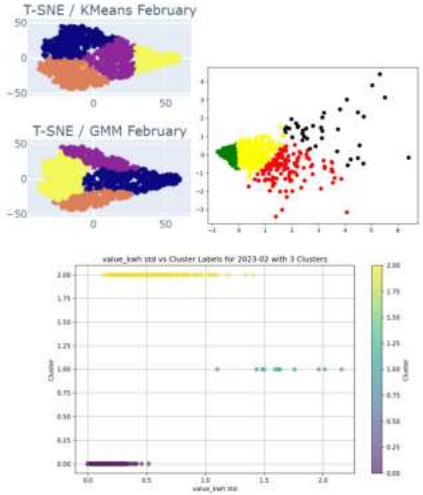
.diff()

.groupby(["id","year-month"])
```

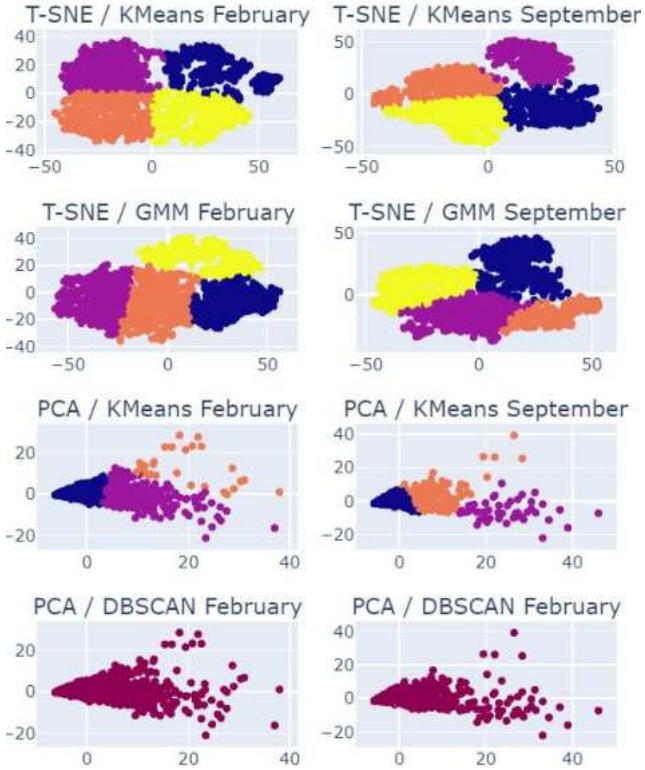
silhouette or elbow method to find optimal amount

Clustering

selecting features



Clustering over all calculated features with different dimensionality reduction and clustering algorithms



Used features:

- everything !

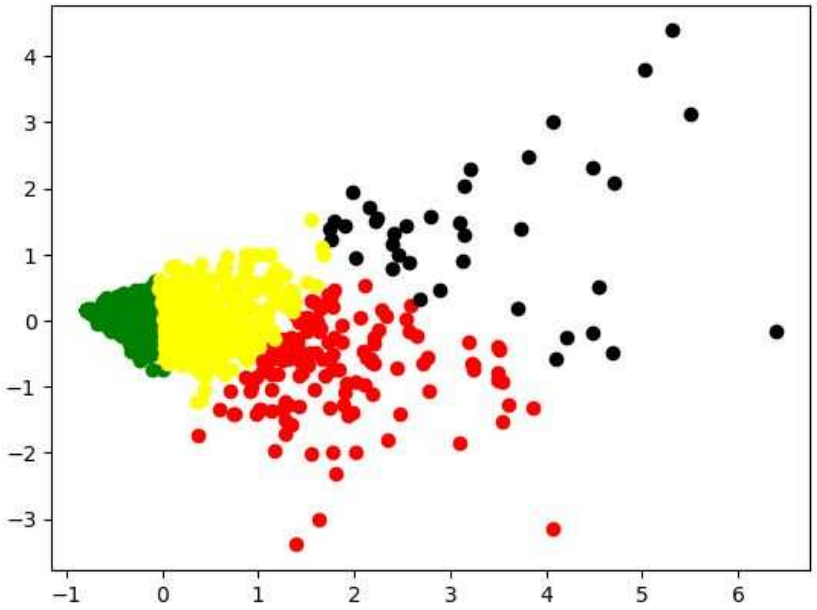
Interpretation:

- More does not equal better in this case :(

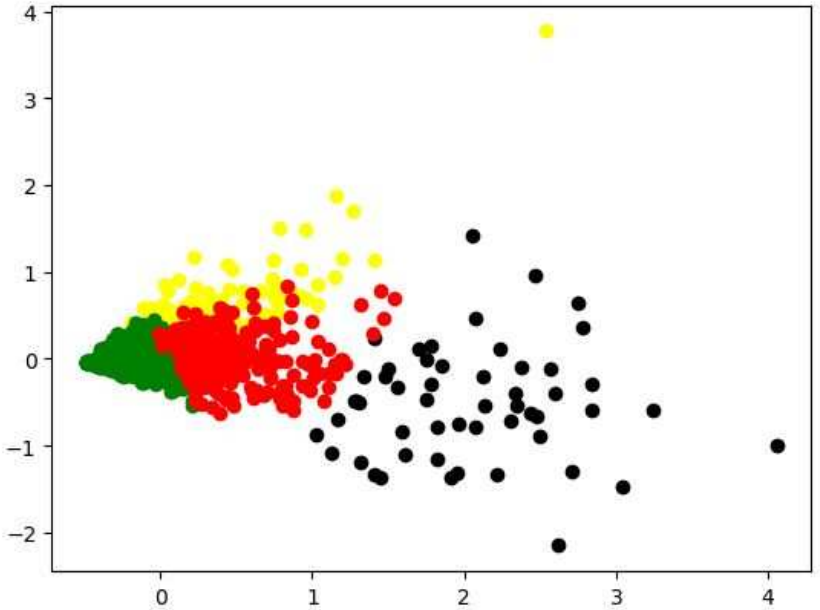
ENERGY DATA HACKDAYS

Clustering with weather features

February 2023



September 2023



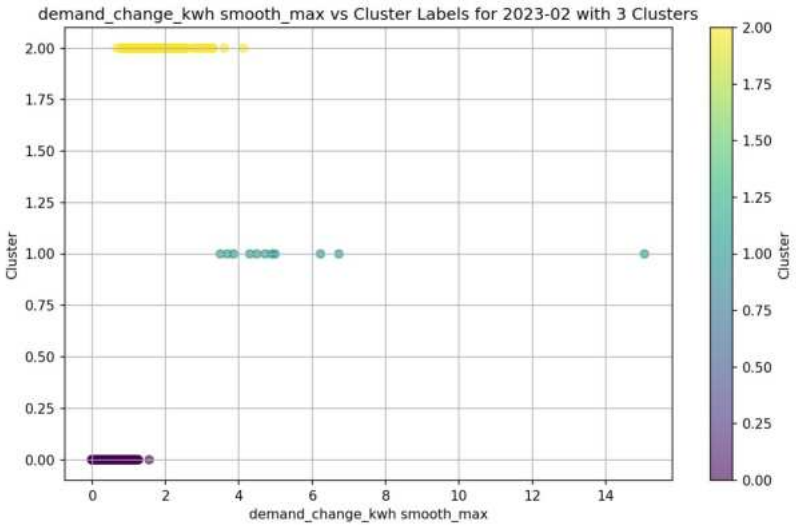
ENERGY DATA HACKDAYS

Clustering with weather features

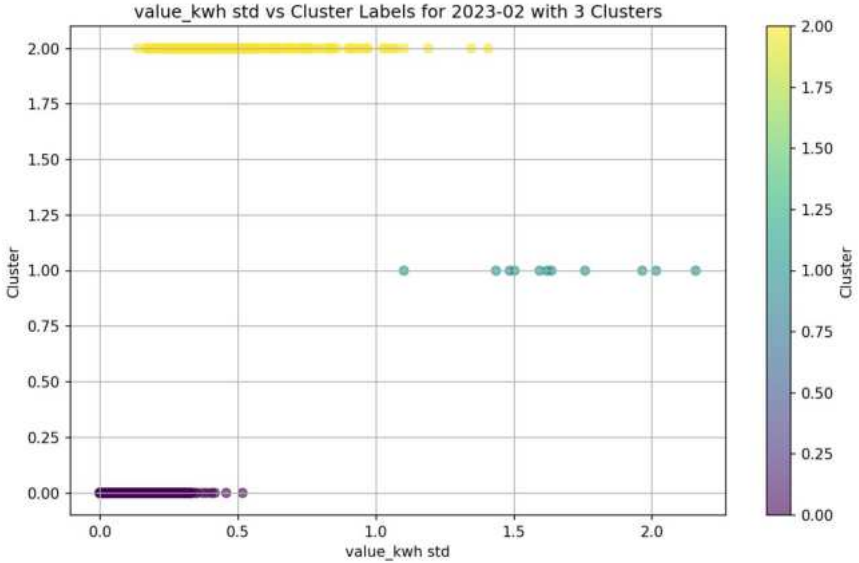
	FEBRUARY 2023			
SEPTEMBER 2023	**Cluster 1**	**Cluster 2**	**Cluster 3**	**Cluster 4**
Cluster 1	2.33%	8.9%	25%	35%
Cluster 2	**89.6%**	33.5%	**40%**	38.46%
Cluster 3	0.4%	29%	2.26%	**0%**
Cluster 4	7.5%	28%	32.6%	26%

Proportions of February cluster members in September cluster

Clustering with Hourly Demand Change

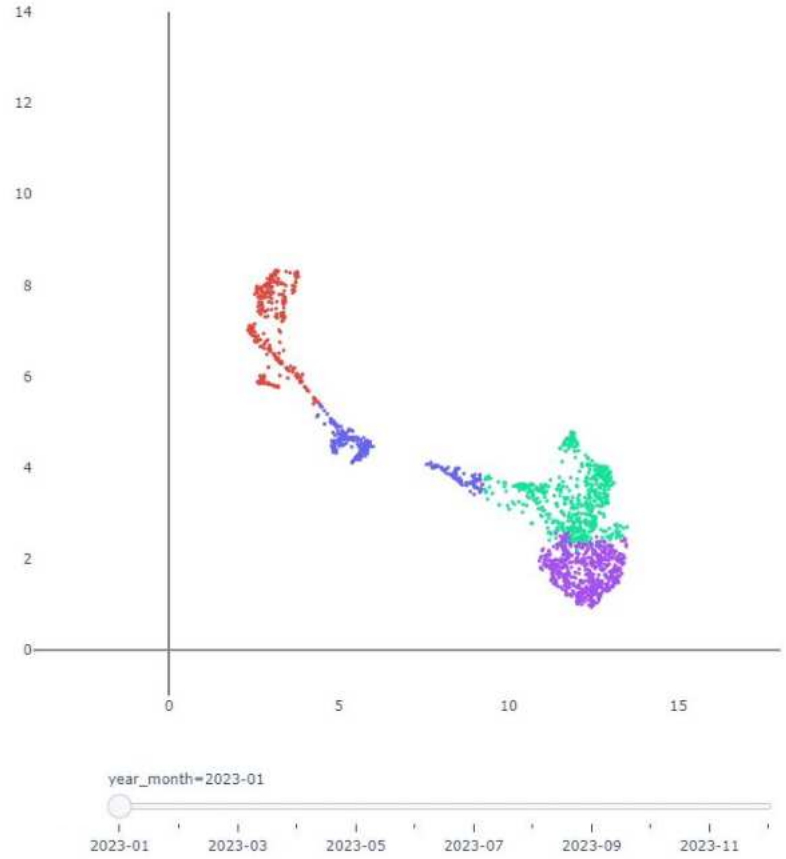


Maximum demand change over an hour



Standard deviation of demand

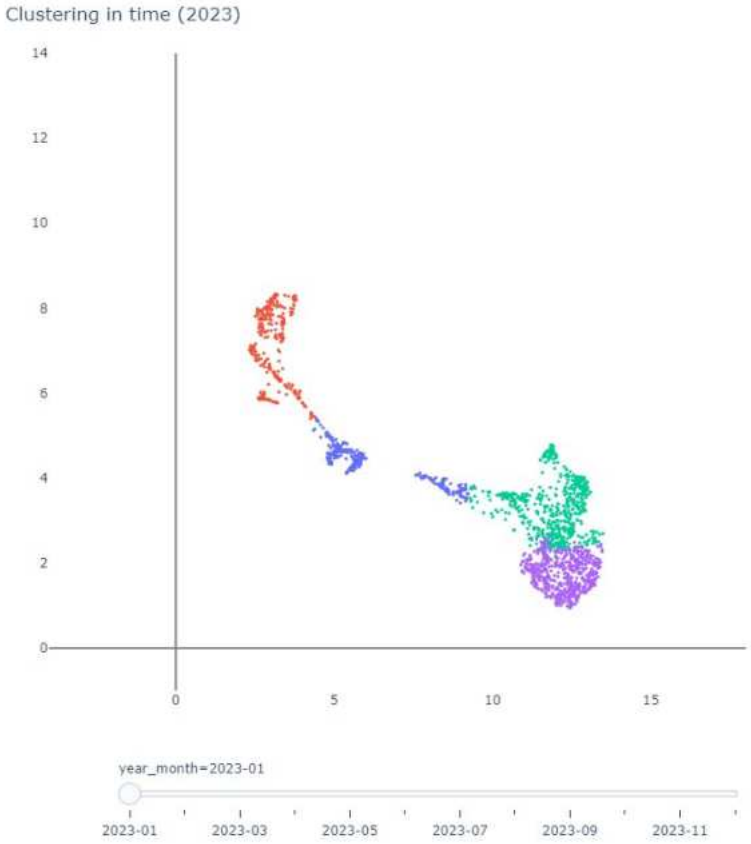
K-Means with UMAP using various features



Used features [50]:

- sunshine_correlation
- temperature_correlation
- day_of_week_mean [0-23]
- mean
- standard deviaton
- min
- max
- weekday_mean [0-6]
- weekday_stddv [0-6]
- cnt_outlier_max
- cnt_outlier_min
- TSMI features

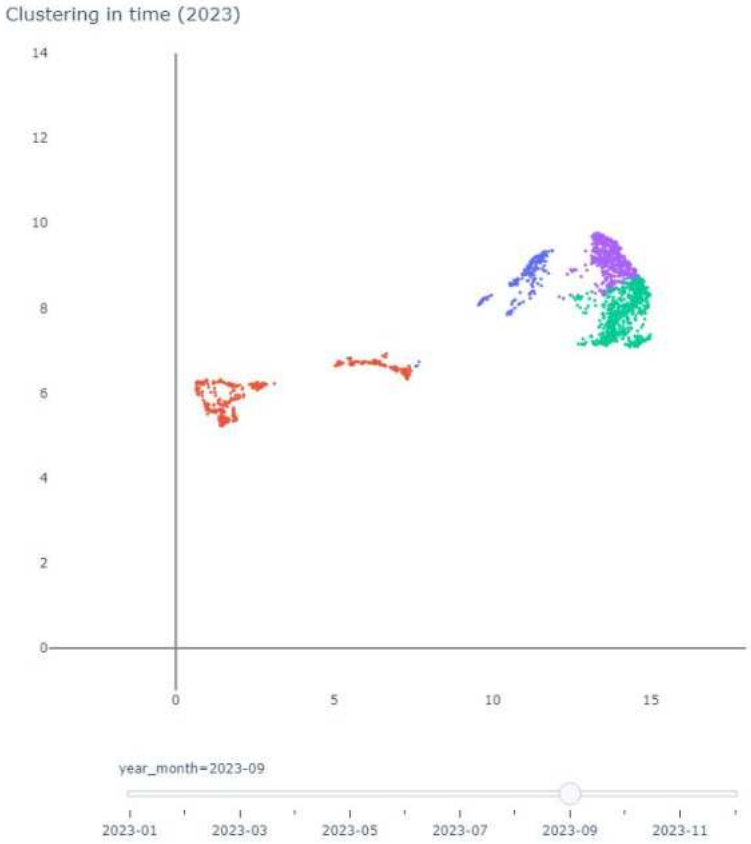
K-Means with UMAP using various features



Used features [50]:

- sunshine_correlation
- temperature_correlation
- day_of_week_mean [0-23]
- mean
- standard deviaton
- min
- max
- weekday_mean [0-6]
- weekday_stddv [0-6]
- cnt_outlier_max
- cnt_outlier_min
- TSMI features

K-Means with UMAP using various features

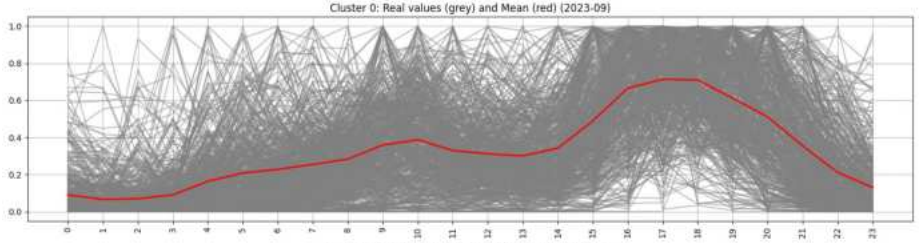


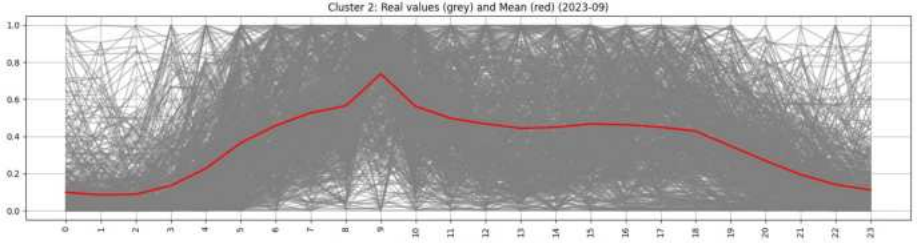
Used features [50]:

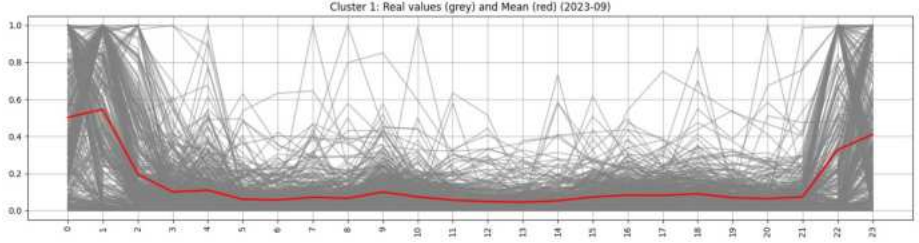
- sunshine_correlation
- temperature_correlation
- day_of_week_mean [0-23]
- mean
- standard deviaton
- min
- max
- weekday_mean [0-6]
- weekday_stddv [0-6]
- cnt_outlier_max
- cnt_outlier_min
- TSMI features

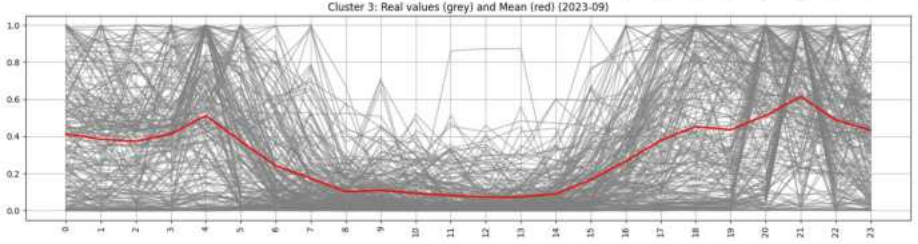
ENERGY DATA HACKDAYS

K-Means with UMAP using various features









ENERGY DATA HACKDAYS

Learnings

Databricks is awesome - Locally computation can be challenging

Features are nice, but be selective!

