Optimizing Electricity Rates

Research and analysis on the **tariff structures** of the electricity market to promote **efficiency, equity, and sustainability**

Headed by: **Professor Patricia Hidalgo-Gonzalez**
Postdoctoral Advisor: **Rajni Bansal, Ph.D.**

About

Katelyn Wong

4th Year Mathematics & Economics Student, UC San Diego
Incoming M.S. in Data Science Student, Halicioglu Data Science Institute

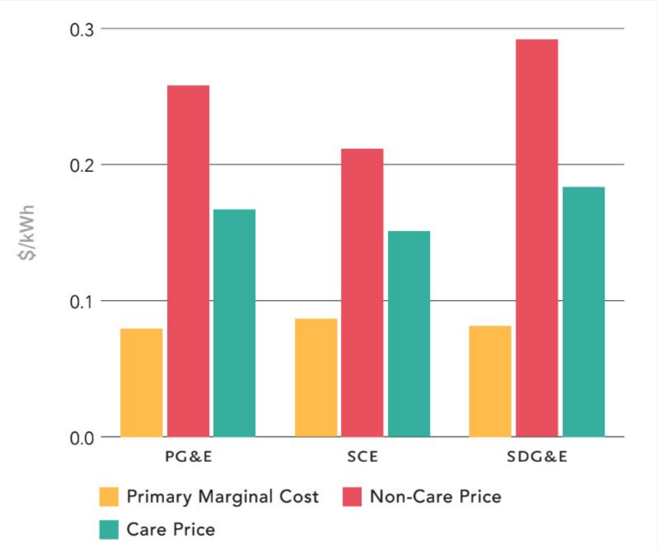
- Student researcher at the **Renewable Energy & Advanced Mathematics Lab** (Fall 2024 - present)
- Instructional assistant for the **UCSD Math Department** (2024 - present)
- Economics student researcher at **Hitotsubashi University**, Japan (2024)
- Project management and finance intern at the **San Diego Association of Governments** (2022-2023)

The Electricity Pricing Problem

Consumers often pay significantly more than the marginal cost of electricity

→ market inefficiency and inequities
→ inefficient energy use
→ slow renewable energy adoption
→ economic burden on low-income households

Marginal Cost v. Retail Prices for Electricity ($/kWh) in 2019



Hinchberger, Andrew J., et al. The Efficiency of Dynamic Electricity Prices. 4 Sept. 2024.

Challenges in Pricing

and the current research gap

Complexities from

- wide-reaching impact on communities
- diverse utility providers
- income disparities
- regional weather variation
- volatile electricity costs
- previously overlooked fixed costs and long-term pricing trends

Research Purpose

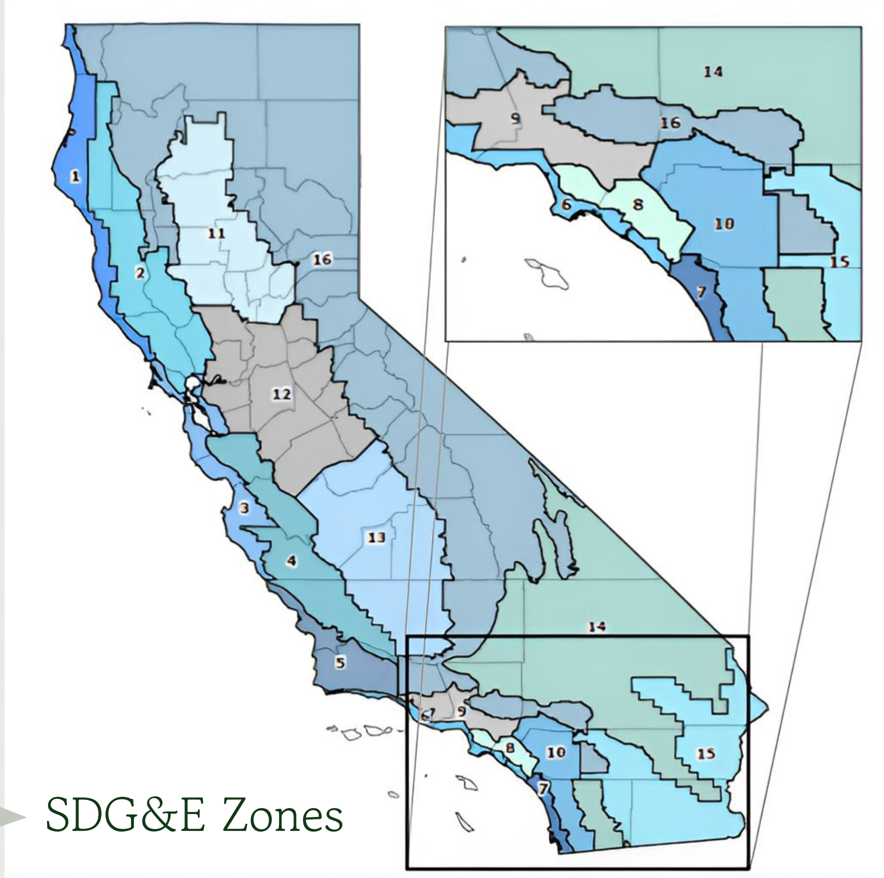
Analyzing electricity markets to optimize pricing strategies that

1. minimize the gap between price and marginal costs
2. promote the feasible and affordable adoption of renewable energy
3. address fixed costs and long-term pricing trends

Outline

Preliminary data analysis	Conducting **EDA for Energy + Environmental Economics (E3) Avoided Cost Data**
Benchmark model building	Creating baseline models for **Energy Value ($/MWh) x Hour** data to compare predictability and robustness extremes
Hybrid model building	Developing hybrid regression **models that integrate multiple baseline approaches**
Multivariate model building	Including other electricity value determinants to model fixed, marginal, and total electricity value
Future research direction	Create **multivariable regression models** for fixed, variable, and total electricity value Extrapolating created models for **real world application to policy and other grid regions**

Current Scope of Research



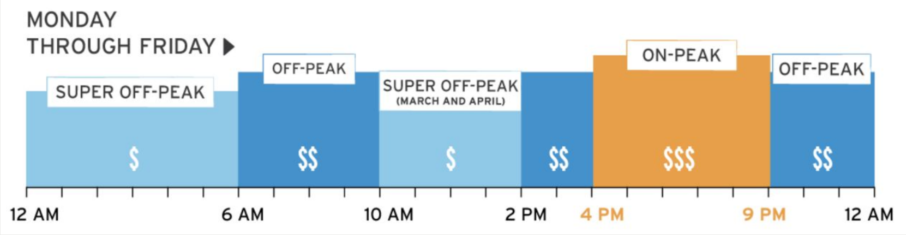
[1]California Masonry Contractors Association. "Concrete Masonry Construction." CMACN, 2021. [Online Image]. Available at: https://cmacn.org/u/2021/06/image042.png.

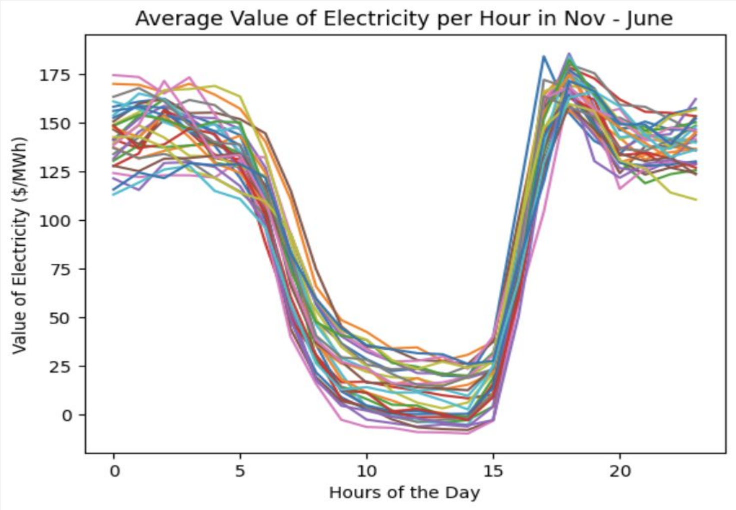
Preliminary Data Analysis

and a closer look at current electric utility pricing models

- Independent variable: **hour of the day**
- Dependent variable: **Average market value of electricity in $/MWh** (from E3 Avoided Cost Calculator data):

Aligns with how **energy companies represent time-of-use electricity pricing.** For example:



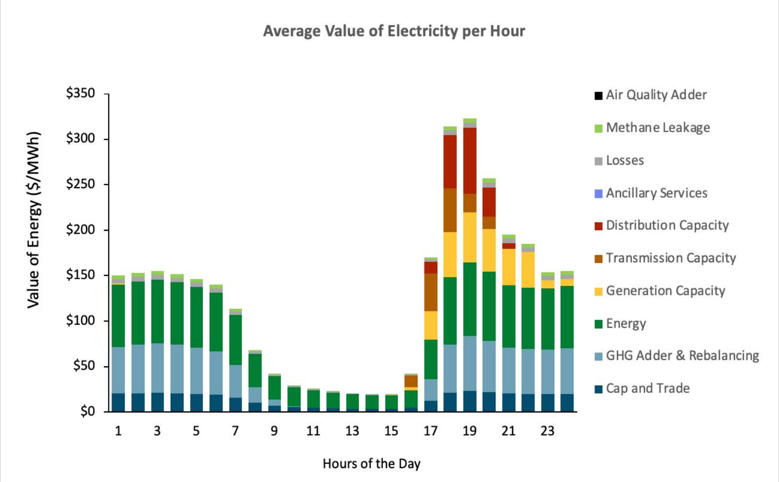


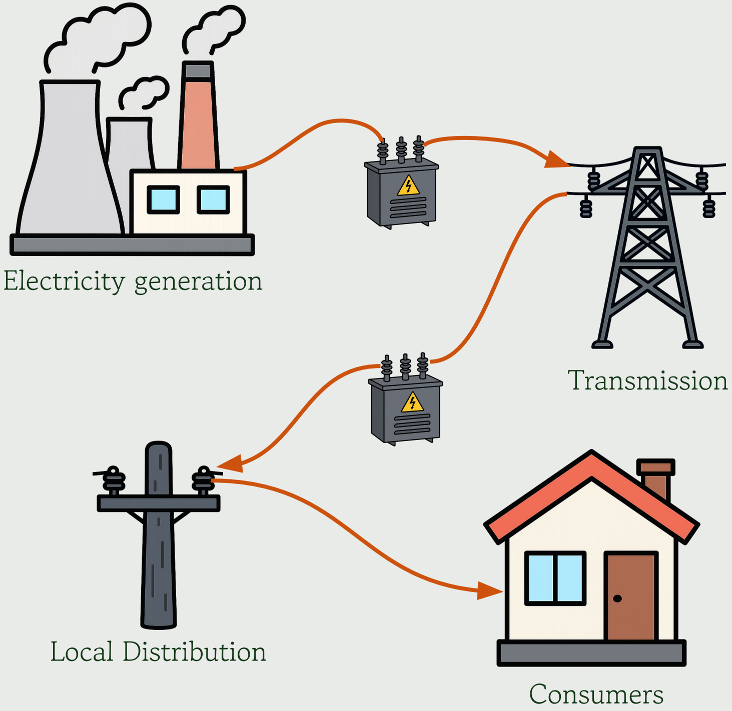
[1]San Diego Gas & Electric. "Time-of-Use (TOU) Pricing Plans." SDG&E, [Mar. 4, 2025]. [https://www.sdge.com/sites/default/files/NEW_TOUDR1_LargeFont.png].

Preliminary Data Analysis

and a closer look at current electric utility pricing models

- Independent variable: **hour of the day**
- Dependent variable: **Average market value of electricity in $/MWh** (from E3 Avoided Cost Calculator data):





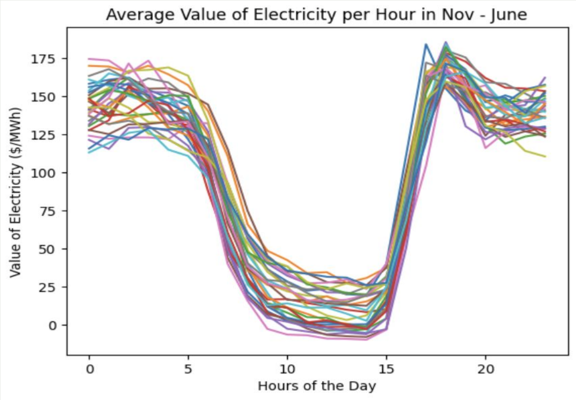
E3 (Energy and Environmental Economics, Inc.). (2024). Average Value of Electricity per Hour [Graph]. In 2024 Avoided Cost Calculator (ACC): Electric Model Version V1b. Retrieved from https://www.ethree.com/public_proceedings/energy-efficiency-calculator/

Preliminary Data Analysis

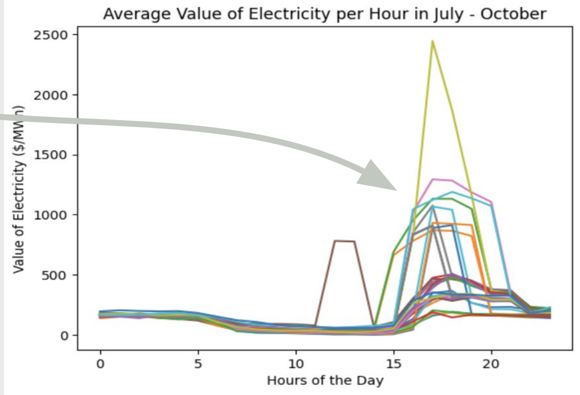
and a closer look at current electric utility pricing models

- Independent variable: **hour of the day**
- Dependent variable: **Average market value of electricity in $/MWh** (from E3 Avoided Cost Calculator data):

- Outliers in July-October data likely from
    - Peak AC and electricity demand
    - Grid congestion
    - Activation of expensive peaker plants
    - Reduction in solar generation
    - Extreme weather events, heat, and wildfires



Aggregated Non Outlier Data

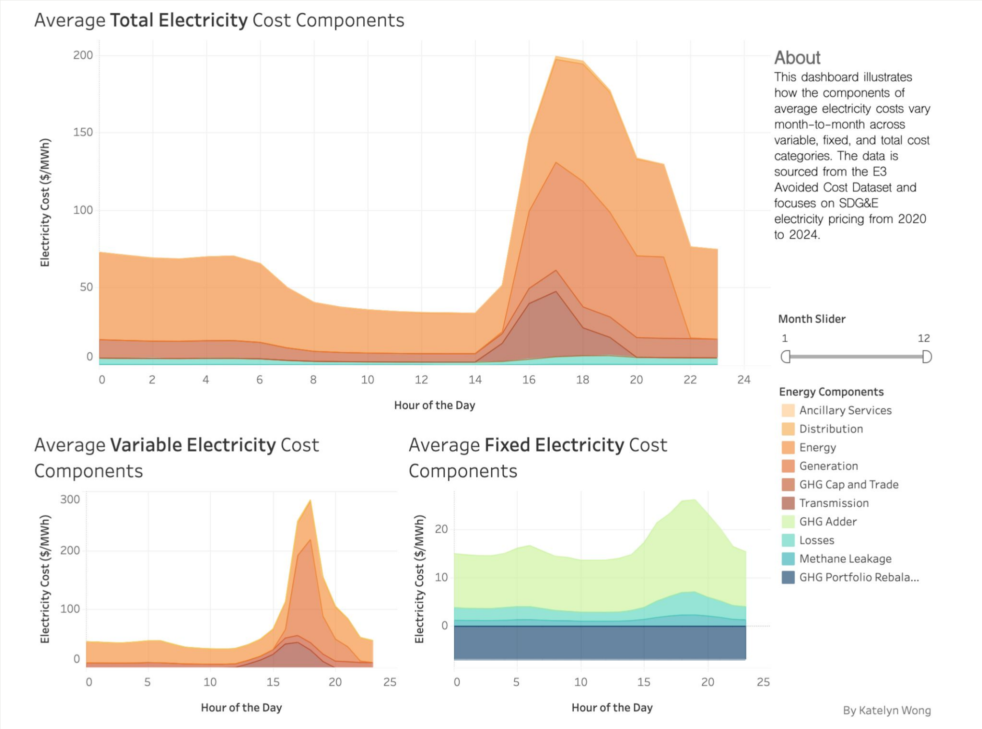


Aggregated Outlier Data

E3 (Energy and Environmental Economics, Inc.). (2024). 2024 Avoided Cost Calculator (ACC): Electric Model Version V1b. Retrieved from https://www.ethree.com/public_proceedings/energy-efficiency-calculator/

Breakdown of **Marginal and Fixed Components** of Electricity

(click to access the interactive dashboard below)



Benchmark Model Building

Developing baseline models to serve as benchmarks for evaluating more complex hybrid regression approaches.

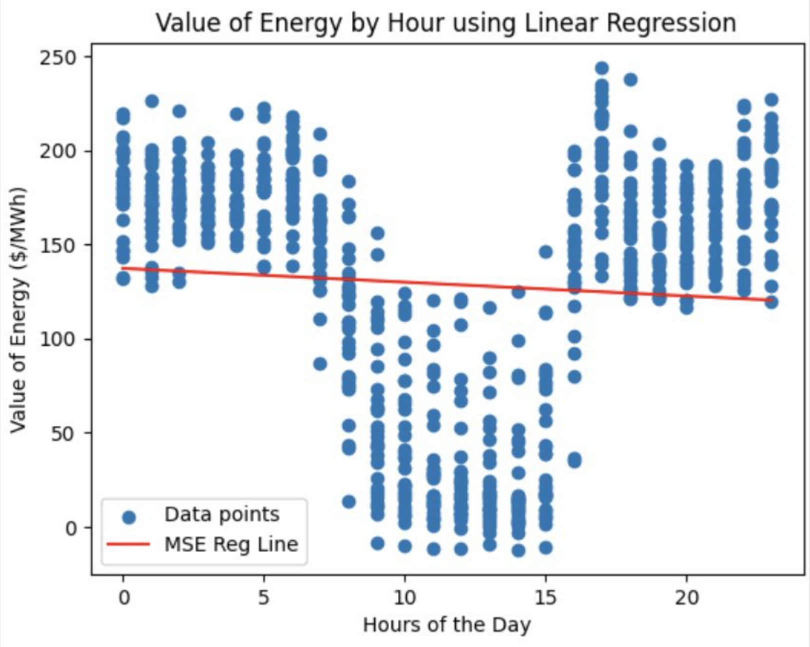
Naive Baseline Model

Initial regression model building for **Value of Electricity ($/MWh) x Hour** data

Linear regression model:

- Naive model provides a **simple baseline for error assessment.**
- Reflects some common constant linear pricing structures used by utility companies

January:



Evaluating model fit using root mean squared error (RMSE)**: ~67.09** ($/MWh)

✓: Predictable and generalizable to other months and years

✗ : Severe underfitting, **misses patterns in data**

Complex Benchmark Model 1

Initial regression model building for **Value of Electricity ($/MWh) x Hour** data

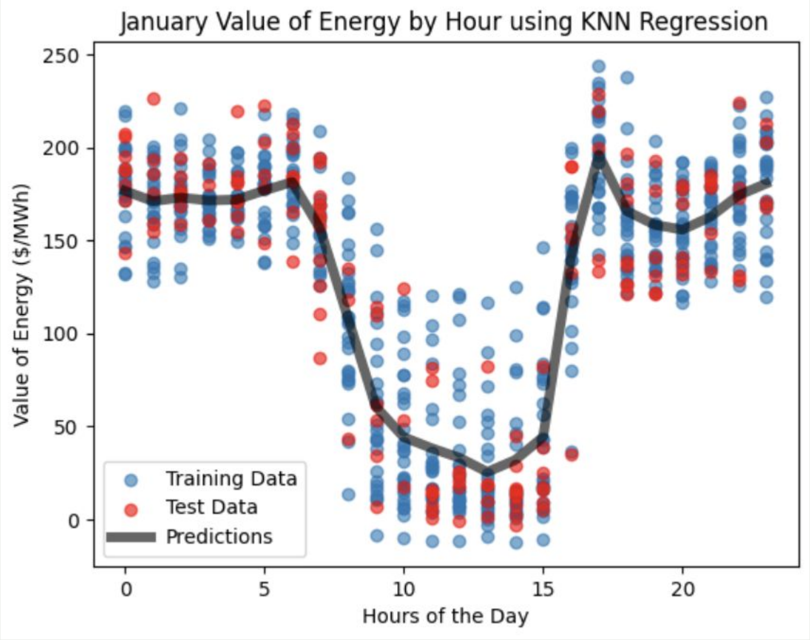
K-Nearest Neighbor model:

- Predicts a data point's value based on the average of its k nearest neighbors.
- **Overfitting provides insight into an ideal error level** for complex model selection (opposite of the naive linear model)

Evaluating model fit using RMSE: **~29.33** ($/MWh)

✓: Captures most of the patterns in the data
✗: Affected by noise so model **not easily generalizable** across months

January:



Complex Benchmark Model 2

Initial regression model building for **Value of Electricity ($/MWh) x Hour** data

Gaussian Basis Modeling

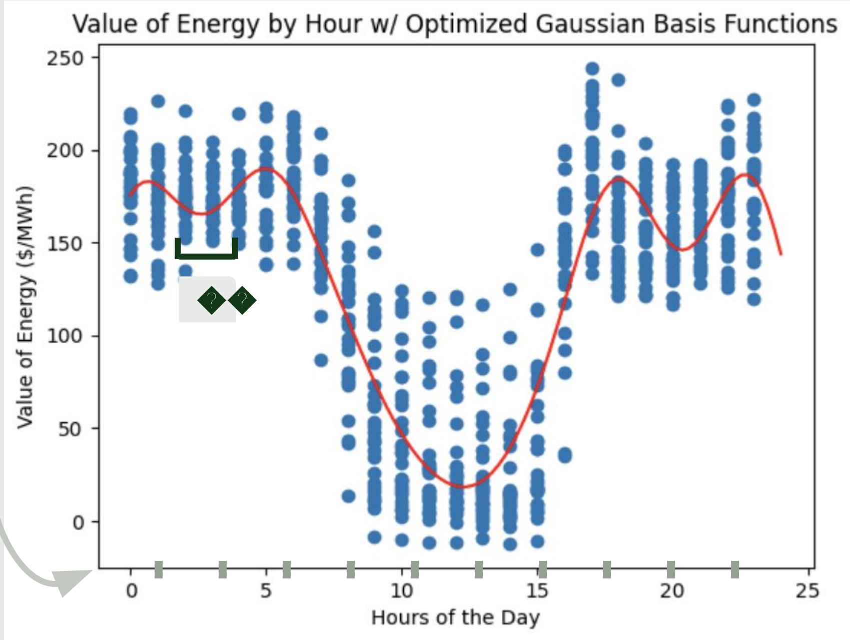
- Transforms non-linear data features into weighted sums of Gaussian curves
- Optimized weights:
    - # of basis functions (n): 10 equally spaced throughout x-axis
    - Width of curves (σ):
        - Different for each basis function

Evaluating model fit using RMSE: **~31.91** ($/MWh)

✓: Controls for **overfitting and learns the data's patterns**

✗: Not generalizable across months

January:



Initial Benchmark Modeling Results

Evaluation of Electricity Value x Hour of Day models: table of **RMSE of each baseline model** for 6 sample months

Sample Months	Linear (naive model)	KNN (optimized/overfit)	Gaussian Bases (overfit)
January	67.09	***29.33***	31.91
February	78.24	37.60	***36.05***
May	66.34	34.45	***33.54***
June	69.15	***27.45***	29.80
August	564.59	511.94	***496.04***
September	346.80	485.61	***323.49***

Base model	**30%** error reduction!	**39%** error reduction!

Initial Benchmark Modeling Results

Evaluation of Electricity Value x Hour of Day models: table of **RMSE of each baseline model** for 6 sample months

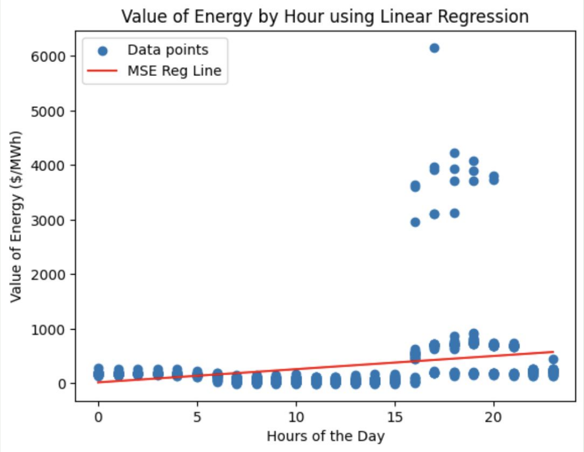
Sample Months	Linear (naive model)	KNN (optimized/overfit)	Gaussian Bases (overfit)
January	67.09	***29.33***	31.91
February	78.24	37.60	***36.05***
May	66.34	34.45	***33.54***
June	69.15	***27.45***	29.80
August	564.59	511.94	***496.04***
September	346.80	485.61	***323.49***

Model Strengths	Linearity allows for monthly **predictability/generalizability.**	**Captures most of the data noise.** Visually shows the patterns of the data.	**Flexibility controls for overfitting. Learns the patterns** of the data. Less error than KNN for outlier months
Model Weaknesses	**Severe underfitting.** Misses patterns in data.	**Not generalizable** across months.	**Not as generalizable** across months.

Addressing Months with Extreme Outliers

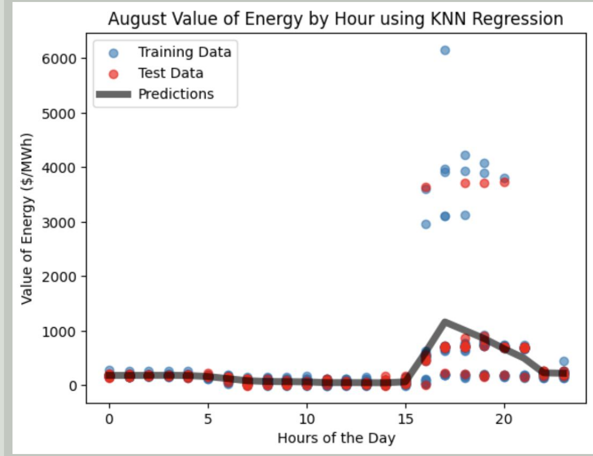
Baseline regression model comparisons for August (month with most outliers):

Linear (naive model)



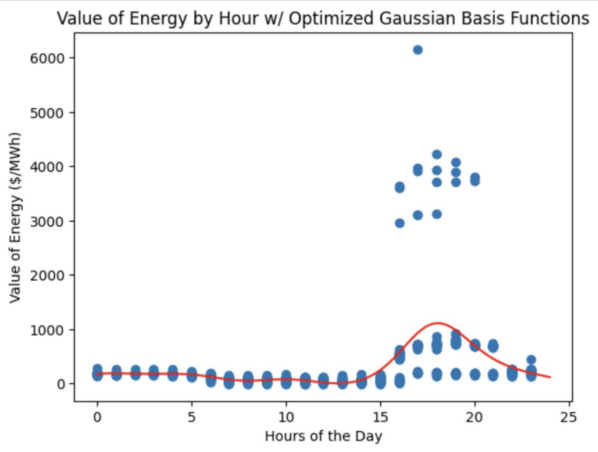
RMSE = 564.59

KNN (overfit)



RMSE = 511.94

Gaussian Bases (overfit)



RMSE = 496.04

Hybrid Model Building

Developing the ideal model that can be easily applied across different months, years, and regions, while remaining flexible enough to account for outliers.

Hybrid Model 1

Complex regression model building for **Value of Electricity ($/MWh) x Hour** data

Clustering + linear regression

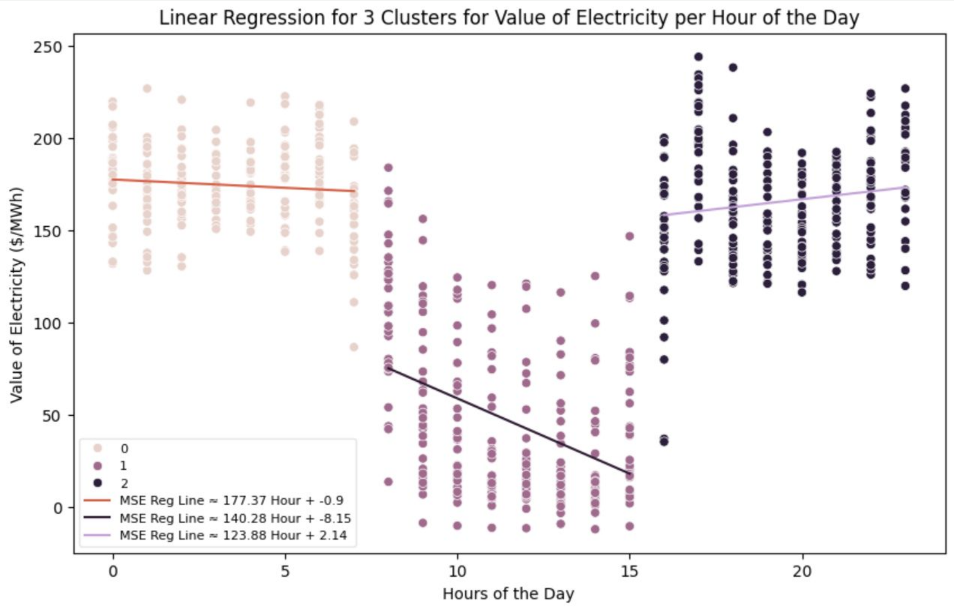
- **3 clusters to capture patterns** in Electricity Value:
  - two clusters for higher Value periods, one cluster for lower Value periods.

Evaluating model fit using RMSE: **~31.69** ($/MWh)

✓: **Linear regression allows for generalizability to other months**

✓: Captures the cyclical pattern of electricity value without overfitting

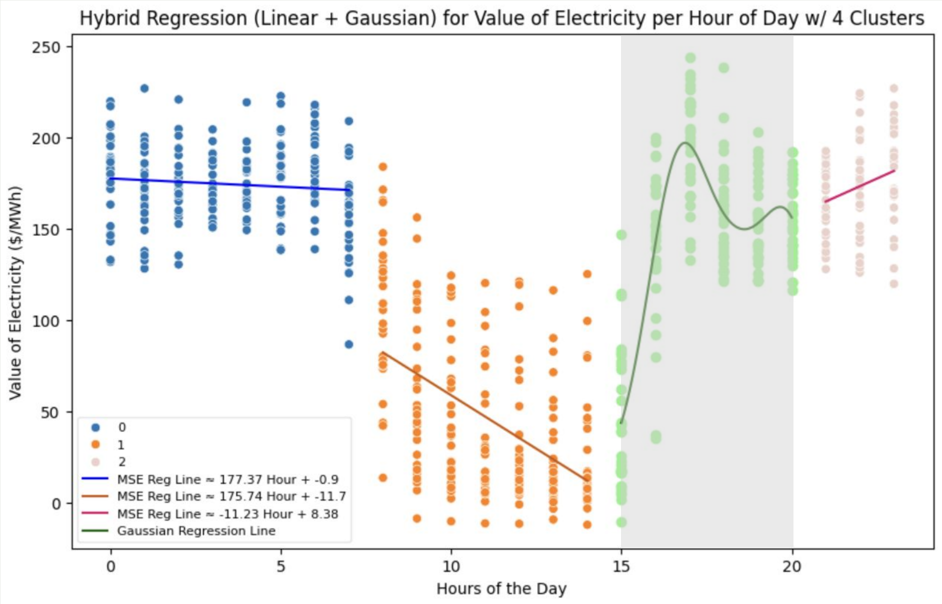
January:



Hybrid Model 2

Complex regression model building for **Value of Electricity ($/MWh) x Hour** data

January:



Clustering + linear + Gaussian regression

- Uses Gaussian for peak hours (15:00 to 20:00)
- Uses Linear Regression for non peak hours
  - 2 clusters before 15:00, 1 cluster after 20:00

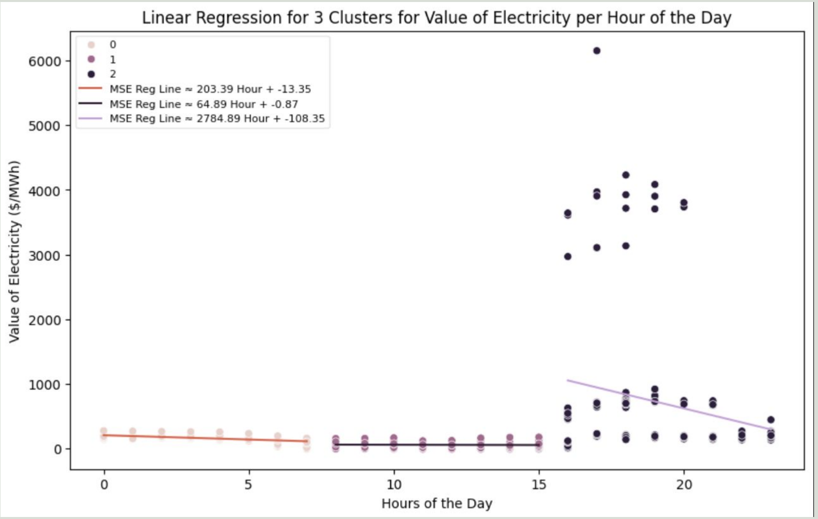
Evaluating model fit using RMSE: **~29.79** ($/MWh)

✓: **Predictability** to other months while **capturing cyclical pattern** during non-peak hours
✓: Accommodates for extreme pricing spikes

Addressing Months with Extreme Outliers

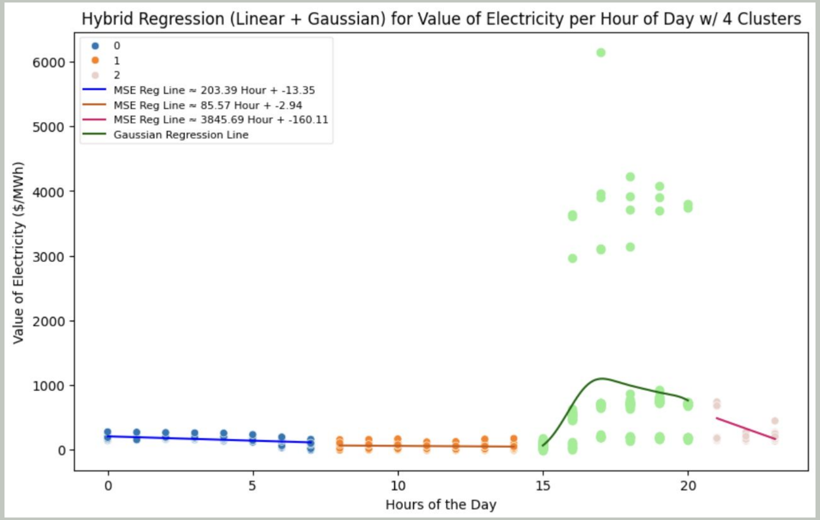
Hybrid model comparisons for August:

Hybrid Model 1: Clustering + linear reg



RMSE = 502.98

Hybrid Model 2: Clustering + linear reg + Gaussian bases



RMSE = 491.99

Final Model Evaluation Comparisons

Evaluation of Electricity Value x Hour of Day models: **RMSE comparison table** for each model for 6 sample month

	Benchmark Models			Proposed Models	
Sample Months	Linear	KNN	Gaussian Bases	Linear + Cluster	Linear + Gaussian + Cluster
January	67.09	***29.33***	31.91	31.69	29.79
February	78.24	37.60	36.05	38.99	***34.91***
May	66.34	34.45	***33.54***	42.81	33.81
June	69.15	***27.45***	29.80	40.48	31.92
August	564.59	511.94	496.04	502.98	***491.99***
September	346.80	485.61	323.49	326.67	***322.78***
		30% reduction	39% reduction	33% reduction!	39% reduction!

Preliminary conclusions

Discussion and **key takeaways and implications** from completed analysis

1. **Clustered Linear + Gaussian Model most optimized for most months (notably outlier months)**
   - Further research can exploring other types hybrid models, using horizontal regression lines for non-peak data and other complex models (e.g. polynomial) for peak periods to improve feasibility.

2. **Clustered Linear Model best generalizable model while capturing data's cyclicality**
   - Easier pricing scheme for utility companies to practically implement and for consumers to interpret.

Creating multivariable regression models

Developing a flexible model for fixed, marginal, and total electricity costs that accommodates to temporal, locational, and outlier factors for SDG&E zones.

Marginal Electricity Value Model

Using E3 ACC data from 2020, 2021, 2022, 2024 for SDG&E Climate Zones

Model 1: Info that Utilities, ISOs and Planners Have

Determinants: Year, Hour, Summer x Peak, Energy, Generation, Transmission, Distribution
RMSE: 1.77 $/MWh
R^2: Model accounts for 100% of variation in energy value

OLS Regression Results

Dep. Variable:	Total	R-squared:	1.000
Model:	OLS	Adj. R-squared:	1.000
Method:	Least Squares	F-statistic:	8.772e+07
Date:	Tue, 22 Apr 2025	Prob (F-statistic):	0.00
Time:	20:36:46	Log-Likelihood:	-62506.
No. Observations:	28032	AIC:	1.250e+05
Df Residuals:	28022	BIC:	1.251e+05
Df Model:	9		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	4.1884	0.038	110.737	0.000	4.114	4.263
Year[T.2021]	-1.7161	0.038	-44.975	0.000	-1.791	-1.641
Year[T.2022]	-0.9293	0.038	-24.389	0.000	-1.004	-0.855
Year[T.2024]	-1.4101	0.039	-35.936	0.000	-1.487	-1.333
Hour	0.0198	0.002	9.725	0.000	0.016	0.024
SummerxPeak	0.3953	0.054	7.345	0.000	0.290	0.501
Energy	1.1413	0.001	1892.225	0.000	1.140	1.142
Generation	0.9993	4.17e-05	2.4e+04	0.000	0.999	0.999
Transmission	0.9997	0.000	8988.128	0.000	0.999	1.000
Distribution	0.9726	0.002	438.539	0.000	0.968	0.977

Model 2: Info that Consumers Have

Determinants: Year, Hour, Summer x Peak, Energy, Generation
RMSE: 114.98 $/MWh ← also much lower than earlier simple regressions
R^2: Model accounts for 89.6% of variation in energy value

OLS Regression Results

Dep. Variable:	Total	R-squared:	0.896
Model:	OLS	Adj. R-squared:	0.896
Method:	Least Squares	F-statistic:	3.446e+04
Date:	Tue, 22 Apr 2025	Prob (F-statistic):	0.00
Time:	20:21:27	Log-Likelihood:	-1.7441e+05
No. Observations:	28032	AIC:	3.488e+05
Df Residuals:	28024	BIC:	3.489e+05
Df Model:	7		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-4.1797	2.047	-2.042	0.041	-8.192	-0.168
Year[T.2021]	-1.8048	2.066	-0.873	0.382	-5.855	2.245
Year[T.2022]	16.3053	2.061	7.912	0.000	12.266	20.345
Year[T.2024]	-0.4871	2.124	-0.229	0.819	-4.651	3.677
Hour	-0.0326	0.111	-0.295	0.768	-0.249	0.184
SummerxPeak	57.6618	2.871	20.081	0.000	52.034	63.290
Energy	1.3097	0.033	40.211	0.000	1.246	1.374
Generation	1.0389	0.002	463.663	0.000	1.035	1.043

Fixed Electricity Value Model

Using E3 ACC data from 2020, 2021, 2022, 2024 for SDG&E Climate Zones

Fixed Value determinants: Greenhouse Gas Rebalancing Values, Losses, Methane Leakage
RMSE: 14.41 $/Mwh
R^2: Model accounts for ~89% of variation in energy value

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                  Total   R-squared:                       0.886
Model:                            OLS   Adj. R-squared:                  0.886
Method:                 Least Squares   F-statistic:                 7.276e+04
Date:                Tue, 22 Apr 2025   Prob (F-statistic):               0.00
Time:                        21:12:09   Log-Likelihood:            -1.1642e+05
No. Observations:               28032   AIC:                         2.329e+05
Df Residuals:                   28028   BIC:                         2.329e+05
Df Model:                           3
Covariance Type:            nonrobust
==================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------
Intercept        -17.4637      0.223    -78.246      0.000     -17.901     -17.026
GHGRebalancing    -0.7385      0.024    -30.580      0.000      -0.786      -0.691
Losses             8.6528      0.060    143.116      0.000       8.534       8.771
MethaneLeak       20.1569      0.098    205.568      0.000      19.965      20.349
==============================================================================
```

Total Electricity Value Model

Using E3 ACC data from 2020, 2021, 2022, 2024 for SDG&E Climate Zones

Fixed Value determinants: Hour, Energy, Generation, Transmission, Distribution, Methane Leak
RMSE: 4.51
R^2: Model accounts for 100% of variation in energy value

OLS Regression Results

Dep. Variable:	Total	R-squared:	1.000
Model:	OLS	Adj. R-squared:	1.000
Method:	Least Squares	F-statistic:	2.589e+07
Date:	Tue, 22 Apr 2025	Prob (F-statistic):	0.00
Time:	21:27:34	Log-Likelihood:	-85410.
No. Observations:	28032	AIC:	1.708e+05
Df Residuals:	28025	BIC:	1.709e+05
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-0.6696	0.071	-9.443	0.000	-0.809	-0.531
Hour	0.0271	0.005	5.956	0.000	0.018	0.036
Energy	1.2726	0.001	981.401	0.000	1.270	1.275
Generation	0.9988	9.37e-05	1.07e+04	0.000	0.999	0.999
Transmission	0.9992	0.000	3993.526	0.000	0.999	1.000
Distribution	0.9597	0.005	192.883	0.000	0.950	0.969
MethaneLeak	6.2117	0.019	332.843	0.000	6.175	6.248

Multivariable Regression Discussion

For marginal, fixed, and total values of electricity

1. **Energy value data is consistent across all climate zones within each utility** (SDG&E, PG&E, and SCE)
   - Indicates that the proposed multivariable models are **adaptable across different geographic locations** within each utility's service area

2. **Primary drivers of marginal electricity value are:**
   - Energy, Generation, Transmission, and Distribution costs
   - Yearly decline observed in marginal electricity value, indicating a general downward trend over time.
   - The **interaction between Summer and Peak Hours is a strong (Summer x Peak), significant predictor**, although summer alone not statistically significant.

3. **Primary drivers of fixed electricity value are:**
   - Greenhouse Gas Rebalancing, Losses, Methane Leakage
   - Strong collinearity between variables and GHG Adder show that they are all closely related

4. **Multivariate models are effective for understanding price dynamics, but challenging for consumers to interpret**
   - Future models can consider **disaggregation by location/utility** and **temporal segmentation**

Future research direction

Refining complex models and **real-world data application**

1. **Continue developing interpretable multivariate regression models**
   - Decide on good model evaluation statistics and evaluate model using each statistic
   - Exploring alternative dynamic pricing methods using machine learning predictions

2. **Validation and case studies**
   - Testing models on real energy market data in regions outside of CAISO
   - Comparing historical pricing trends across years v. proposed model to predict electricity pricing for upcoming years
   - Conducting surveys that take into account customer input

3. **Application to public policy**
   - Evaluate regulatory constraints from established public policy that may affect pricing model implementation

Thank You!