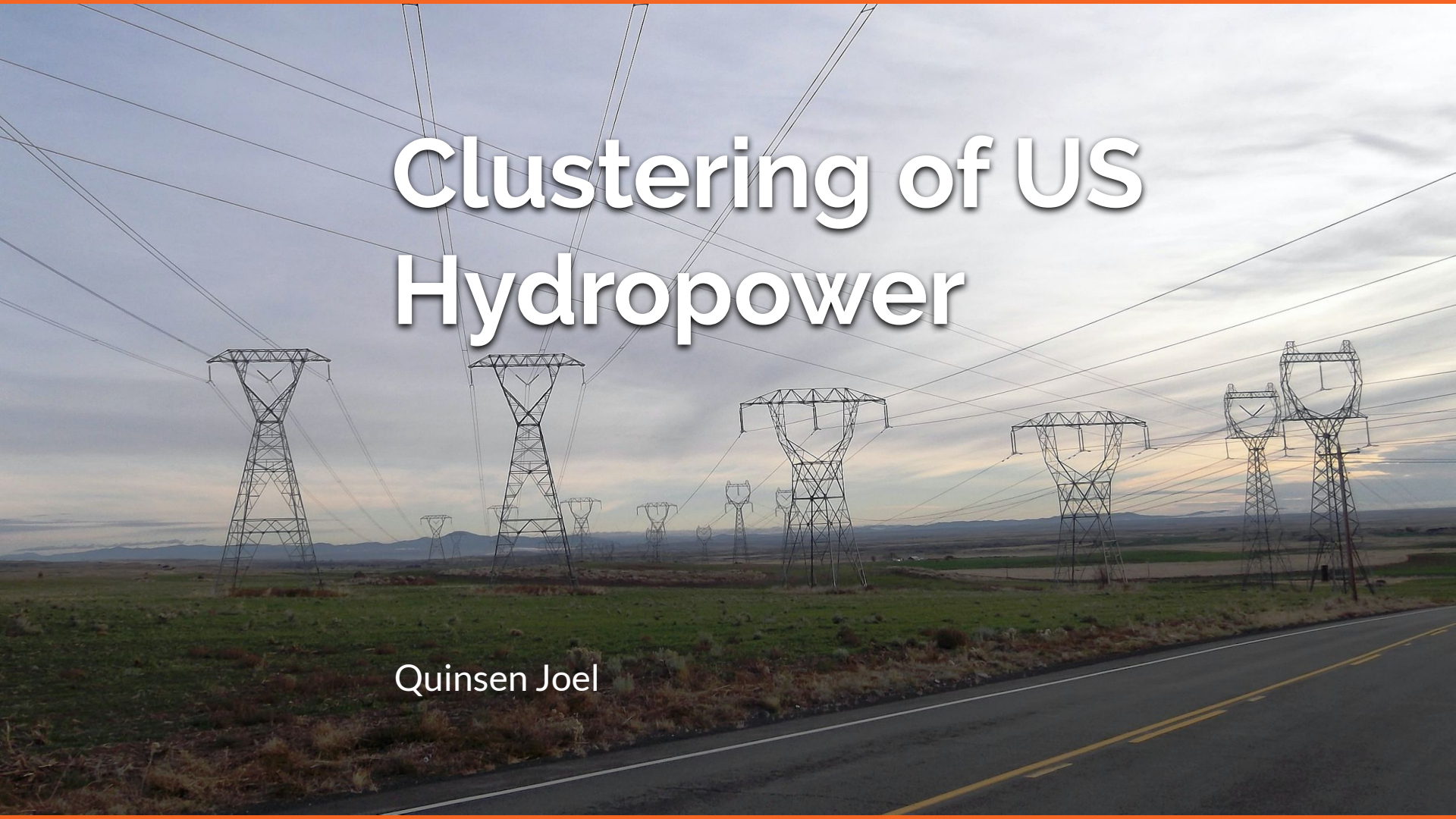


Clustering of US Hydropower

Quinsen Joel

Introduction

Hydropower is a major player in the transition to a renewable energy future

Segmentation is the division of multi-variable data into distinct clusters or classes. Segmenting existing hydropower sites can help with two broad objectives:

1. narrowing future studies to existing sites with comparable characteristics (for example, studying more about what makes a particular set of hydropower sites produce a certain level of efficiency) and
2. identifying gaps in coverage of relevant variables and help prioritize new research and development (for example, realizing we'd like to encourage more hydropower sites owned by private companies to generate high capacity).

Goal: to discover segments of hydropower sites using various common clustering algorithms including K-Means Clustering and Principal Components Analysis. Generally, segmentation will be determined by means of clustering results that exhibit the most "within-cluster" similarity and the most "between-cluster" differences.

ORNL EHA Data

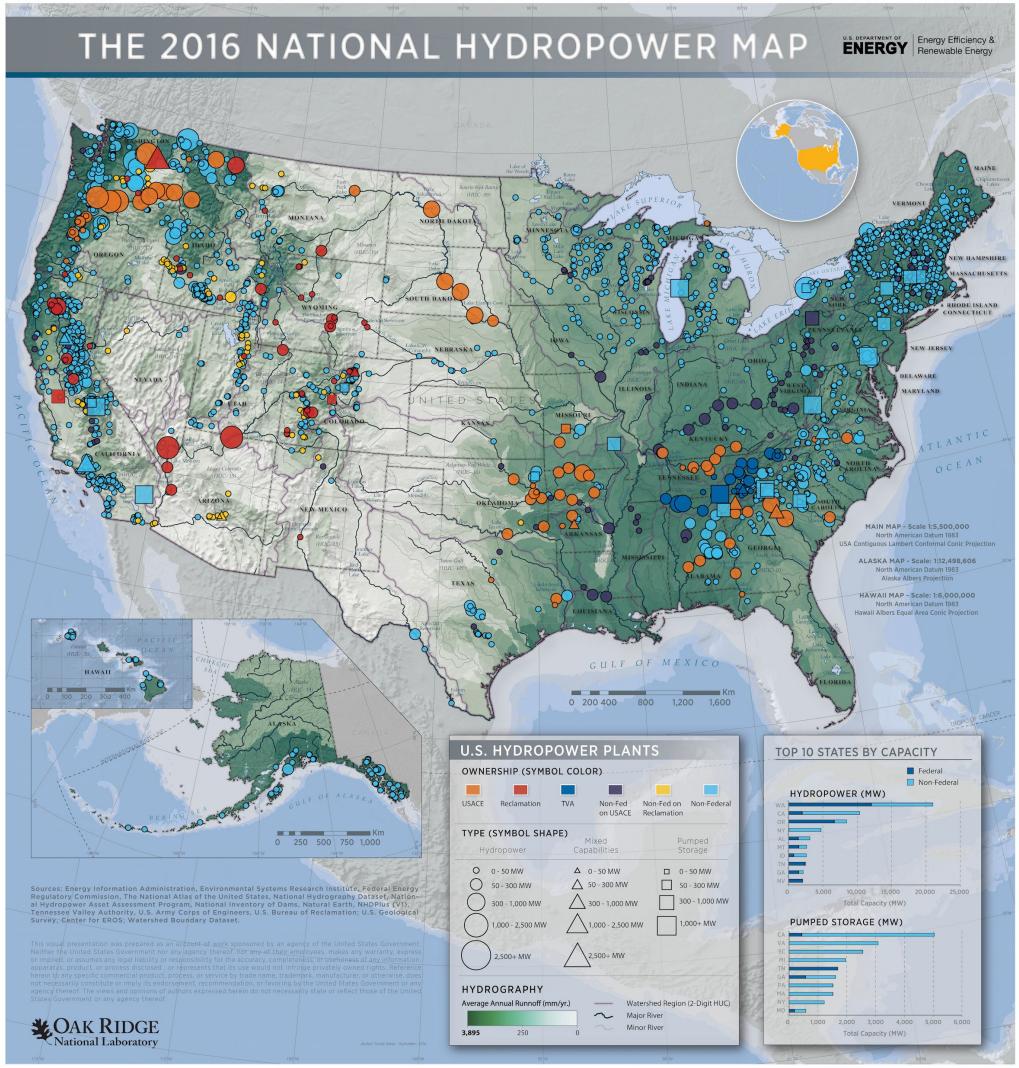
The National Hydropower Asset Assessment Program (NHAAP) funded by the Department of Energy (DOE) is the nation's leading effort to assess and expand hydropower technology.

The dataset is a product of Oak Ridge National Laboratory (ORNL) consisting of all existing hydropower assets (EHA) currently operating in the US

Features and Observations

The dataset comes with 34 features and a total of 2310 observations (hydropower assets). The features can be divided into three categories;

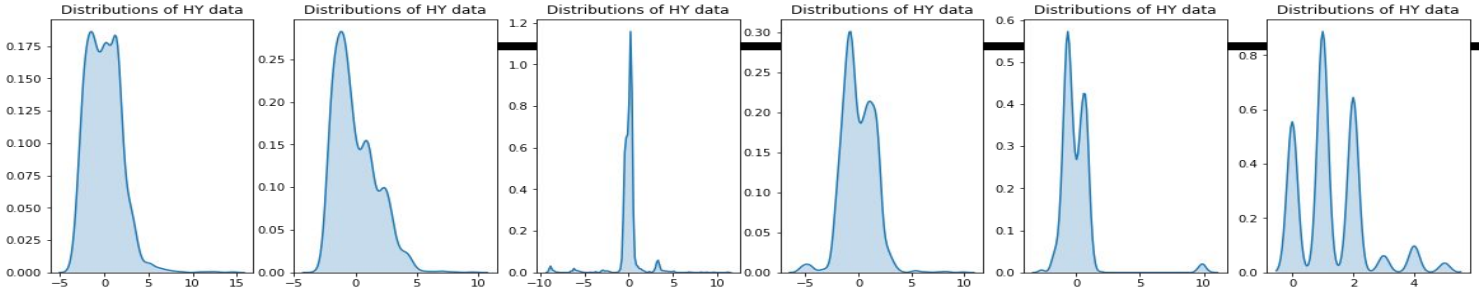
1. ***"local"*** (features that describe geography, owners, or other highly "localized" features),
2. ***"global"*** (features that describe widely comparable qualities such as energy capacity, ownership type, or sector) and
3. ***"irrelevant"*** features (features such as identifiers, license numbers,).


THE 2016 NATIONAL HYDROPOWER MAP
U.S. DEPARTMENT OF ENERGY
Energy Efficiency & Renewable Energy
MAIN MAP - Scale 1:5,500,000
North American Datum 1983
USA Contiguous Lambert Conformal Conic Projection
ALASKA MAP - Scale: 1:12,498,606
North American Datum 1983
Alaska Albers Projection
HAWAII MAP - Scale: 1:6,000,000
North American Datum 1983
Hawaii Albers Equal Area Conic Projection
U.S. HYDROPOWER PLANTS
OWNERSHIP (SYMBOL COLOR)
USACE
Reclamation
TVA
Non-Fed on USACE
Non-Fed on Reclamation
Non-Federal
TYPE (SYMBOL SHAPE)
Hydropower
Mixed Capabilities
Pumped Storage
0 - 50 MW
50 - 300 MW
300 - 1,000 MW
1,000 - 2,500 MW
2,500+ MW
1,000+ MW
HYDROGRAPHY
Average Annual Runnoff (mm/yr.)
3,895
250
0
Watershed Region (2-Digit HUC)
Major River
Minor River
TOP 10 STATES BY CAPACITY
Federal
Non-Federal
HYDROPOWER (MW)
WA
CA
OR
NY
AL
MT
ID
TN
CO
WI
0
5,000
10,000
15,000
20,000
25,000
Total Capacity (MW)
PUMPED STORAGE (MW)
CA
VA
SC
NJ
PA
GA
PA
MA
NY
MD
0
1,000
2,000
3,000
4,000
5,000
6,000
Total Capacity (MW)
Sources: Energy Information Administration, Environmental Systems Research Institute, Federal Energy Regulatory Commission, The National Atlas of the United States, National Hydrography Dataset, National Hydropower Asset Assessment Program, National Inventory of Dams, Natural Earth, NHDPlus (V1), Tennessee Valley Authority, U.S. Army Corps of Engineers, U.S. Bureau of Reclamation, U.S. Geological Survey, Center for EROS, Watershed Boundary Dataset.
This visual presentation was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.
OAK RIDGE National Laboratory
CANADA
MEXICO
UNITED STATES
PACIFIC OCEAN
ATLANTIC OCEAN
GULF OF MEXICO
HAWAII
ALASKA
0 200 400 800 1,200 1,600 Km
0 100 200 300 400 Km
0 250 500 750 1,000 Km

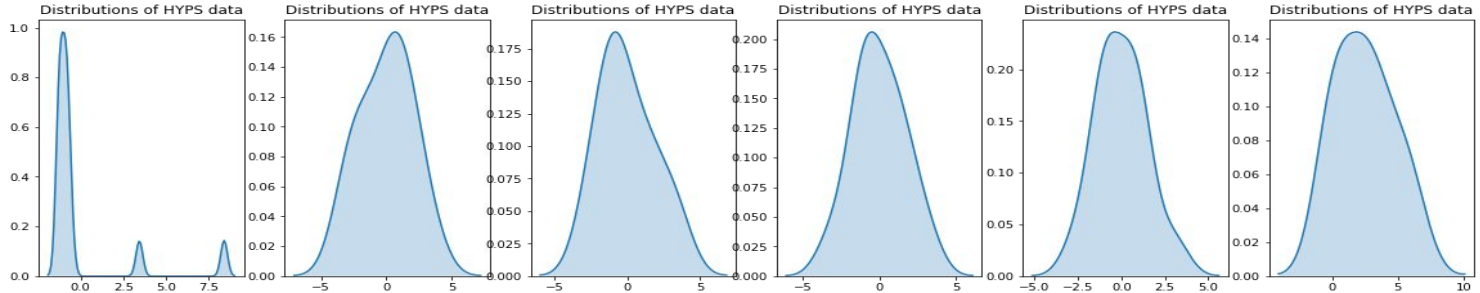
Short Name	Unit	Is Unique?	Description	Source
Sector	N/A	No	Sector Name	EHA, EIA
OwType	N/A	No	Hydropower plant facility owner type.	EHA, EIA, FERC, Reclamation, HMR, web searches
Type	N/A	No	Type of hydropower facility.	EHA, EIA, FERC, HMR
Mode	N/A	No	Mode of hydropower plant operation.	https://onlinelibrary.wiley.com/doi/abs/10.1002/rra.3004, EHA, FERC, web searches
Pm_Type	N/A	No	Permitting agency	EHA, FERC, Reclamation
Number_of_Units	number	No	Total number of generators within each plant.	EHA, EIA
CH_MW	Megawatt	No	Total capacity from hydraulic turbine-generator units within each plant. Excludes capacity from PS turbine-generator units. PS turbine-generator units are those which can consume (instead of generating) power to pump water to an upper reservoir for later use.	EHA, EIA, FERC, Reclamation, HMR, web searches
CH_MWh	Megawatt-Hour	No	Average annual net hydropower generation.	EHA, EIA, FERC, Reclamation
CH_Pf	Percent	No	(Total HY net generation [gross HY generation - plant consumption])/(HY capacity * number of hours in a year).	EHA, EIA, FERC, Reclamation
CH_OpYear	Four digit year	No	Year that the first generator came online for conventional hydropower.	EHA, EIA, HMR
PS_MW	Megawatt	No	Total capacity from PS turbine-generator units in hydropower and pumped storage plants that are licensed, exempt, or currently active but awaiting relicensing.	EHA, EIA, FERC, Reclamation, HMR, web searches
PS_MWh	Megawatt-Hour	No	Pumped storage gross generation in megawatt hours for each powerplant.	EHA, EIA, FERC, Reclamation
PS_Pf	Percent	No	(Total PS gross generation)/(HY capacity * number of hours in a year).	EHA, EIA, FERC, Reclamation
PS_OpYear	Four digit year	No	Year that the first generator came online for pumping units.	EHA, EIA, HMR
Water	N/A	No	Waterway where the hydropower facility is located	EHA, FERC, HMR, web searches
HUC	number	No	10-digit hydrologic unit code.	EHA, NHDPlus, V1
ReEDSPCA	number	No	Regional Energy Deployment System classified by National Renewable Energy Laboratory (NREL)	EHA, NREL
NERC	Acronym	No	Regional Entity Complaince Programs	EHA, EIA
BACode	Acronym	No	Balancing Authority Code	EHA, EIA
Dam_Own	N/A	No	Dam ownership where hydropower is located.	EHA, EIA, FERC, Reclamation, HMR, web searches
Trans	Acronym	No	Transmission or Distribution System Owner	EHA, EIA
County	N/A	No	U.S. county	EHA, EIA, FERC, Reclamation
State	U.S. state abbreviation	No	U.S. state abbreviation.	EHA, EIA, FERC, Reclamation
Lat	Decimal Degrees	No	Latitude.	EHA, EIA, FERC, Reclamation, NID, NHD, AK, HI, AEA
Lon	Decimal Degrees	No	Longitude.	EHA, EIA, FERC, Reclamation, NID, NHD, AK, HI, AEA
Pt_Own	N/A	No	Hydropower plant facility ownership.	EHA, EIA, FERC, Reclamation, HMR, web searches
FcIssue	mm/dd/year	No	Most recent date of FERC license issuance	EHA, FERC
FcExpire	mm/dd/year	No	Most recent date of FERC license expiration	EHA, FERC
NID_ID	N/A	No	National Inventory of Dam's identification number	NID, HMR, FERC
EIA_PtID	N/A	No	EIA plant identification code.	EHA, EIA
FcDocket	N/A	No	FERC license docket number.	EHA, FERC
EHA_PtID	N/A	Yes	Identification code of each hydropower plant feature. Provides functional linkage to other datasets within EHA.	EHA
PtName	N/A	No	Hydropower plant name.	EHA, EIA, FERC, Reclamation, HMR, web searches

"Global" features highlighted in blue, "local" features highlighted in green, and "irrelevant" features highlighted in orange.

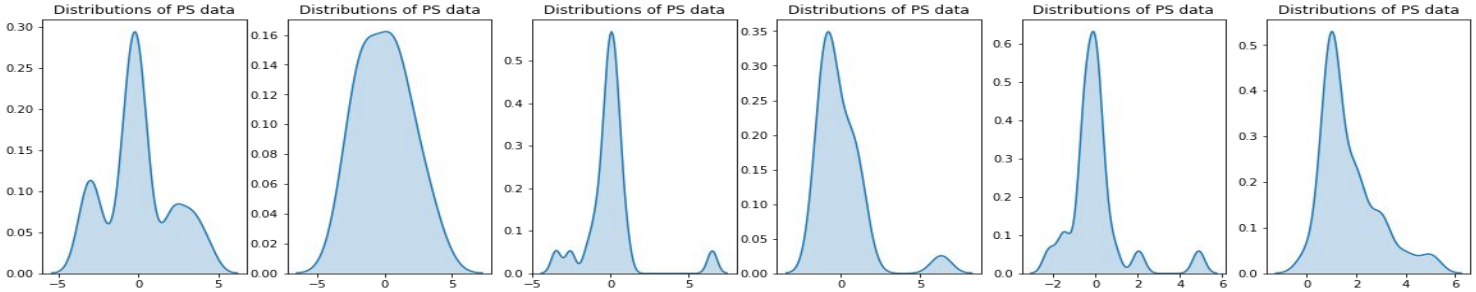
HY


Distributions of HY data
Distributions of HY data
Distributions of HY data
Distributions of HY data
Distributions of HY data
Distributions of HY data

HYPS


Distributions of HYPS data
Distributions of HYPS data
Distributions of HYPS data
Distributions of HYPS data
Distributions of HYPS data
Distributions of HYPS data

PS


Distributions of PS data
Distributions of PS data
Distributions of PS data
Distributions of PS data
Distributions of PS data
Distributions of PS data

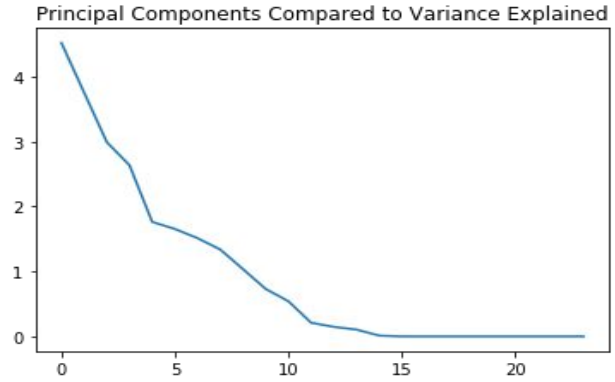
Data Preprocessing

Choose clustering features by testing clustering on various sets of numerical and categorical variables for completeness.

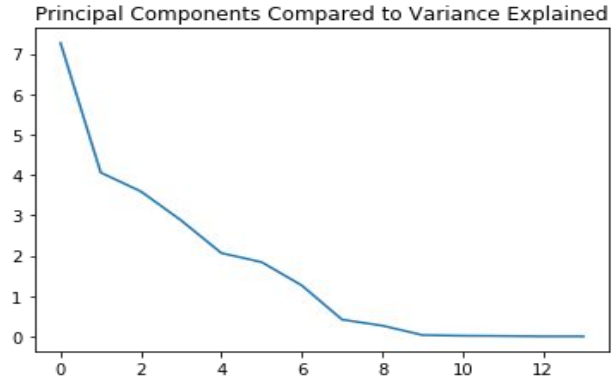
Choose categorical variables and ***"one-hot" recode*** as numeric;

Transform data, scale all features in order to more accurately represent distances in the clustering algorithms, and apply PCA to reduce the amount of features required in analysis.

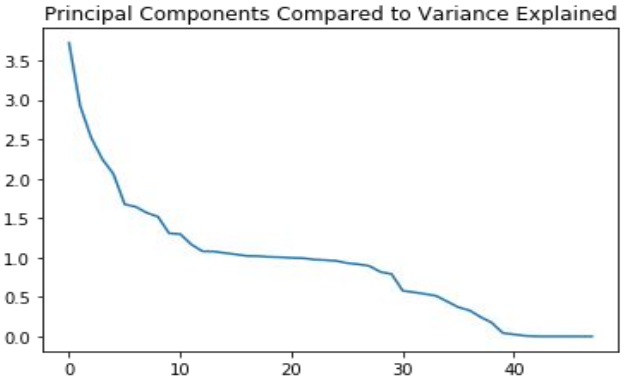
PCA Plots



HY



HYPS



PS

Methods

K means clustering “Sweeping” from k = 6:8

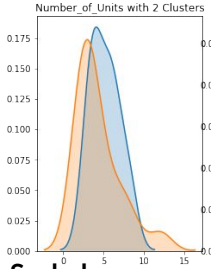
Looked at ***Pseudo F Score*** and ***Silhouette score*** to evaluate clusters

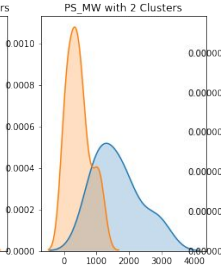
Tested clustering on three subsets of data: Hydropower plants (***HY***), Hybrid Hydropower-Pumped Storage plants (***HYPS***), and Pumped Storage plants (***PS***).

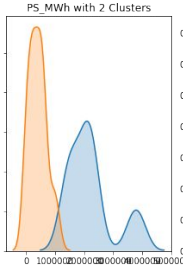
Sci-Kit Learn for nearly all of it.

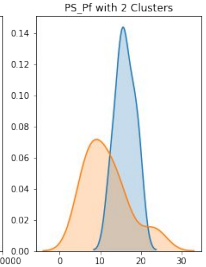
Results - HY

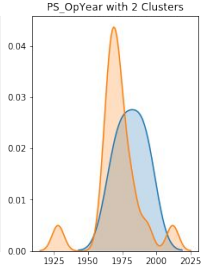
Original

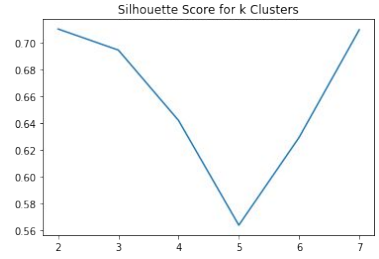


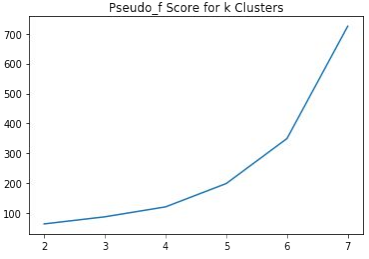




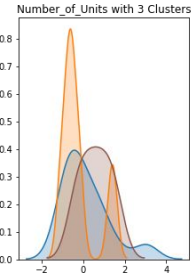


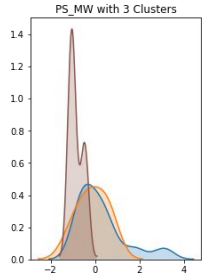


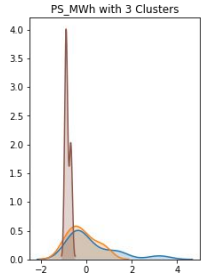


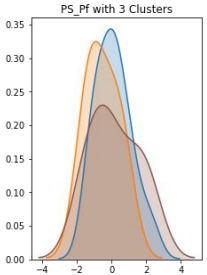


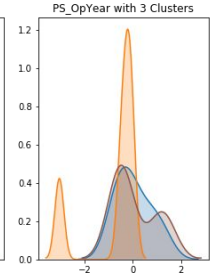
Scaled

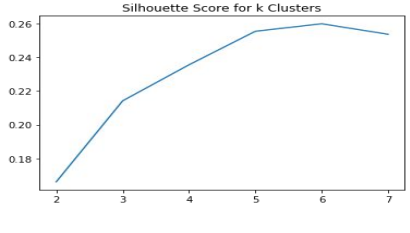


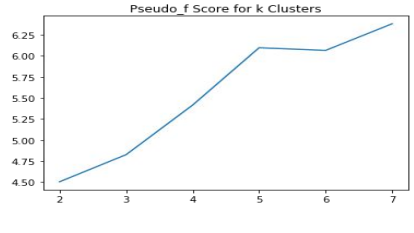




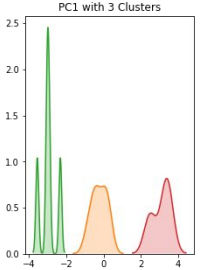


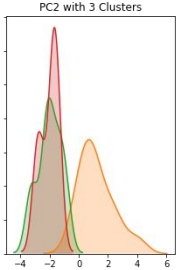


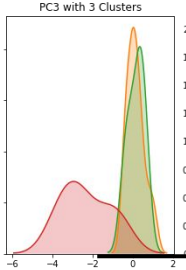


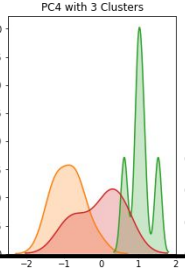


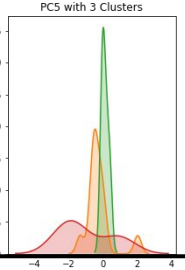
PCA

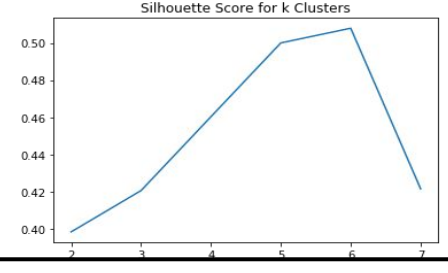


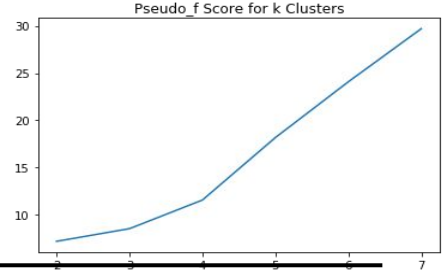






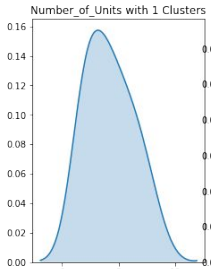


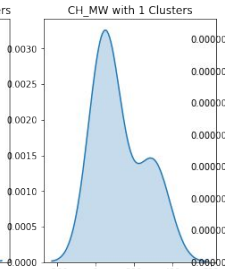


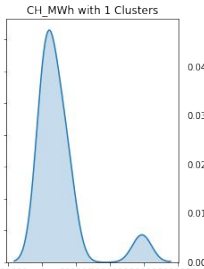


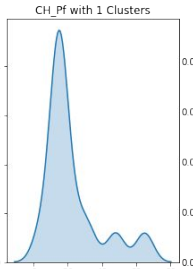
Results - HYPS

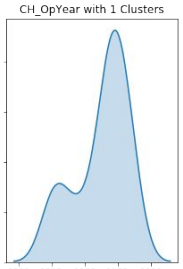
Original

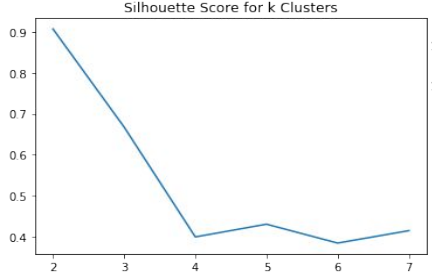


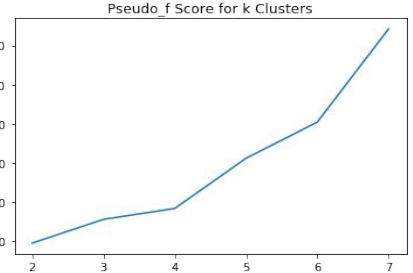




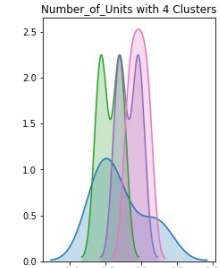


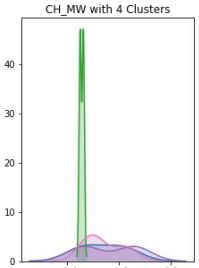


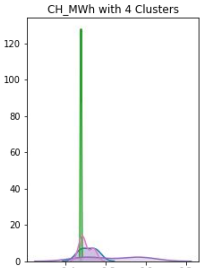


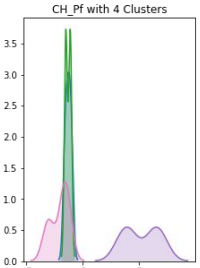


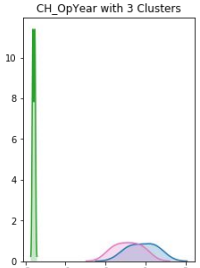
Scaled

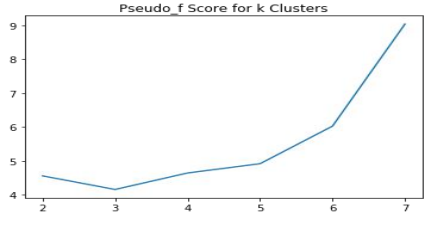


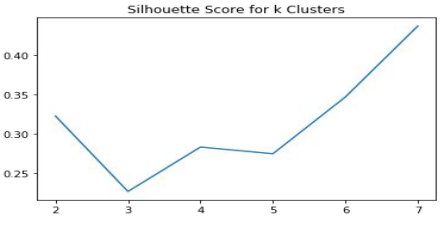




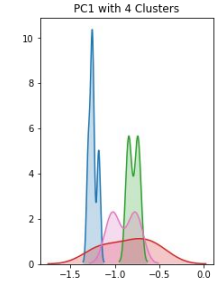


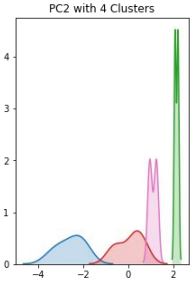


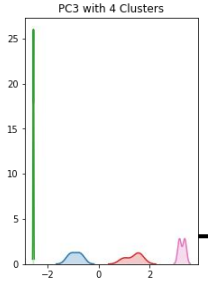


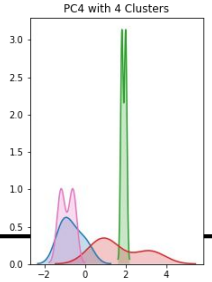


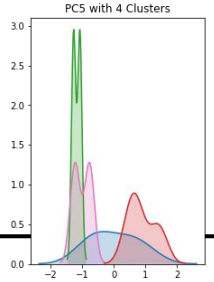
PCA

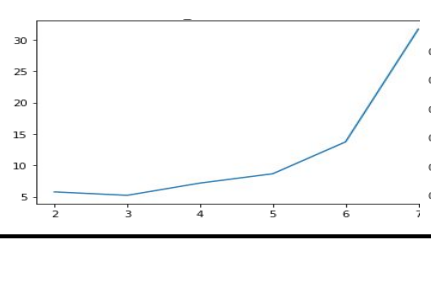


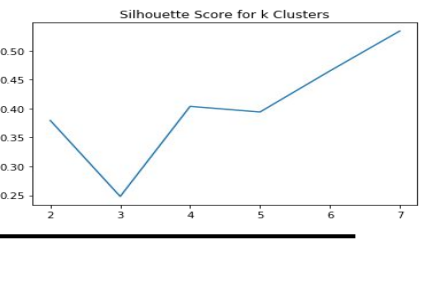






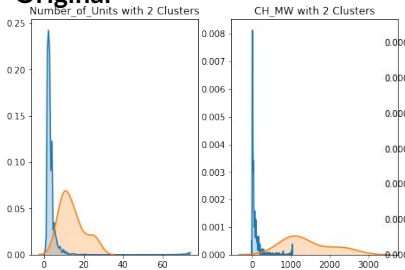


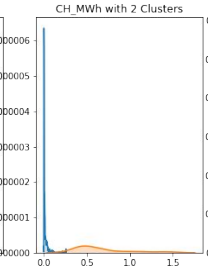


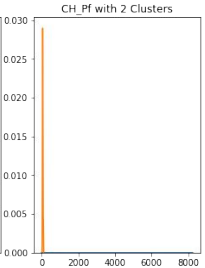


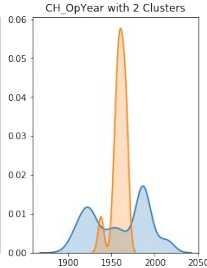
Results - PS

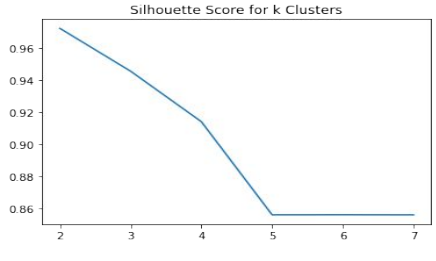
Original

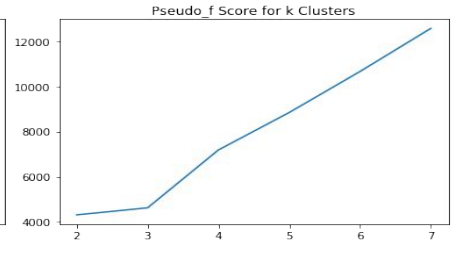




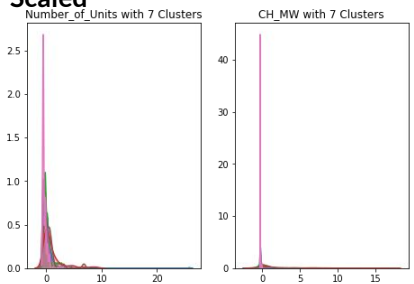


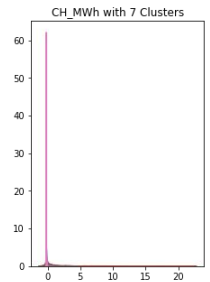


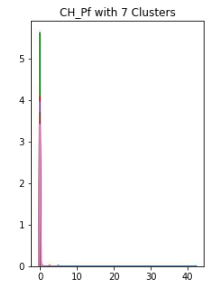


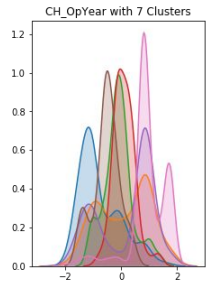


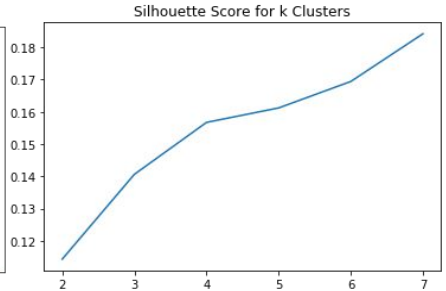
Scaled

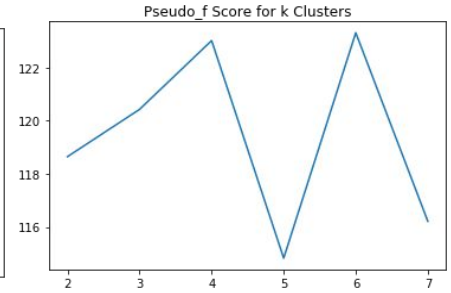




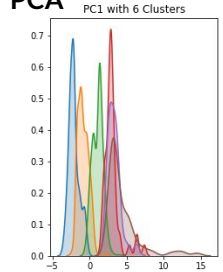


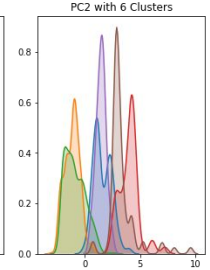


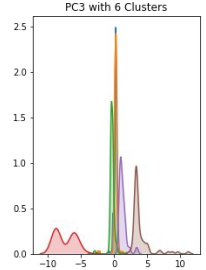


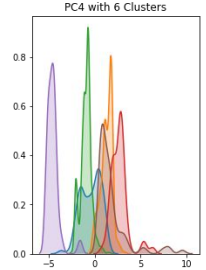


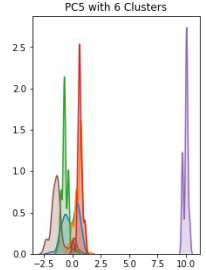
PCA

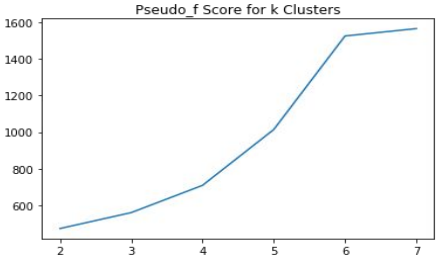


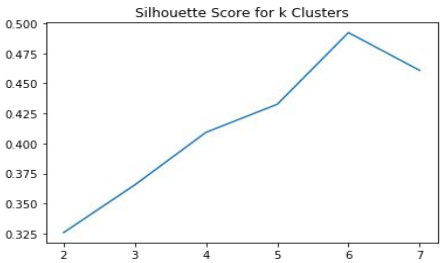












Analysis

Some of the resulting clusters are ***“noise clusters”*** - ie. clusters with just a few samples that are spurious products of the algorithm.

The remaining large-sample clusters should be studied further to ***check their validity***.

Descriptions of each of the clusters on the next page tell us the ***feature characteristics*** of each cluster.

Analysis

HY

cluster_label	index	Number_of_Units	CH_MW	CH_MWh	CH_Pf	CH_OpYear	OwType_Cooperative	OwType_Industrial	OwType_Investor-Owned Utility	ACE	Pm_Type_Unknown	Sector_Commercial CHP	Sector_Commercial Non-CHP	Sector_Electric Utility	Sector_IPP	Sector_IPP Non-CHP	Sector_Industrial	Sector_Industrial CHP	Secto
0	1593.815851	1.638695	2.305239	6.658185e+03	42.551812	1986.221445	0.018648	0.027972	0.011655	0000	0.044289	0.004662	0.023310	0.107226	0.002331	0.326340	0.000000	0.002331	
1	815.889807	2.495868	33.868124	1.081260e+05	53.331394	1943.207989	0.033058	0.001377	0.601928	0000	0.112948	0.000000	0.000000	0.938017	0.000000	0.002755	0.000000	0.000000	
2	27.705882	2.705882	145.113725	3.628409e+05	36.379187	1956.333333	0.000000	0.000000	0.000000	0000	0.000000	0.000000	0.000000	1.000000	0.000000	0.000000	0.000000	0.000000	
3	1174.886364	2.568182	10.886556	4.088029e+04	46.658840	1959.062500	0.001894	0.070076	0.028409	0000	0.032197	0.000000	0.005682	0.003788	0.000000	0.732955	0.001894	0.028409	
4	235.063291	6.341772	377.305063	1.482769e+06	36.607217	1961.215190	0.012658	0.000000	0.000000	5076	0.012658	0.000000	0.000000	1.000000	0.000000	0.000000	0.000000	0.000000	
5	324.321429	3.928571	126.721429	4.672039e+05	37.200398	1941.535714	0.000000	0.000000	0.000000	0000	0.000000	0.000000	0.000000	1.000000	0.000000	0.000000	0.000000	0.000000	

HYPS

cluster_label	index	Number_of_Units	CH_MW	CH_MWh	CH_Pf	CH_OpYear	PS_MW	PS_MWh	PS_Pf	PS_OpYear	ACE	Pm_Type_Unknown	Sector_Commercial CHP	Sector_Commercial Non-CHP	Sector_Electric Utility	Sector_IPP	Sector_IPP Non-CHP	Sector_Industrial	Sector_Industrial CHP	Sector_Industria Non-CHP
0	182.250000	4.0	154.25	2.128849e+05	15.405608	1979.0	159.250000	1.876130e+05	16.459650	1983.75	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
1	23.000000	33.0	6495.00	1.979430e+07	34.790200	1941.0	314.000000	1.254156e+06	45.595054	1973.00	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
2	599.000000	5.0	191.80	8.307790e+05	56.463624	1968.0	187.800000	5.146336e+04	2.699327	1968.00	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
3	65.500000	3.0	19.45	2.493078e+04	15.288378	1926.5	77.050000	1.196652e+05	17.839704	1971.50	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
4	322.000000	2.0	70.60	1.987529e+05	32.136957	1940.0	95.000000	9.069878e+04	10.898676	1956.00	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
5	790.666667	6.0	156.20	1.514701e+05	10.563194	1972.0	694.033333	3.809649e+05	9.566292	1972.00	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
6	17.000000	3.0	86.00	1.878778e+05	24.938652	1954.0	8.500000	3.528683e+04	47.390317	1954.00	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0

PS

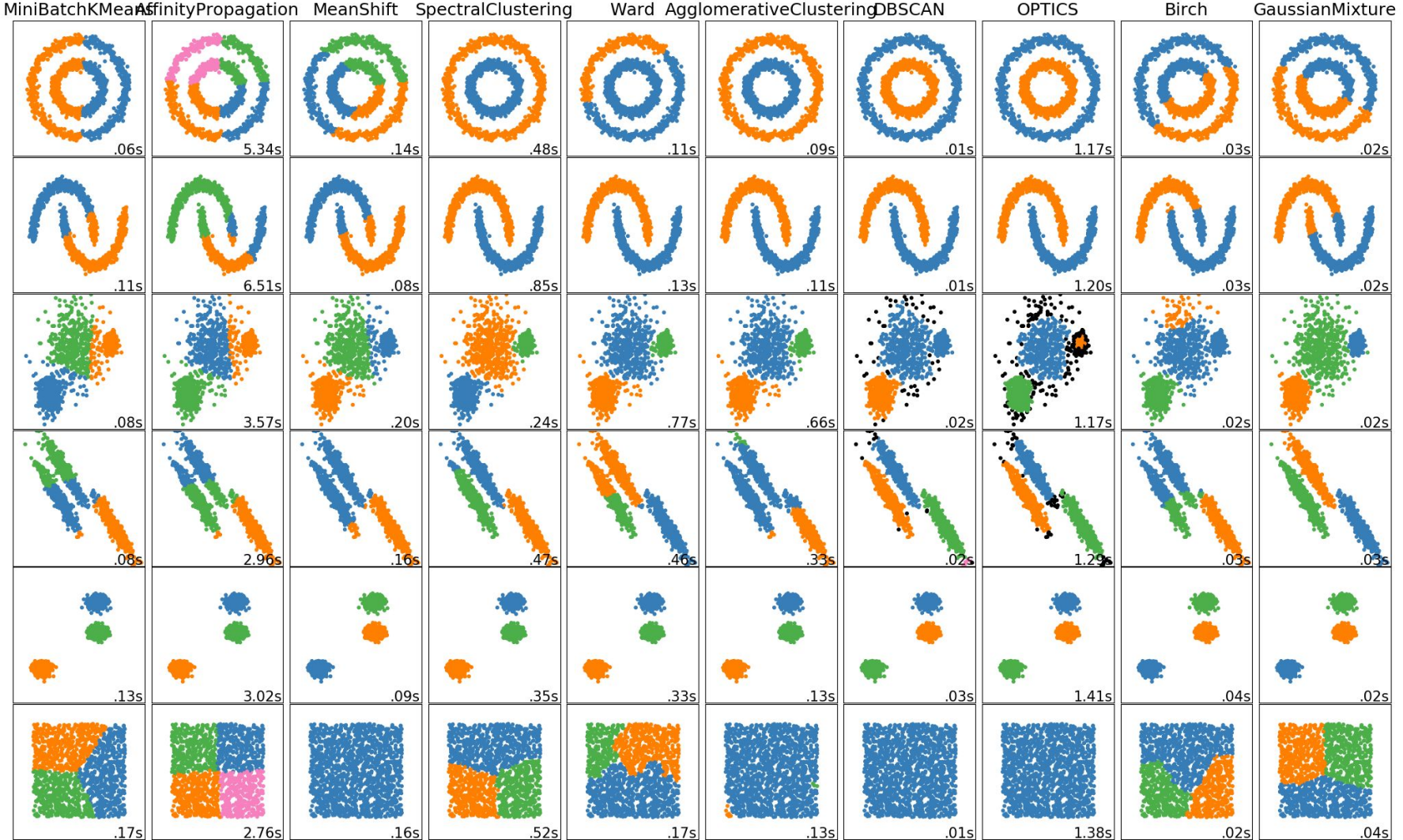
cluster_label	index	Number_of_Units	PS_MW	PS_MWh	PS_Pf	PS_OpYear	OwType_Cooperative	OwType_Industrial	OwType_Investor-Owned Utility	ACE	Pm_Type_Unknown	Sector_Commercial CHP	Sector_Commercial Non-CHP	Sector_Electric Utility	Sector_IPP	Sector_IPP Non-CHP	Sector_Industrial	Sector_Industrial CHP	Sector_Industrial Non-CHP
0	332.000000	4.000000	1713.600000	2.271913e+06	15.134849	1978.000000	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
1	896.769231	4.846154	927.615385	1.092820e+06	13.289370	1975.076923	0.076923	0.0	0.692308	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
2	791.200000	4.000000	668.000000	6.805055e+05	10.107084	1962.200000	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
3	2163.000000	2.000000	42.000000	3.487157e+04	9.478031	2012.000000	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
4	53.666667	6.000000	163.066667	1.092348e+05	13.694170	1975.666667	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
5	41.000000	2.000000	200.000000	2.938665e+05	16.773200	1981.000000	0.000000	0.0	0.000000	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0

Recommendations

Try different **unsupervised learning** algorithms.

Pre-label clusters and **test hypotheses**.

Try different clustering **feature schemes**/ pre-segment data in interesting ways.


MiniBatchKMeans
AffinityPropagation
MeanShift
SpectralClustering
Ward
AgglomerativeClustering
DBSCAN
OPTICS
Birch
GaussianMixture
.06s
5.34s
.14s
.48s
.11s
.09s
.01s
1.17s
.03s
.02s
.11s
6.51s
.08s
.85s
.13s
.11s
.01s
1.20s
.03s
.02s
.08s
3.57s
.20s
.24s
.77s
.66s
.02s
1.17s
.02s
.02s
.08s
2.96s
.16s
.47s
.46s
.33s
.02s
1.29s
.03s
.03s
.13s
3.02s
.09s
.35s
.33s
.13s
.03s
1.41s
.04s
.02s
.17s
2.76s
.16s
.52s
.17s
.13s
.01s
1.38s
.02s
.04s

References & Github

“EXISTING HYDROPOWER ASSETS.” *HydroSource*, hydrosource.ornl.gov/market-info-and-data/existing-hydropower-assets.

“Scikit-Learn.” *Scikit*, scikit-learn.org/stable/.

https://qjoel6398.github.io