

ELECTRIC CONSUMPTION RATE

PYTHON KHIVA LIBRARY

info@shapelets.io

INTRODUCTION

In this use case we analyse the energy data from 100 commercial/ industrial sites and find out that the consumption rates assigned to each site are correct.

We use a predictive modelling to perform this analysis.

The analysis performed to find out the correct use of electric consumption rates is based on:

- Geographical research.
- Features extraction to identify the time series where to apply a predictive modelling.
- Similarity research.

We are considering that consumption rates are assigned according to the kind of sub-industry sites are established.

Khiva library is used to re-dimension the data, extract the features and compose the features matrix for the predictive modelling and obtain similarities between sites of interest. All algorithms are executed in GPU*.

ELECTRIC RATES

- Commercial Real Estate – Commercial Property
- Shopping Center/Shopping Mall – Commercial Property
- Business Services – Commercial Property
- Bank/Financial Services – Commercial Property
- Food Processing – Light Industrial
- Manufacturing – Light Industrial
- Other Light Industrial – Light Industrial
- Grocer/Market – Food Sales & Storage
- Primary/Secondary School – Education

	SITE_ID	INDUSTRY	SUB_INDUSTY
0	6	Commercial Property	Shopping Center/Shopping Mall
1	8	Commercial Property	Shopping Center/Shopping Mall
2	9	Commercial Property	Corporate Office
3	10	Commercial Property	Shopping Center/Shopping Mall
4	12	Commercial Property	Business Services
5	13	Commercial Property	Commercial Real Estate
6	14	Commercial Property	Business Services
7	21	Commercial Property	Shopping Center/Shopping Mall
8	22	Commercial Property	Bank/Financial Services
9	25	Commercial Property	Shopping Center/Shopping Mall
10	29	Commercial Property	Shopping Center/Shopping Mall
11	30	Commercial Property	Bank/Financial Services
12	31	Commercial Property	Commercial Real Estate
13	32	Commercial Property	Shopping Center/Shopping Mall
14	36	Commercial Property	Business Services
15	41	Commercial Property	Corporate Office
16	42	Commercial Property	Shopping Center/Shopping Mall
17	44	Commercial Property	Shopping Center/Shopping Mall

Figure 1: Metadata

WHERE ARE THE SITES LOCATED?

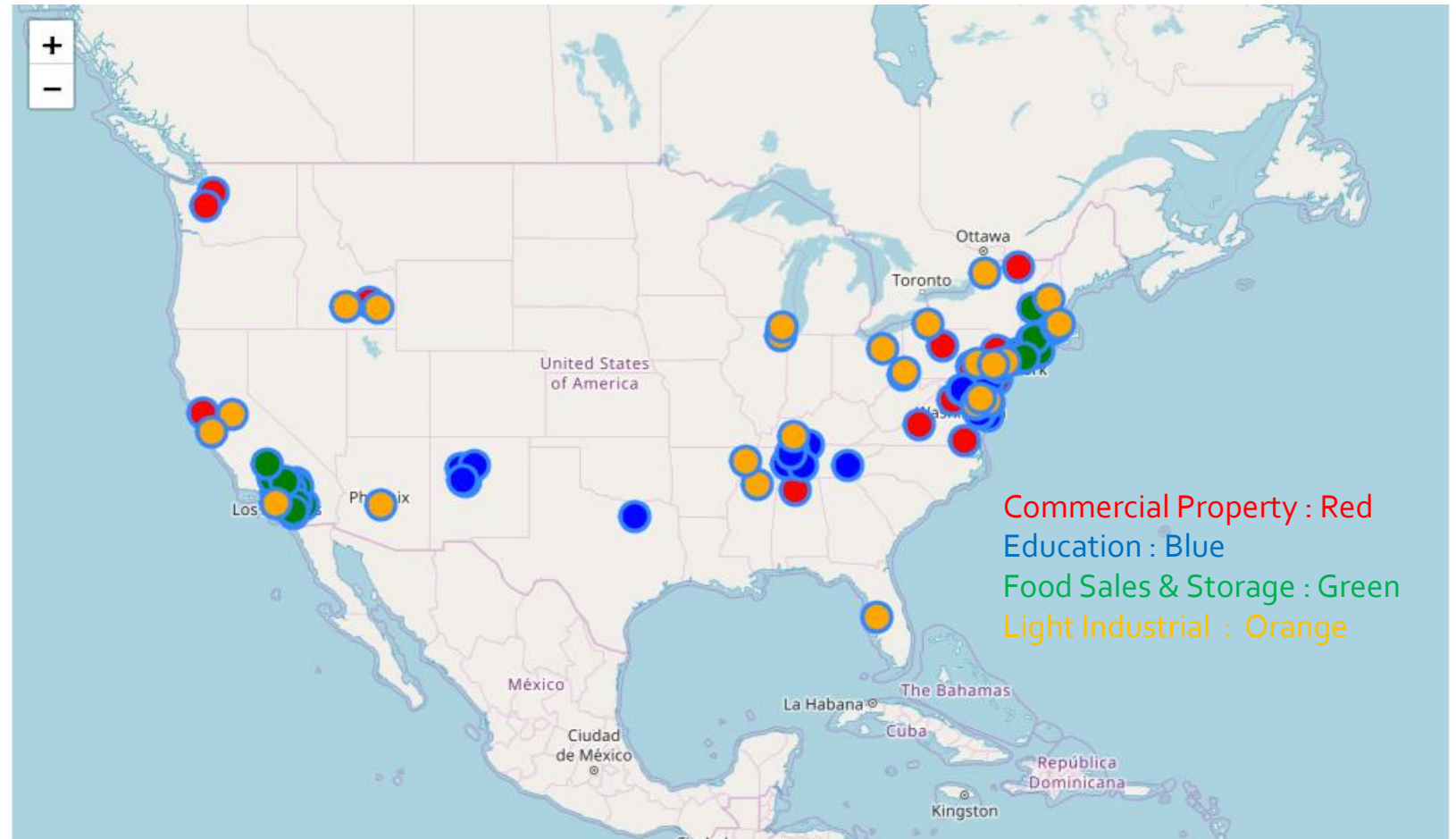


Figure 2: Local locations organised by Industry

WHAT KIND DATA ARE WE GOING ANALYSE?

- Data corresponding to electrical consumption during year 2012 of 100 sites.
- Data being collected each 5 minutes. 12 points per hour.
- 10,531,288 Data points in total.
- Re-dimension of the time series of each local using Visvalingam's algorithm. The target is going to simulate a data collection of 2 points per hour. Being the resulting data per local equal to 16,666.
- 1,666,600 Data points in total.

FEATURES EXTRACTION. FEATURES MATRIX & LABELS MATRIX

Features extracted:

- | | |
|-------------------------------|--|
| 1. Abs_energy | 15. Maximum |
| 2. Absolute_sum_of_changes | 16. Mean_absolute_change |
| 3. Count_above_mean | 17. Minimum |
| 4. Count_below_mean | 18. Number_crossing_m |
| 5. First_location_of_maximum | 19. Mean |
| 6. First_location_minimum | 20. Median |
| 7. Has_duplicates | 21. Mean_change |
| 8. Has_duplicate_max | 22. Ratio_value_number_to_time_series_length |
| 9. Kurtosis | 23. Skewness |
| 10. Last_location_of_máximum | 24. Standard_deviation |
| 11. Last_location_of_mínimum | 25. Sum_of_reoccurring_values |
| 12. Has_duplicate_min | 26. Sum_values |
| 13. Longest_strike_above_mean | 27. Variance |
| 14. Longes_strike_below_mean | 28. Variance_larger_tan_sdtandard_deviation |

Performance: 12.84 seconds in extracting 28 features of 100 time series of 16,666 data points → 1,666,600 points in total.

Label matrix : Sub-industries (Electric consumption rate)

PREDICTIVE MODELLING

- Pre-processing:
 - Shuffle
 - Scale of the Features Matrix
- Model selection (Based on Scikit Learn recommendations) :

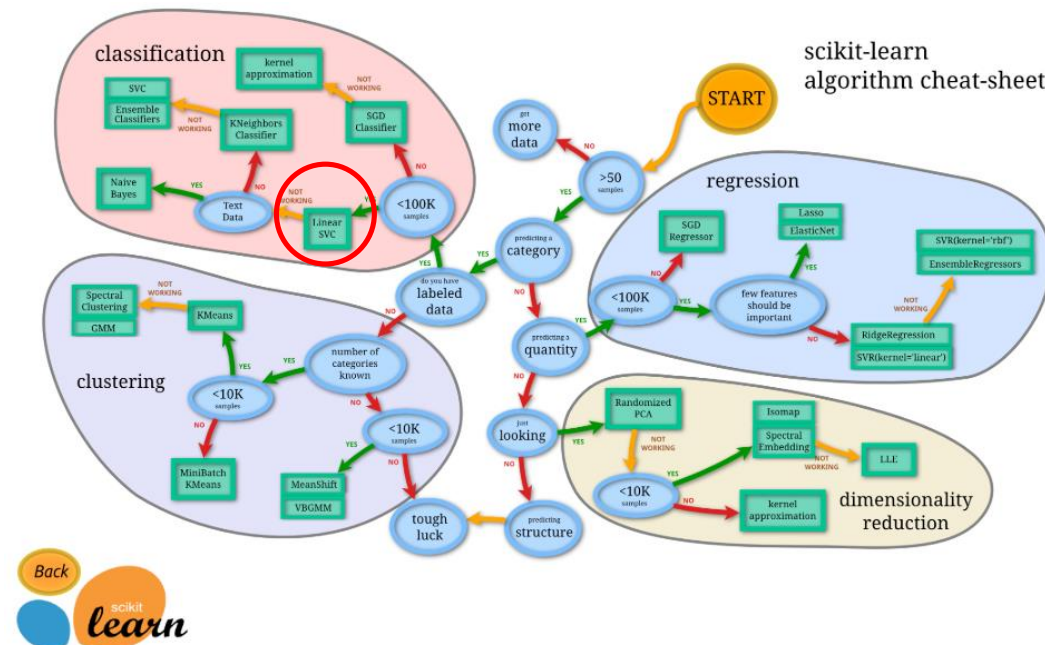


Figure 3: Model selection recommendation by Scikit-learn

PREDICTIVE MODELLING. RESULTS

- Results:
 - TEST: ['Grocer/Market' 'Primary/Secondary School' 'Food Processing' 'Food Processing' 'Grocer/Market' 'Primary/Secondary School' 'Grocer/Market' 'Food Processing' 'Shopping Center/Shopping Mall' 'Shopping Center/Shopping Mall' 'Manufacturing' 'Primary/Secondary School' 'Shopping Center/Shopping Mall' 'Primary/Secondary School' 'Grocer/Market']
 - PREDICTION: ['Grocer/Market' 'Primary/Secondary School' 'Food Processing' 'Food Processing' 'Grocer/Market' 'Primary/Secondary School' 'Grocer/Market' 'Food Processing' 'Shopping Center/Shopping Mall' 'Shopping Center/Shopping Mall' 'Food Processing' 'Primary/Secondary School' 'Primary/Secondary School' 'Food Processing' 'Grocer/Market']
- NUMBER OF ERRORS: 3
- ERROR RATE: 20 %
- ACCURACY: 80%

PREDICTION DETAILS

- **USE CASE**

- SITE 92 → Pred: Food Processing.
Test: Primary/Secondary School.
Different Industry.
- SITE 49 → Pred: Primary/Secondary School.
Test: Shopping Center/Shopping Mall.
Different Industry.
- SITE 761 → Pred: Food Processing.
Test: Manufacturing.
Same Industry.

	industry_test	industry_pred	y_test \
0	Food Sales & Storage	Food Sales & Storage	Grocer/Market
1	Education	Education	Primary/Secondary School
2	Light Industrial	Light Industrial	Food Processing
3	Light Industrial	Light Industrial	Food Processing
4	Food Sales & Storage	Food Sales & Storage	Grocer/Market
5	Education	Education	Primary/Secondary School
6	Food Sales & Storage	Food Sales & Storage	Grocer/Market
7	Light Industrial	Light Industrial	Food Processing
8	Commercial Property	Commercial Property	Shopping Center/Shopping Mall
9	Commercial Property	Commercial Property	Shopping Center/Shopping Mall
10	Light Industrial	Light Industrial	Manufacturing
11	Education	Education	Primary/Secondary School
12	Commercial Property	Education	Shopping Center/Shopping Mall
13	Education	Light Industrial	Primary/Secondary School
14	Food Sales & Storage	Food Sales & Storage	Grocer/Market

	y_pred	file
0	Grocer/Market	496
1	Primary/Secondary School	144
2	Food Processing	731
3	Food Processing	673
4	Grocer/Market	281
5	Primary/Secondary School	197
6	Grocer/Market	400
7	Food Processing	766
8	Shopping Center/Shopping Mall	10
9	Shopping Center/Shopping Mall	45
10	Food Processing	761
11	Primary/Secondary School	213
12	Primary/Secondary School	49
13	Food Processing	92
14	Grocer/Market	399

Figure 4: Prediction and test comparisons.

CASE I. GEOGRAPHICAL RESEARCH

SITE 92 →

Pred: Food Processing.
Test: Primary/Secondary School.
Different Industries.

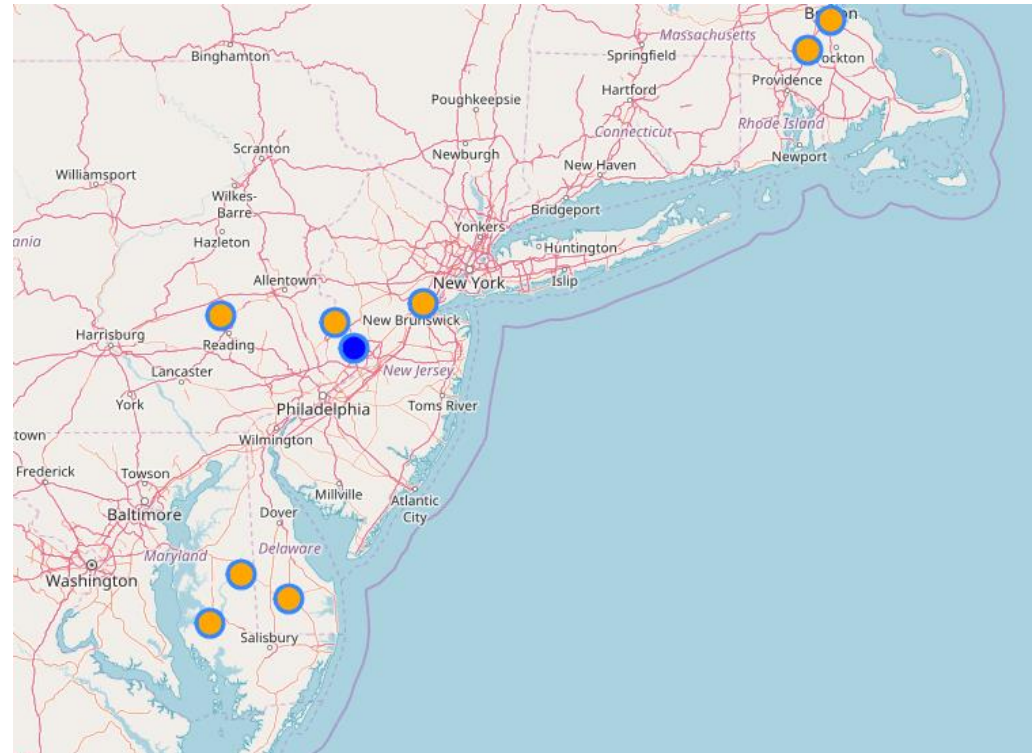


Figure 5: Location of site 92 and other Food Processing sites.

CASE I. QUICK SIMILARITY COMPARISON. FIRST MONTH

SITE 92 →

Pred: Food Processing.
Test: Primary/Secondary School.
Different Industries.

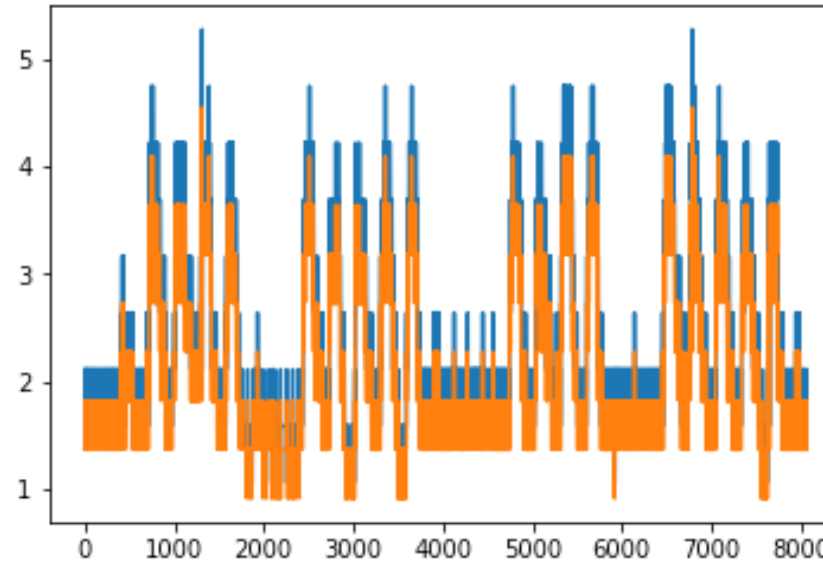


Figure 6: First month of site 92 against site 887

SITE_ID	INDUSTRY	SUB_INDUSTRY	SQ_FT	LAT	LNG	TIME_ZONE	TZ_OFFSET
26	92	Education	Primary/Secondary School	105530	40.243922	America/New_York	-04:00
99	887	Light Industrial	Food Processing	381939	38.83195	America/New_York	-04:00

Figure 7: Metadata of site 92 and 887.

CASE I. CONCLUSIONS

SITES 92 →

Pred: Food Processing.

Test: Primary/Secondary School.

Different Industries.

According to our analysis and due to the very similar behaviour between site 92 and 887 we decide to send an inspector to check both sites and find out any anomaly that may lead to a change in the electric consumption rate.

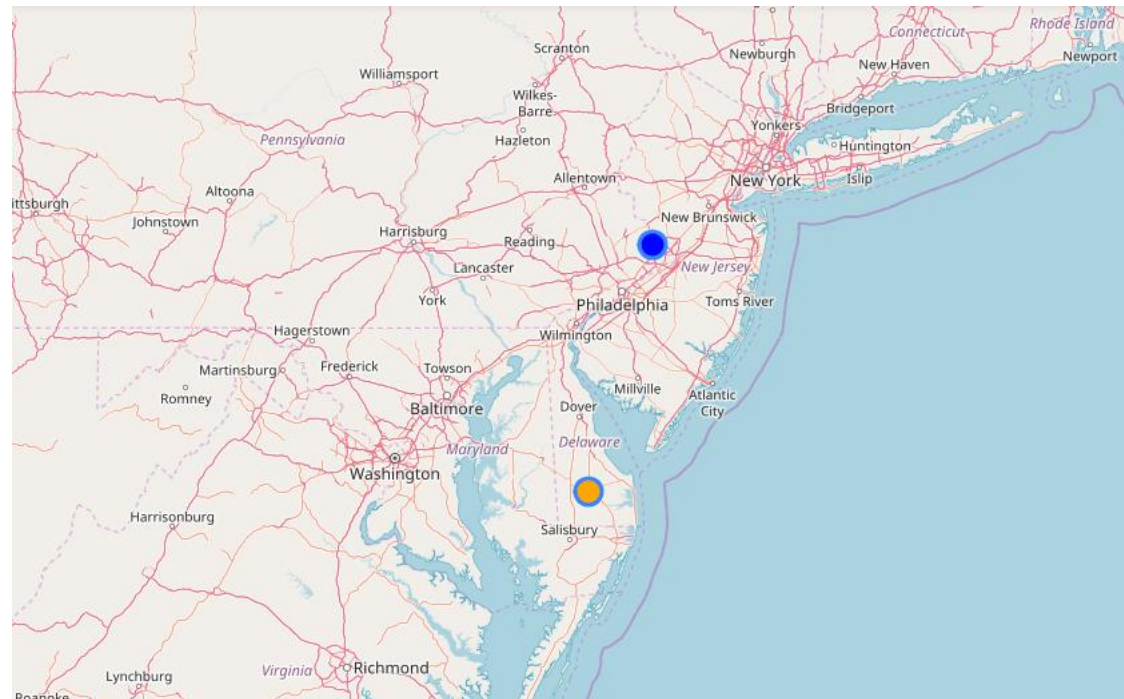


Figure 8: Locations of site 92 and site 887

CASE II. GEOGRAPHICAL COMPARISON

SITE 49 →

Pred: Primary/Secondary School.
Test: Shopping Center/Shopping Mall.
Different Industry.

Let's compare the location of this site and the rest of Primary/Secondary Schools.

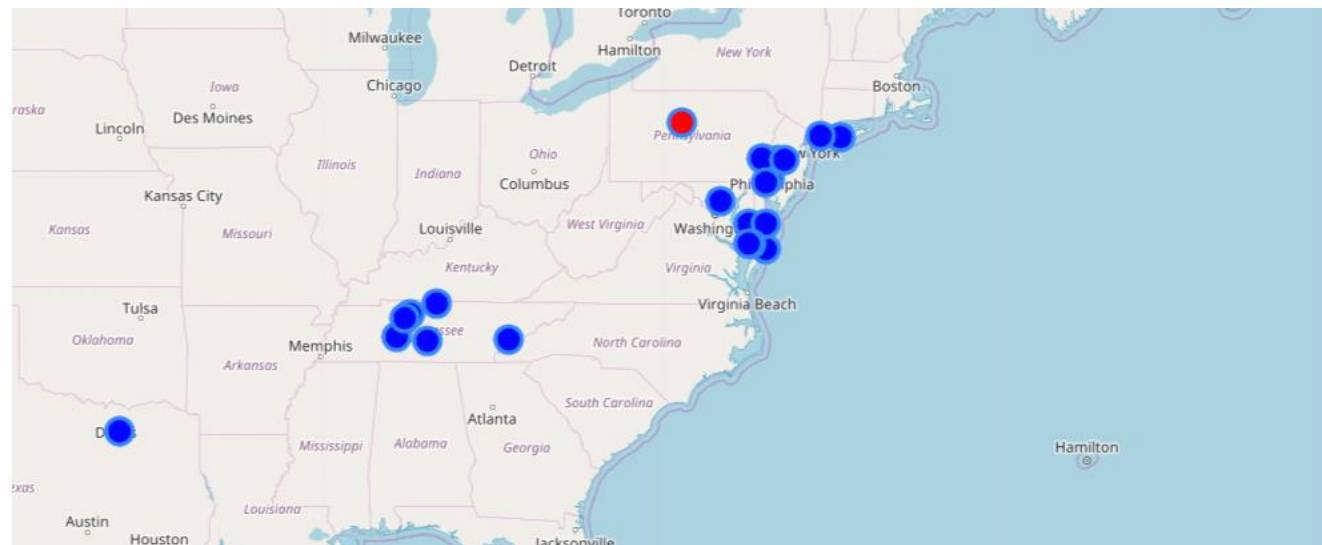


Figure 9: Location of site 49 and other Primary/Secondary Schools.

CASE II. QUICK SIMILARITY COMPARISON. FIRST MONTH

SITE 49 →

Pred: Primary/Secondary School.
Test: Shopping Center/Shopping Mall.
Different Industry.

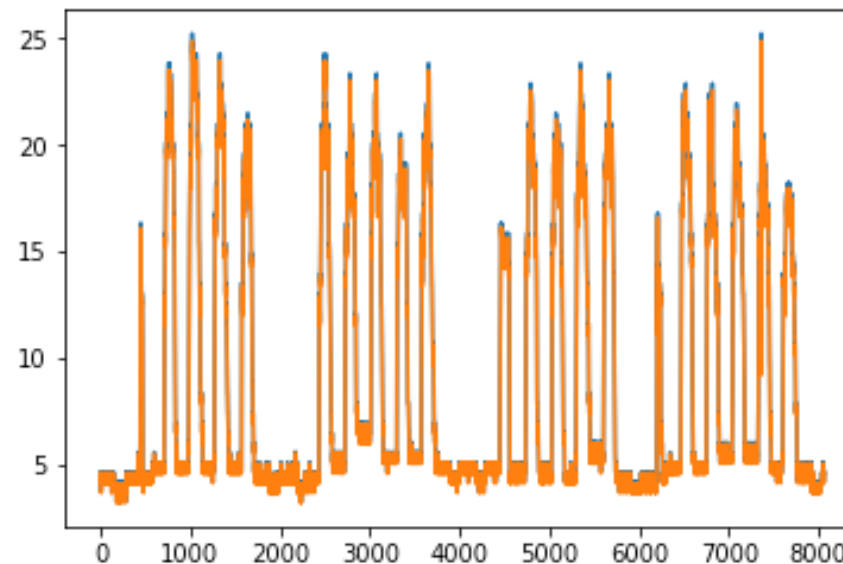


Figure 10: First month of site 49 against site 275

	SITE_ID	INDUSTRY	SUB_INDUSTRY	SQ_FT	\
19	49	Commercial Property	Shopping Center/Shopping Mall	497092	
	LAT	LNG	TIME_ZONE	TZ_OFFSET	
19	41.216176	-78.148092	America/New_York	-04:00	
	SITE_ID	INDUSTRY	SUB_INDUSTRY	SQ_FT	LAT \
49	275	Education	Primary/Secondary School	108405	40.269343
	LNG	TIME_ZONE	TZ_OFFSET		
49	-74.745143	America/New_York	-04:00		

Figure 11: Metadata of site 49 and site 275

CASE II. CONCLUSIONS

SITE 49 →

Pred: Primary/Secondary School.
Test: Shopping Center/Shopping Mall.
Different Industry.

Sites 49 and 275 have exactly the same behaviour, so an inspector should visit these sites and check both electric rates.

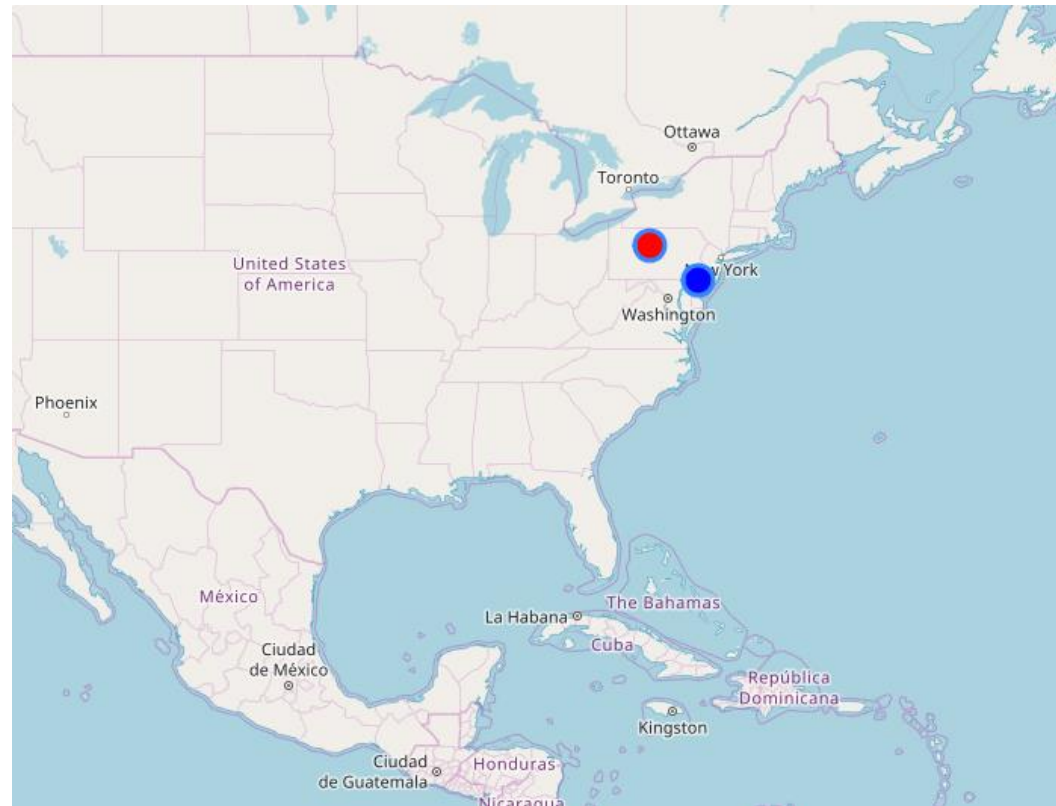


Figure 12: Location of site 49 (red) and site 275 (blue).

CASE III. GEOGRAPHICAL RESEARCH

SITE 761 →

Pred: Food Processing.
Test: Manufacturing.
Same Industry.

Let's plot the location of this site against the location of other Food Processing sites.

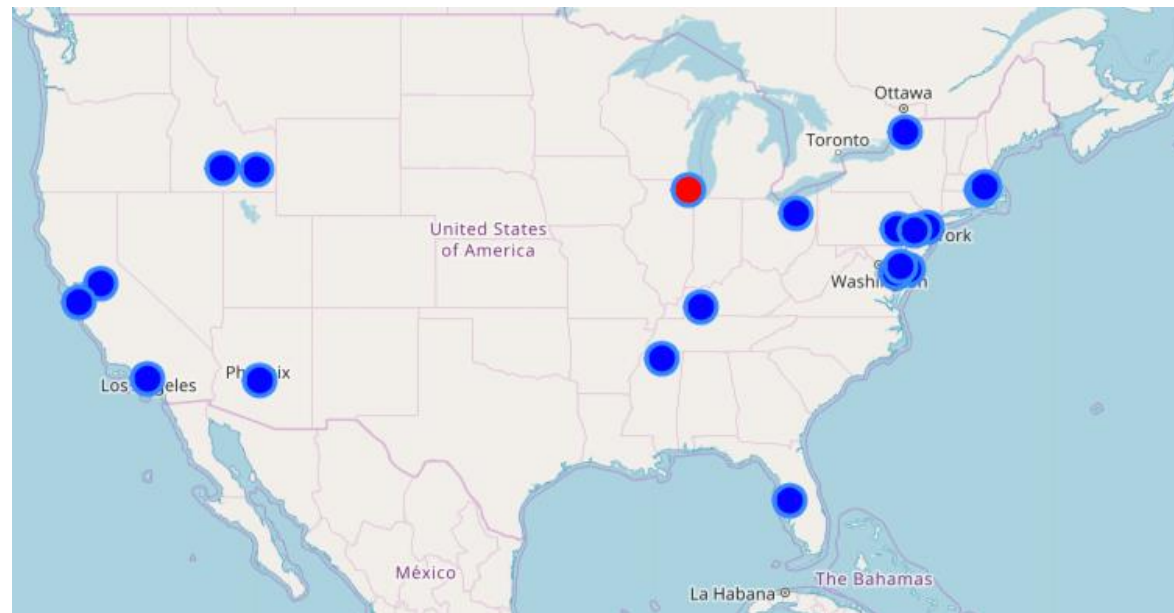


Figure 13: Location of site 761 against other Food processing sites.

CASE III. QUICK SIMILARITY COMPARISON. FIRST MONTH

SITE 761 →

Pred: Food Processing.

Test: Manufacturing.

Same Industry.

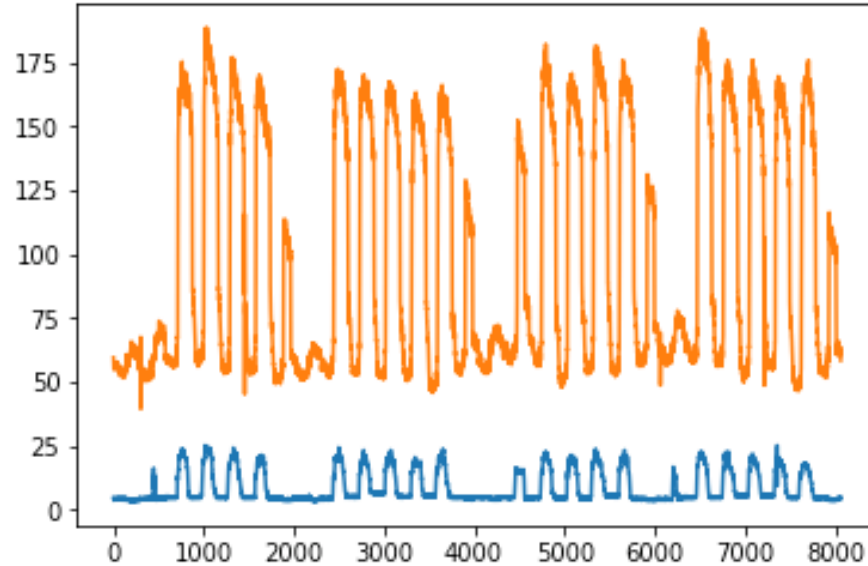


Figure 14: First month of site 761 against site 887.

	SITE_ID	INDUSTRY	SUB_INDUSTRY	SQ_FT	LAT	LNG	\
90	761	Light Industrial	Manufacturing	127899	42.107627	-87.789483	
		TIME_ZONE	TZ_OFFSET				
90	America/Chicago		-05:00				
	SITE_ID	INDUSTRY	SUB_INDUSTRY	SQ_FT	LAT	LNG	\
99	887	Light Industrial	Food Processing	381939	38.83195	-75.825022	
		TIME_ZONE	TZ_OFFSET				
99	America/New_York		-04:00				

Figure 15: Metadata of sites 761 and 887.

CASE III. CONCLUSIONS

SITE 761 →

Pred: Food Processing.

Test: Manufacturing.

Same Industry.

After comparing site 761 with the rest of manufacturing locals, we have concluded that site 832 is the most similar one. However the visualisation shows a very different behaviour.

We consider that there might not be enough Manufacturing samples to train the model, or maybe manufacturing sites have not a similar behaviour between them.

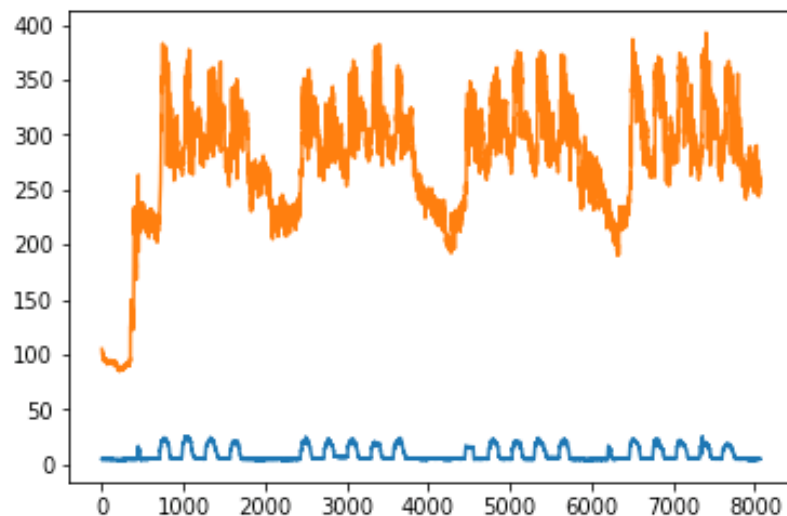


Figure 16: First month of site 761 against site 832

	SITE_ID	INDUSTRY	SUB_INDUSTY	SQ_FT	LAT	LNG	\
90	761	Light Industrial	Manufacturing	127899	42.107627	-87.789483	
		TIME_ZONE	TZ_OFFSET				
90	America/Chicago		-05:00				
	SITE_ID	INDUSTRY	SUB_INDUSTY	SQ_FT	LAT	LNG	\
98	832	Light Industrial	Manufacturing	186614	35.8111	-89.969061	
		TIME_ZONE	TZ_OFFSET				
98	America/Chicago		-05:00				

Figure 17: Metadata of site 761 and site 832

Decision: We could offer another payment rate for this site so it is not treated as a manufacturing site.