NIDIS-Impact of Drought on Energy Sector

BA887: BU-NIDIS - Capstone Project Final Report

Team A1: Chinar Boolchandani, Cordell Williams, Subhiksha Sivasubramanian, Yapei Xiong, Yuxuan Mei

Table of Contents:

Introduction and Business Problem:

Drought could be one of the most serious consequences of climate change from a human and an economic perspective. Drought has plagued civilization for millennia and with the explosive population growth it is placing much more pressure on water supplies. Combined with projections that parts of the globe will become significantly drier in coming decades, drought will likely be much more of a serious issue in the future than it has in the past. Hence, understanding the impact of drought on different sectors of the economy is now more important than ever. Low water levels and high temperatures present unique challenges to the energy sector which are crucial to analyze and help energy producers and consumers take preventive measures.Our team wants to use the analytical and statistical modeling techniques to understand the impact of drought conditions on the energy production and communities who consume the electricity generated in the Northeast region of the United States.

Hence, the end goal of this project is to increase the dissemination of information related to energy production in varying drought conditions. This data gathering and modeling will provide energy grid operators with an expanded set of tools to anticipate changes in energy prices and consumption demand. Furthermore, it will provide academics with additional unsourced data to further research on the subject of the impact of dry conditions on the energy supply chain. For this project, we have aggregated data from a variety of federal and state government agencies related to drought conditions, as well as energy consumption, production and pricing. In addition, we have created a predictive model that forecasts optimal electricity prices based upon expected demand, and we have aggregated our insights into a dynamic dashboard.

Research Methodology:

To begin our research into the impact of reduced water supply on energy production, we conducted an initial knowledge gathering phase to read through established and new research on the energy market as a whole and ground our understanding in the threat of drought conditions on the energy sector. Through this initial background research, we learned that one of the main roles water plays in the power production supply chain is through its use as a method of cooling. Power plants draw large amounts of water from nearby bodies of water and the water table to boil and generate steam that propels the turbines to generate electricity. After the steam is turned back into water, pass-through cooling systems will allow the water to runoff back into the environment, while recirculating cooling systems re-use the water a number of times before release.

After concluding our initial research phase, the project broke off into two segments: subject matter expert engagement; and, data exploration and extraction. Given our lack of knowledge or acuity around this topic, we reached out to three industry experts within the Boston University (BU) ecosystem to increase our understanding of the topic on a fundamental level, and seek guidance on accessible datasets that would further our goal. This series of interviews helped us narrow down our research focus to understand the correlation between water use and availability with energy prices. In addition, they helped guide us to data that we were able to compile from a variety of federal and local government agencies and nonprofit organizations. We focused our data search on sources that showed monthly rates of energy prices and production for various power plants throughout the northeast, average water usage rates by industry for each county in the US, usage rates for natural gas and thermonuclear energy, precipitation and water table levels, and other supporting data sources.

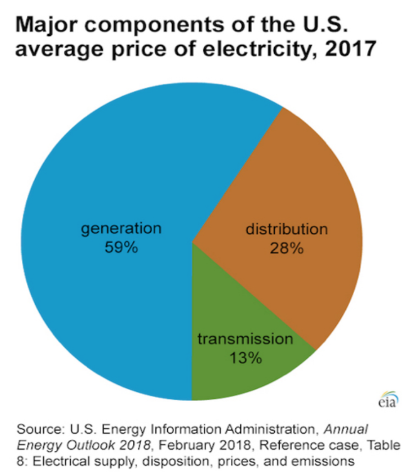
Business Impact:

The power industry is a free market economy where electricity prices are determined by a number of factors including supply and demand constraints, regulatory control, and rising costs of doing business. As the planet has begun to experience more of the extreme adverse impacts of climate change, this has begun to directly impact the energy sector. In 2015, the US Department of Energy (DoE) released a climate impact assessment that listed rising temperatures and reduced water as the top growing threat to nearly every method of energy production. For the Northeast region of the country, the DoE specifically listed concerns over the increasing temperatures leading to greater reliance on air condition penetration will heavily drive electricity consumption. This background knowledge helped us understand the future impact of drought conditions on the region, and framed our analysis of the role weather conditions, such as temperature and precipitation, have in electricity supply and demand.

The dashboard that we have created takes aggregated data, such as historic weather conditions and demographic trends, and forecasts electricity prices and consumption in an accessible and dynamic format. This will provide electricity generating utilities with a powerful tool in understanding some of the most important aspects of electricity price volatility, and help navigate decision-making when determining appropriate weather-related adjustments to future prices.

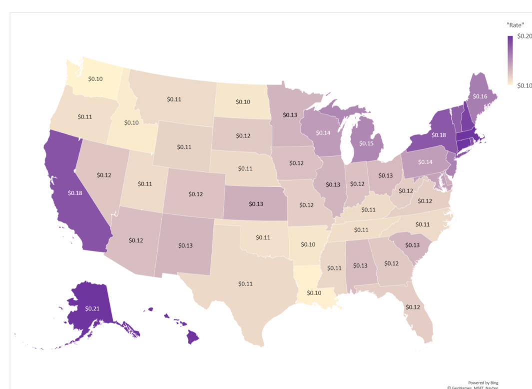
Community Impact and Benefits:

In the US, the largest component of the average price of electricity was the process required to actually generate the electricity, followed by the distribution method. As temperatures begin to rise, so does the reliance on power-generated methods of cooling. This increased demand will result in higher energy prices to offset the increase in generation. Soaring energy prices will impact every aspect of our community. Below are two of the customer segments that we believe are most vulnerable to the impacts of drought conditions on electricity production, and could be better served with the results of our analysis.



Low-income Households

Low-income households are the most vulnerable to soaring electricity prices and need to be well informed and prepared with alternative options of subsidized electricity options by the government. For low-income households, the burden of electricity prices going up can be life changing. The map on the right shows that average residential retail electricity prices are the highest in most of the north-east states(highlighted in dark purple) across the US apart from California. With the forecasting of electricity prices based on drought related factors, power suppliers and governments can take measures to subsidize and provide electricity at lower rates to these impacted households and communities and better support their functioning.



Hospitals and Community Care

Large nursing and care giving facilities' costs also go up when electricity suppliers raise their prices. A predictive analysis of electricity prices and actions to avoid inflation due to these rising prices can help these social communities plan for future expenses and mitigate unexpected impacts of drought conditions. Given the vulnerability of these populations, and their heavy reliance on energy withdrawals, it is imperative that they be provided with clear insight into the role that future rising temperatures could play in the stability of operations.

Dataset Collection and Information:

We collected data from different federal and state government agencies, including US Energy Information Administration (EIA), US Geological Survey (USGS), National Oceanic and Atmospheric Administration (NOAA) and Centers for Disease Control (CDC). In addition, we contacted four industry experts for primary professional knowledge and guidance. The industry experts include: David Jermain, Senior Fellow in BU Institute for Sustainable Energy; Michael Caramanis, BU Professor of Mechanical and Systems Engineering; Lauri Erik Pekkala, New Energy Products Growth Lead at National Grid; and Alan Pisano, BU Assistant Professor in Department of Electrical and Computer Engineering. Professional experts suggested our team research the interconnection between power grids in the Northeast with other regions. Furthermore, we researched and aggregated data on the core components of generating, distributing, and pricing energy resources for different industries, as well as the trends of energy migration. Lastly, we collected data that would give us insight on the interplay between drought and electricity pricing. Our data set includes features on the types of power plants in the Northeast, their water consumption and usage methods, electricity output and capacity, consumer segments and drought severity indices. The breakdown of our main aggregated datasets, which contain monthly data for the years 2011-2020, include:

- **Power plant level data**: Power plant ids, names, latitude, longitude of plants, net generation(MW) , capacity , water consumption, cooling type and generation type
- **State level electricity data** : Revenue from electricity, consumers, consumption in Mwh, price(Cents/kWh)
- **Drought related data** : Temperature, Precipitation , Drought Severity Indices - D0,D1,D2,D3,D4 and Palmer Severity Drought Index (PDSI)
    - *Note: Please refer to the Appendix for the definition of these indices.*

Overall, we have 51 columns and 840 observations in our aggregated datasets. Below is a snapshot of some our most relevant features in the dataset look like:

Year	Month	State	Revenue(Thousand $)	Consumption(Megawatthours)	Customers(Count)	Price(Cents/kWh)	Precipitation	Temperature	Drought_index	D0	D1	D2	D3	D4	PDSI	CAPACITY_MW	NET_GEN_MWH	CAPACITY_MWH
2011.0	1.0	CT	466863.0	2780309.0	1614219.0	16.79	3.36	23.2	4.0	0.0	0.0	0.0	0.0	0.0	-0.04	362.038889	1.578153e+06	260668.0
2011.0	2.0	CT	387663.0	2390927.0	1572803.0	16.21	3.84	27.1	4.0	0.0	0.0	0.0	0.0	0.0	0.21	362.038889	1.578153e+06	260668.0
2011.0	3.0	CT	398719.0	2436332.0	1614558.0	16.37	4.74	37.0	4.0	0.0	0.0	0.0	0.0	0.0	1.04	362.038889	1.578153e+06	260668.0
2011.0	4.0	CT	370989.0	2260627.0	1612808.0	16.41	5.52	48.9	4.0	0.0	0.0	0.0	0.0	0.0	-0.02	362.038889	1.578153e+06	260668.0
2011.0	5.0	CT	376033.0	2300770.0	1611524.0	16.34	4.87	59.7	4.0	0.0	0.0	0.0	0.0	0.0	0.01	362.038889	1.578153e+06	260668.0

Feature Engineering:

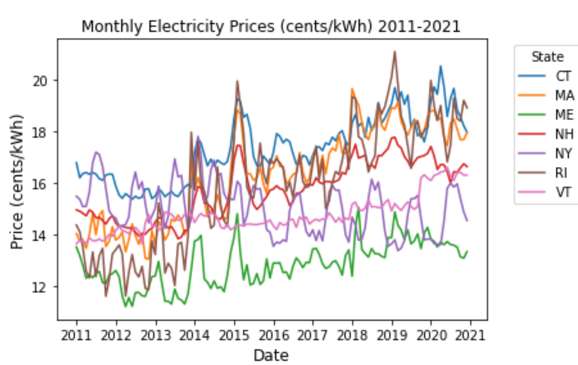
After performing basic data cleaning by removing missing/null values, we imputed data where needed and altered data types and columns to be consistent across our both datasets. We also did some feature engineering on drought and electricity related features to make sure our machine learning models are empowered enough to predict the electricity prices accurately.

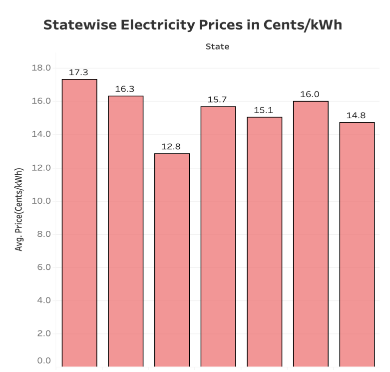
1. We calculated the temperature and precipitation lag
2. Calculated the interactions between our temperature and precipitation variables
3. Consumption per customer
4. Percentage of power generated by power plants compared to their maximum capacity
5. Percentage of electricity consumed by customers compared to net generation by power plants
6. We also calculated the top 25% , bottom 25% and top/bottom 25% percentiles for these important engineered features
7. Calculating Drought Severity and Coverage Index(DSCI) $1(D0) + 2(D1) + 3(D2) + 4(D3) + 5(D4) = DSCI$
8. Encoding categorical variables to dummy variables for modeling
9. We also normalized the data using Standard Scaler

Exploratory Data Analysis and Modeling:

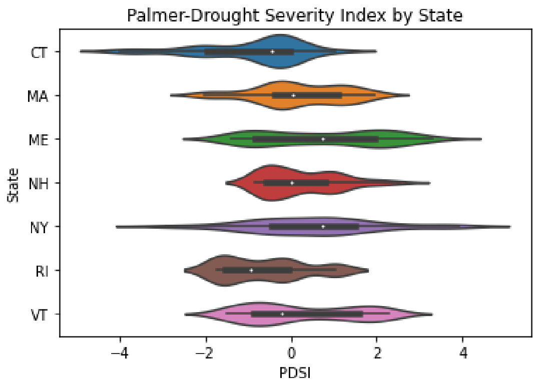
The collected data from different open source platforms were organized and merged to one efficient dataset. In order to process data, categorical columns like 'State','Temperature' and 'Precipitation' across different years were one hot encoded and missing data was imputed.

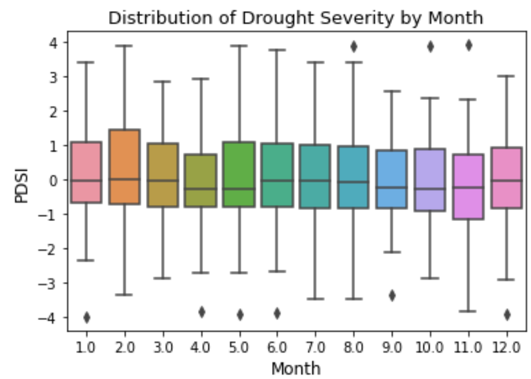
- We visualized the change in electricity prices over the years across all states as shown in the plot below on the left. Over time, electricity prices have trended upwards, although it does appear to be a very seasonal industry with increasing prices in the winter months, and decreasing prices in the summer months. Interestingly, prices per kWh overall have remained within a very tight bound for the New England region.



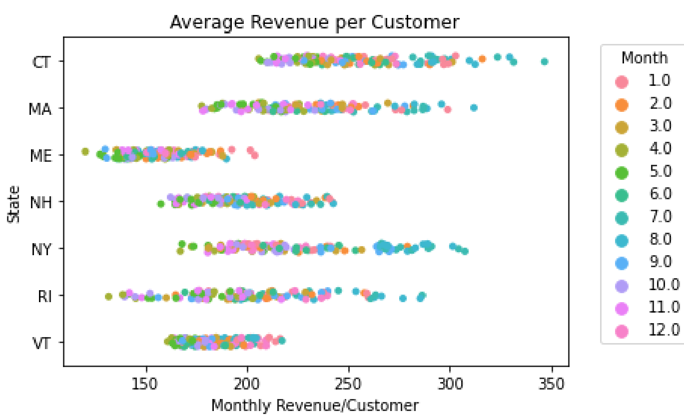


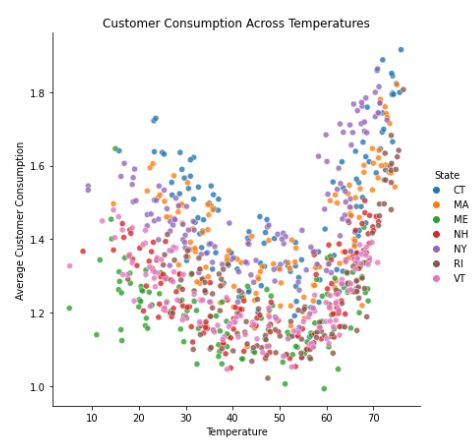
- Next, we wanted to understand the distribution of the Palmer Drought Severity Index, which is our key drought indicator to understand how different states in the Northeast region have been affected by drought on average, and to understand the relationship between drought levels and seasonality within the region. As we can see from both the violin plot and the box-and-whisker plot below, this region has largely been unaffected by drought conditions. Furthermore, seasonality seems to be very loosely related to drought conditions. This is unsurprising given the region; however, we do see some distinctions when looking at drought conditions with respect to the states. For example, Vermont, New Hampshire and Rhode Island show greater signs of dry conditions when compared to states such as Maine and New York.





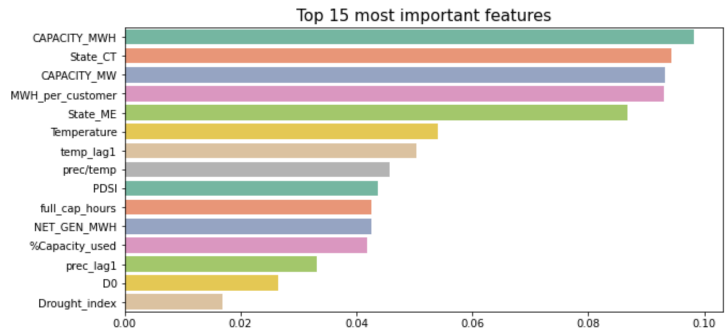
- Next, we wanted to understand some of the basic relationships around electricity demand with respect to seasonality and the environmental factors like temperature. The chart to the left below shows average monthly revenue per customer by State for each of the twelve months. As we can visually glean from the graph, revenues per customer tend to be higher in the winter and summer months, and lower in the spring and fall months. In addition, we can see that Connecticut and Massachusetts tend to have higher average revenues, while New York and Rhode Island are much more dispersed. On the chart to the right below, we show the relationship between average consumption per customer and the temperature. As we can see, energy usage is highest when temperatures are either low or high. This coincides with our findings on average revenue per customer. We see that all states seem to follow this trend, even though the dispersion of revenue per customer is quite different for the states.





Predictive Modeling and Feature Importance:

We used several predictive machine learning algorithms to fit on our final merged dataset consisting of both plant and energy data. We started with a benchmark model of Linear Regression and started adding more complexity to modeling by then using Lasso regression with penalty parameters, Decision Tree, Random Forest and Extreme Gradient Boosting models. With increasing complexity of models, we did see improvements in prediction performance , except XGBoost which did not perform as good in R square as Random Forest since the dataset is not large enough. Our best prediction model was Random Forest since it follows the bagging approach and averages results of different decision trees to reduce variance in the overall model and the worst performing model was OLS(Linear Regression). We also used Grid Search Cross Validation to tune hyperparameters like number of trees, depth of trees etc. Hyperparameter tuning helped with model performance further. Below are the metrics measured from our ML models.



Model	**MAE(Cents/kWh)**	**MSE**	**RMSE**	**Explained Variance Score**
Linear Regression-without hyperparameter tuning	1.09	1.8	1.37	48.51%
Lasso Regression- with tuning	1.34	5.73	2.39	17.82%
Random Forest- without tuning	1.2	4.59	2.14	33.44%
Random Forest- with tuning	**1.18**	**4.75**	**2.18**	**95.94%**
XGBoost- with tuning	1.29	4.31	2.07	66.30%

Conclusions:

We were able to identify, explore, clean and curate varied sources of data related to electricity supply and demand, the design and capacity of Northeast power plant facilities, environmental factors such as temperature, and indices that measure drought severity. We then used this aggregated data to build predictive and explanatory models that provide direct insight into the relationship between drought driven climate change and electricity demand management. After examining the importance of various features in our model, we've concluded that electricity pricing is moderately impacted by current temperature levels to a statistically significant degree. However, the main drivers of electricity pricing today continue to be the overall capacity levels of the power plant (likely an indicator of the size and investment in the facility), and the average MWH consumed per customer. In addition, we believe that data must be easily accessible and ingestible for non-technical audiences to have a true impact. In this accord, and given that our data has been publicly sourced, we have deployed a publicly-available, interactive visualization dashboard on Tableau to encourage the community to engage and learn how drought-related conditions are impacting the Northeast energy sector. This dataset and dashboard will equip utility owners/operators, and impacted communities, with relevant information on the correlation between drought and electricity prices, and help with data-driven decision making.

Challenges Faced:

Throughout our project, we have faced several challenges such as inadequate knowledge in the energy/drought related fields, lack of existing dataset readily available to use, and we were able to resolve some of them through research, reaching out to industrial experts, and our collaborative efforts. Below are some of those challenges and the steps we took to mitigate them:

- ➢ Data Sources were spread out and we had to dig into several websites, reach out to different industry experts which delayed the process initially but eventually we were able to gather a comprehensive list of features to solve the business problem
- ➢ Some of the data had to be collected using API , some was direct download and some was within the maps so it took a lot of time and effort gathering data and merging that into one dataset
- ➢ Limited number of observations available we had 840 observations across 10 years 2011-2020: we tried to add plant level data to the energy dataset and use both in our EDA and modeling approaches
- ➢ Since we wanted to create a dashboard to help visualize the electricity pricing fluctuations with drought factors, we used Tableau which has a very user friendly interface but it only allows one author to work on a dashboard so merging them also took considerable efforts

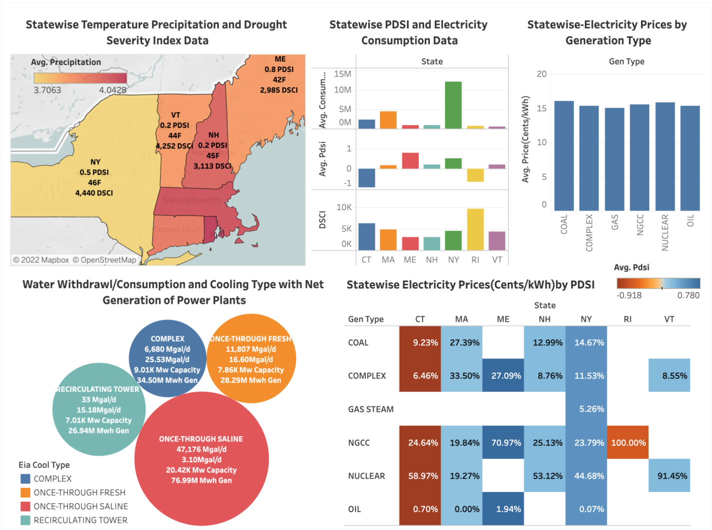
Next Steps:

It is worth mentioning that our project analysis and modeling was on a limited number of features to predict electricity prices. In reality there might be other socio-economic and regulatory factors that impact this price volatility. Additional factors to include could be average snowfall, cost of fuel, specific customer segments, transmission and distribution systems, and changes in government regulations. This project analysis can be scaled up to include above and many other additional factors and be made readily available as a front-end tool for energy utilities and communities looking for electricity price forecasting. The dataset we created, key statistics and insights from our project analysis will make it easier for businesses and individuals to understand and react accordingly in their electricity usage activity. This business analysis can be used as a baseline model to develop insights and statistics for the west and south-west states.

Github Repository and Tableau Dashboards:

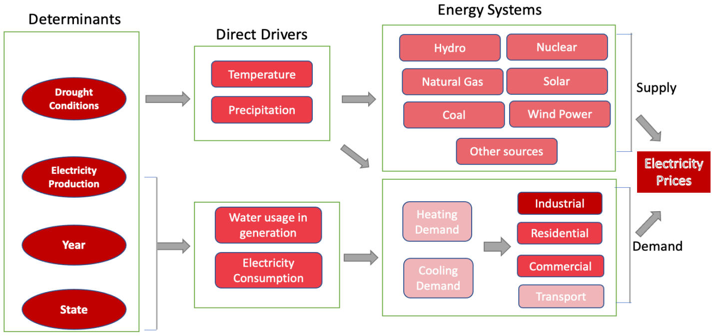
Our Tableau dashboard provides detailed visualizations of complex information on the underlying trends and relationships between drought related conditions and energy supply and demand across the Northeast region from 2011-2020. The dashboard can be accessed using the links below:

- [Github Repository for our Team]()
- [Tableau Dashboard- Drought-Energy Impact]()



Appendix:

1. Business Overview and Impact to Electricity Generating utilities:



2. Community Impact - Customer Segmentation:

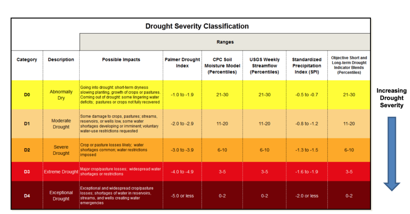
3. Data Dictionary for our final dataset:

   Electricity data that comprises of the following important variables collected on a monthly basis:
   - Revenue from Electricity
   - Electricity Consumption
   - Number of customers by state
   - Precipitation and Temperature
   - D0,D1,D2,D3,D4 - Drought severity with D0, the least and D4, the highest
   - PDSI - Palmer Drought Severity Index

   Power plant based data that comprises the following important variables:
   - Unique power plant ID
   - County and state
   - Capacity in MW
   - Net Generation of each power plant
   - Cooling method of Power plant
   - Water source for cooling the power plant
   - Maximum amount of water withdrawal
   - Temperature (2011-2021)
   - Precipitation (2011-2021)

4. Definition of the indices used:

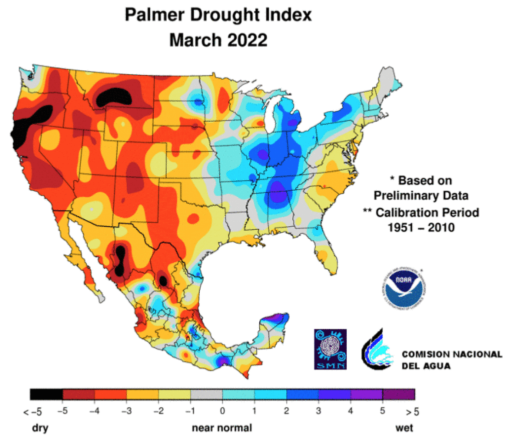
<u>Drought Severity and Coverage Indices :</u>
https://droughtmonitor.unl.edu/About/AbouttheData/DSCI.aspx
Using cumulative Drought Monitor data, we added the percentages for D0 through D4 for a given month to get the Drought Severity and Coverage Index for that month. Formula for calculating DSCI : *1(D0) + 2(D1) + 3(D2) + 4(D3) + 5(D4) =* ***DSCI***



Drought Severity Classification

		Ranges					
Category	Description	Possible Impacts	Palmer Drought Index	CPC Soil Moisture Model (Percentiles)	USGS Weekly Streamflow (Percentiles)	Standardized Precipitation Index (SPI)	Objective Short and Long-term Drought Indicator Blends (Percentiles)
D0	Abnormally Dry	Going into drought: short-term dryness slowing planting, growth of crops or pastures. Coming out of drought: some lingering water deficits; pastures or crops not fully recovered	-1.0 to -1.9	21-30	21-30	-0.5 to -0.7	21-30
D1	Moderate Drought	Some damage to crops, pastures; streams, reservoirs, or wells low, some water shortages developing or imminent; voluntary water-use restrictions requested	-2.0 to -2.9	11-20	11-20	-0.8 to -1.2	11-20
D2	Severe Drought	Crop or pasture losses likely; water shortages common; water restrictions imposed	-3.0 to -3.9	6-10	6-10	-1.3 to -1.5	6-10
D3	Extreme Drought	Major crop/pasture losses; widespread water shortages or restrictions	-4.0 to -4.9	3-5	3-5	-1.6 to -1.9	3-5
D4	Exceptional Drought	Exceptional and widespread crop/pasture losses; shortages of water in reservoirs, streams, and wells creating water emergencies	-5.0 or less	0-2	0-2	-2.0 or less	0-2

Increasing Drought Severity



PDSI(Palmer Drought Severity Index) :
https://www.droughtmanagement.info/palmer-drought-severity-index-pdsi/
Calculated using monthly temperature and precipitation data along with information on the water-holding capacity of soils. It takes into account moisture received (precipitation) as well as moisture stored in the soil, accounting for the potential loss of moisture due to temperature influences.

5. Intermediate Energy Features created by us pertaining to Consumption and Generation of electricity by power plants:

CAPACITY_MWH	temp_lag1	prec_lag1	prec/temp	MWH_per_customer	%Capacity_used	full_cap_hours
260668.0	46.852143	3.895631	0.144828	1.722386	6.054266	10.666093
260668.0	23.200000	3.360000	0.141697	1.520169	6.054266	9.172307
260668.0	27.100000	3.840000	0.128108	1.508978	6.054266	9.346494
260668.0	37.000000	4.740000	0.112883	1.401671	6.054266	8.672438
260668.0	48.900000	5.520000	0.081575	1.427698	6.054266	8.826438
260668.0	59.700000	4.870000	0.091928	1.553871	6.054266	9.601927
260668.0	66.900000	6.150000	0.029285	1.797607	6.054266	11.098980
260668.0	74.100000	2.170000	0.163277	1.859490	6.054266	11.490436
260668.0	70.800000	11.560000	0.117169	1.563704	6.054266	9.662529
260668.0	66.400000	7.780000	0.104869	1.391235	6.054266	8.596049
260668.0	53.400000	5.600000	0.077778	1.335061	6.054266	8.259802
260668.0	45.900000	3.570000	0.122764	1.478860	6.054266	9.154495
260668.0	36.900000	4.530000	0.090221	1.622493	6.054266	10.037765
260668.0	31.700000	2.860000	0.031143	1.497939	6.054266	9.266999
260668.0	35.000000	1.090000	0.028384	1.461525	6.054266	9.036740
260668.0	45.800000	1.300000	0.060362	1.320896	6.054266	8.164976
260668.0	49.700000	3.000000	0.063961	1.356031	6.054266	8.378359
260668.0	61.600000	3.940000	0.064759	1.489152	6.054266	9.195812
260668.0	66.400000	4.300000	0.051012	1.844661	6.054266	11.386166

6. References:

- https://www.drought.gov/sectors/energy#impacts
- https://www.drought.gov/sectors/energy#map
- http://nedews.nrcc.cornell.edu/
- https://www.usgs.gov/mission-areas/water-resources/science/water-use-united-states
- https://www.eia.gov/electricity/data.php
- https://www.energy.gov/sites/default/files/Impacts%20of%20Long-term%20Drought%20on%20Power%20Systems%20in%20the%20US%20Southwest%20%E2%80%93%20July%202012.pdf
- https://www.altexsoft.com/blog/business/price-forecasting-machine-learning-based-approaches-applied-to-electricity-flights-hotels-real-estate-and-stock-pricing/
- https://data.cdc.gov/Environmental-Health-Toxicology/Palmer-Drought-Severity-Index-1895-2016/en5r-5ds4