Long-Term Electricity Price Prediction

Shikhar Verma, Nami Saghaei,
Kun Yang, Randell Espina,
Maria Sheila Ramos,
Ashkan Yousefi

Mentors:
Antonio Vitti, CFO formerly with Merchant Atlas, Inc.
Dr. Steven Gustafson, Chief Scientist at Maana

Objective

Predict the monthly retail price of electricity 3-5 years in advance.

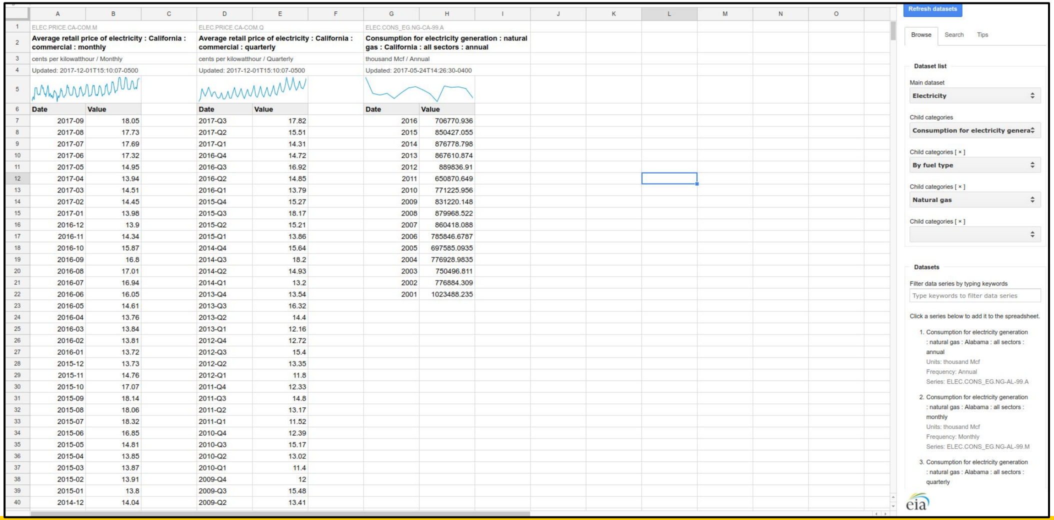
Data Collection and Cleaning

EIA “OpenData” Database

Biggest hiccup was finding enough data…

Spent majority of time selecting, cleaning, formatting, and merging data.

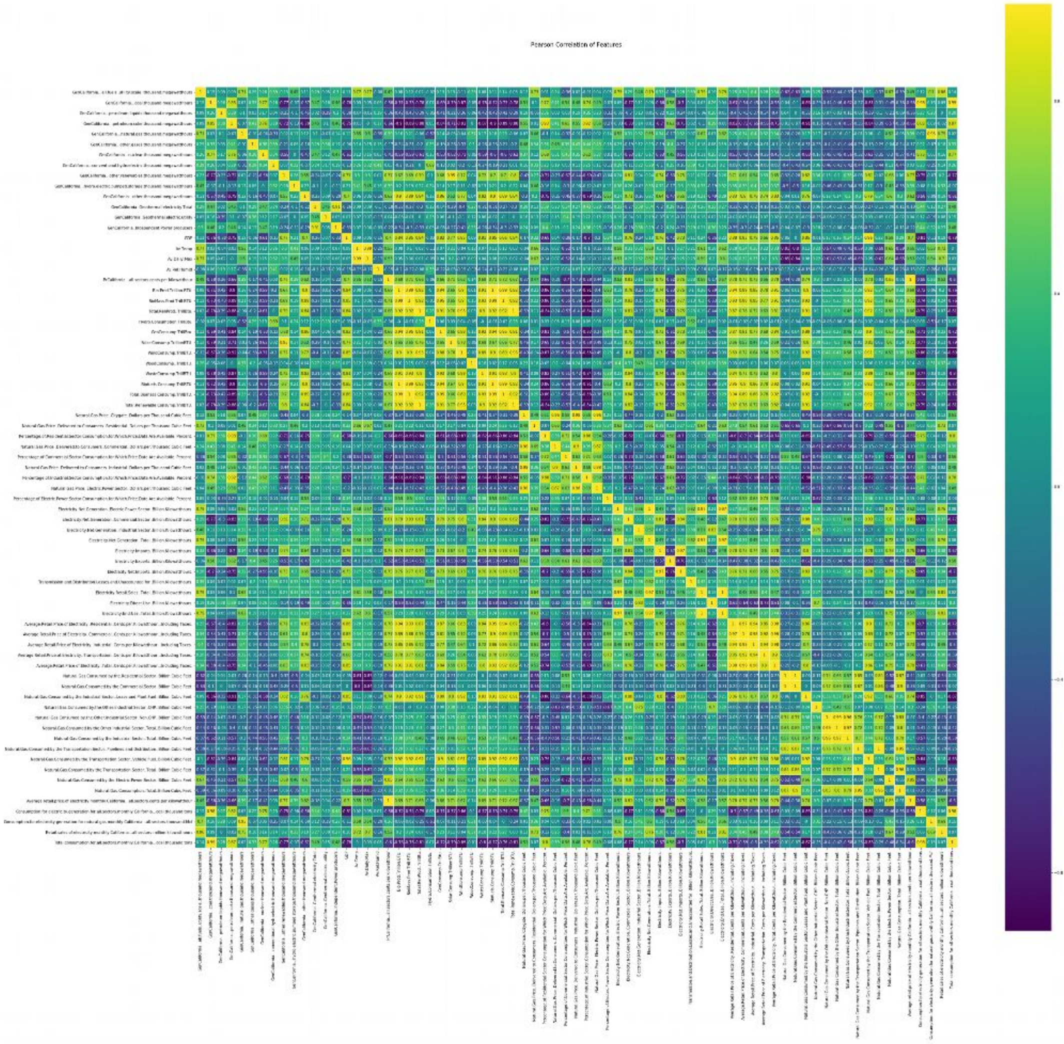
Up to last week we had 13 columns in our potential feature matrix. Now we have 78.



	A	B	D	E	G	H
1	ELEC.PRICE.CA-COM.M		ELEC.PRICE.CA-COM.Q		ELEC.CONS_EG.NG-CA-99.A	
2	**Average retail price of electricity : California : commercial : monthly**		**Average retail price of electricity : California : commercial : quarterly**		**Consumption for electricity generation : natural gas : California : all sectors : annual**	
3	cents per kilowatthour / Monthly		cents per kilowatthour / Quarterly		thousand Mcf / Annual	
4	Updated: 2017-12-01T15:10:07-0500		Updated: 2017-12-01T15:10:07-0500		Updated: 2017-05-24T14:26:30-0400	
5						
6	**Date**	**Value**	**Date**	**Value**	**Date**	**Value**
7	2017-09	18.05	2017-Q3	17.82	2016	706770.936
8	2017-08	17.73	2017-Q2	15.51	2015	850427.055
9	2017-07	17.69	2017-Q1	14.31	2014	876778.798
10	2017-06	17.32	2016-Q4	14.72	2013	867610.874
11	2017-05	14.95	2016-Q3	16.92	2012	889836.91
12	2017-04	13.94	2016-Q2	14.85	2011	650870.649
13	2017-03	14.51	2016-Q1	13.79	2010	771225.956
14	2017-02	14.45	2015-Q4	15.27	2009	831220.148
15	2017-01	13.98	2015-Q3	18.17	2008	879968.522
16	2016-12	13.9	2015-Q2	15.21	2007	860418.088
17	2016-11	14.34	2015-Q1	13.86	2006	785846.6787
18	2016-10	15.87	2014-Q4	15.64	2005	697585.0935
19	2016-09	16.8	2014-Q3	18.2	2004	776928.9835
20	2016-08	17.01	2014-Q2	14.93	2003	750496.811
21	2016-07	16.94	2014-Q1	13.2	2002	776884.309
22	2016-06	16.05	2013-Q4	13.54	2001	1023488.235
23	2016-05	14.61	2013-Q3	16.32		
24	2016-04	13.76	2013-Q2	14.4		
25	2016-03	13.84	2013-Q1	12.16		
26	2016-02	13.81	2012-Q4	12.72		
27	2016-01	13.72	2012-Q3	15.4		
28	2015-12	13.73	2012-Q2	13.35		
29	2015-11	14.76	2012-Q1	11.8		
30	2015-10	17.07	2011-Q4	12.33		
31	2015-09	18.14	2011-Q3	14.8		
32	2015-08	18.06	2011-Q2	13.17		
33	2015-07	18.32	2011-Q1	11.52		
34	2015-06	16.85	2010-Q4	12.39		
35	2015-05	14.81	2010-Q3	15.17		
36	2015-04	13.85	2010-Q2	13.02		
37	2015-03	13.87	2010-Q1	11.4		
38	2015-02	13.91	2009-Q4	12		
39	2015-01	13.8	2009-Q3	15.48		
40	2014-12	14.04	2009-Q2	13.41		

Correlation Analysis

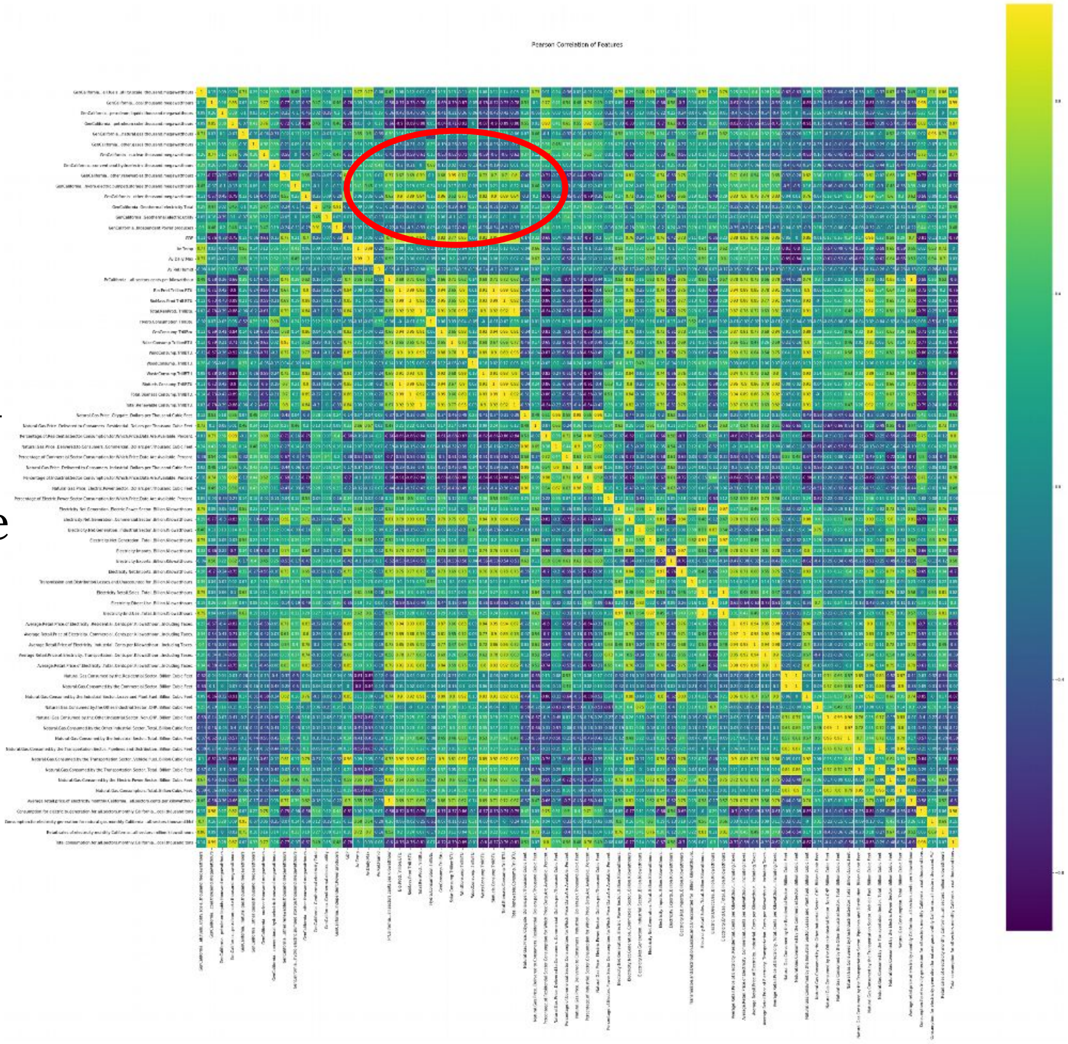
For feature selection


Pearson Correlation of Features

GDP

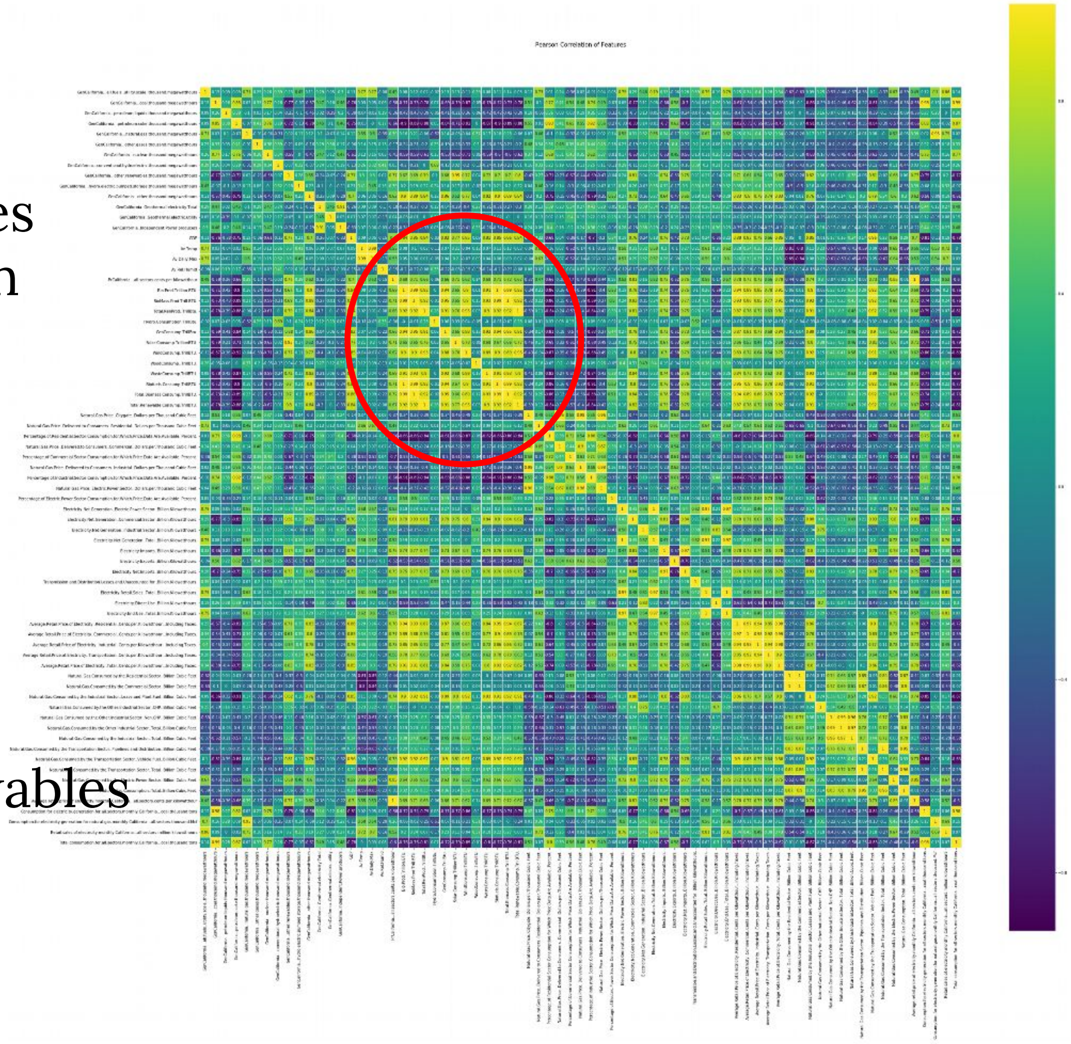
Renewables Consumption

...and Average Retail Electricity Price



Renewables Production

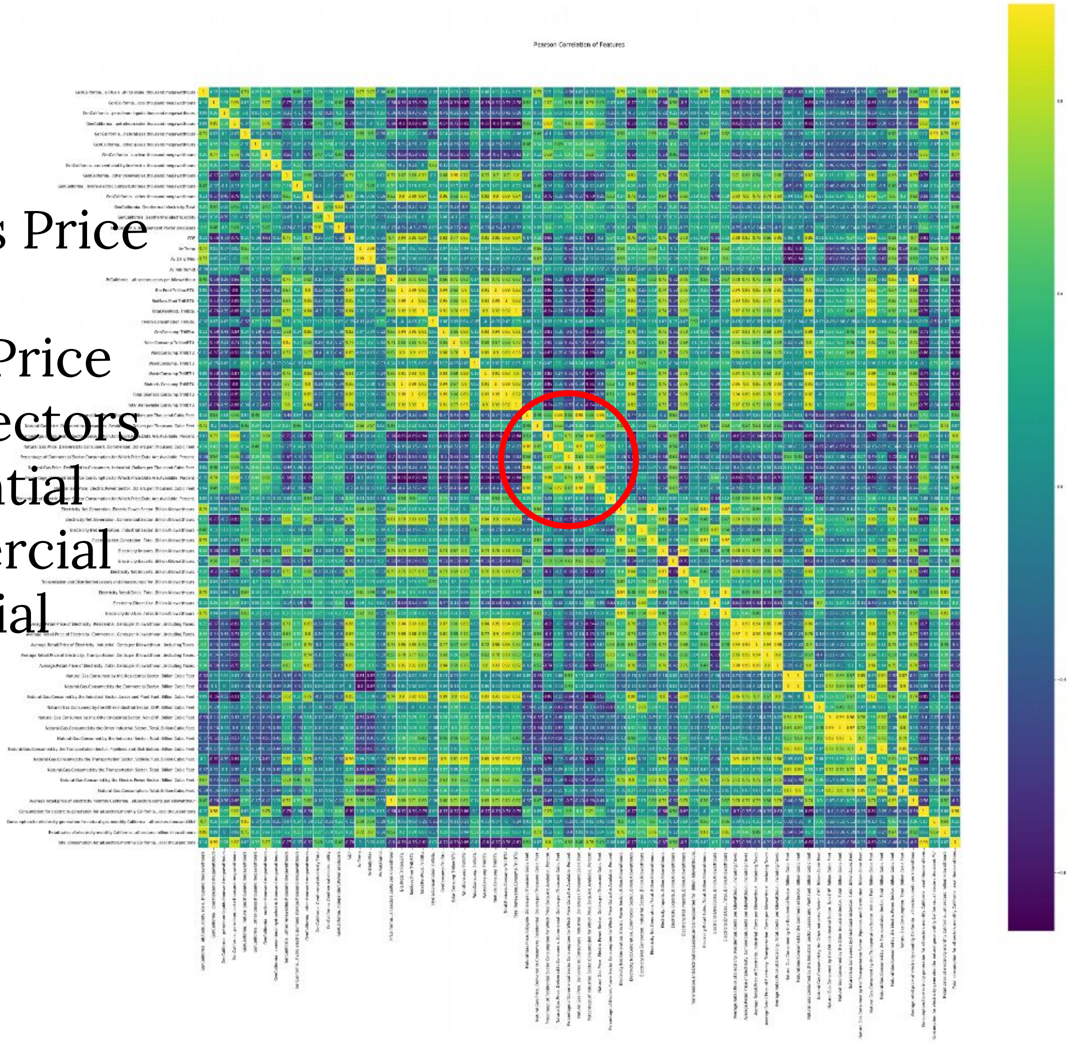
- Bio
- Geo
- Solar
- Wind
- Hydro
- Waste
- Total Renewables



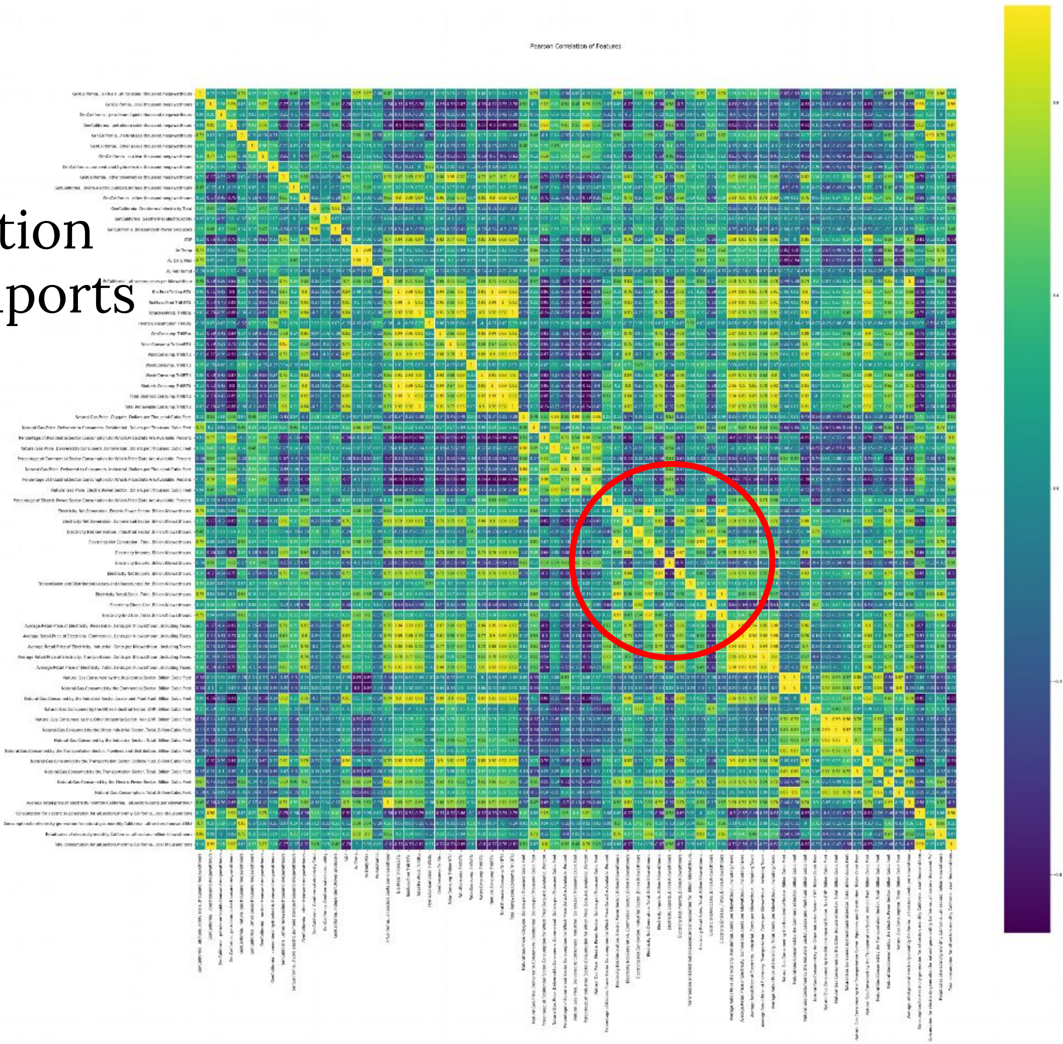
Natural Gas Price

Electricity Price
Different Sectors

- Residential
- Commercial
- Inudstrial

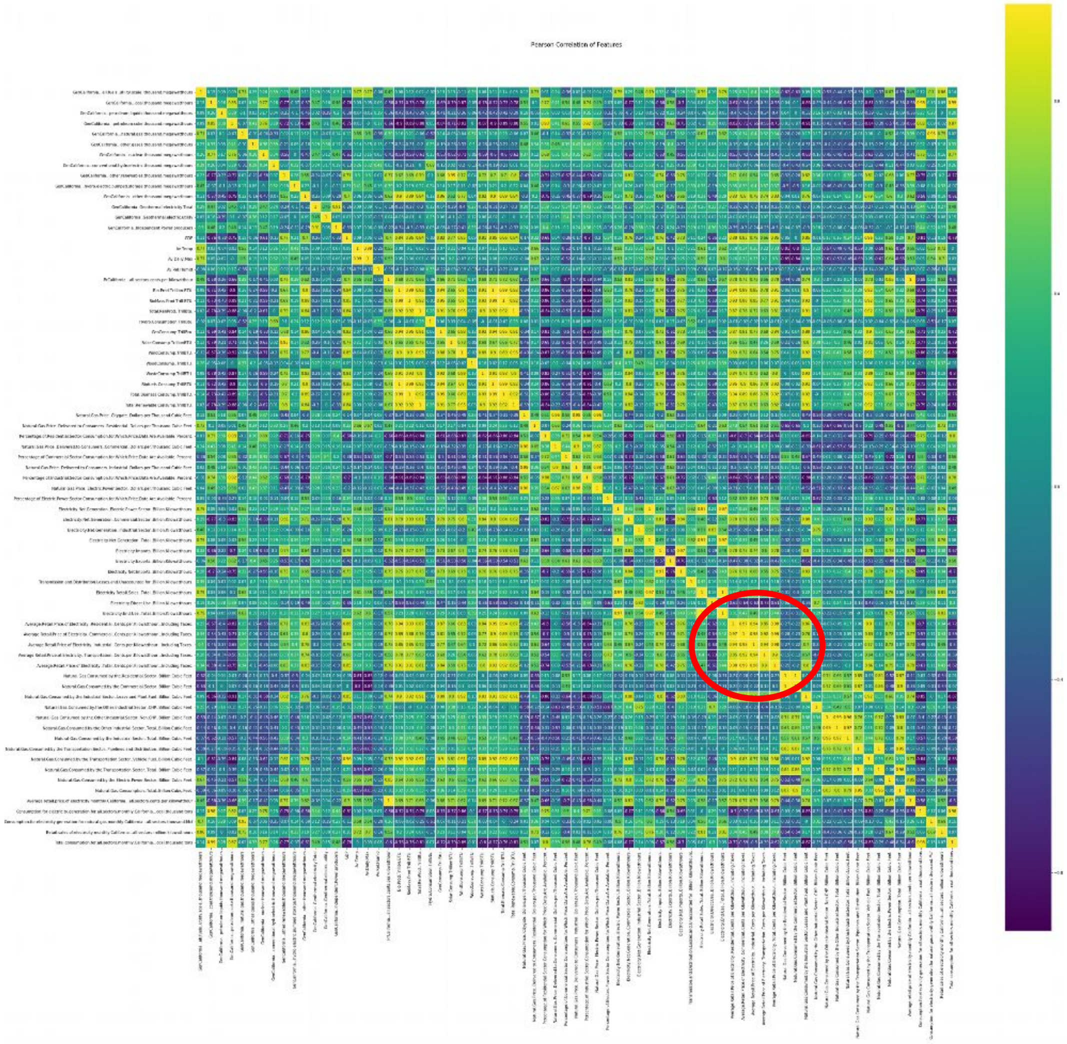


Net Generation Exports/Imports in different sectors

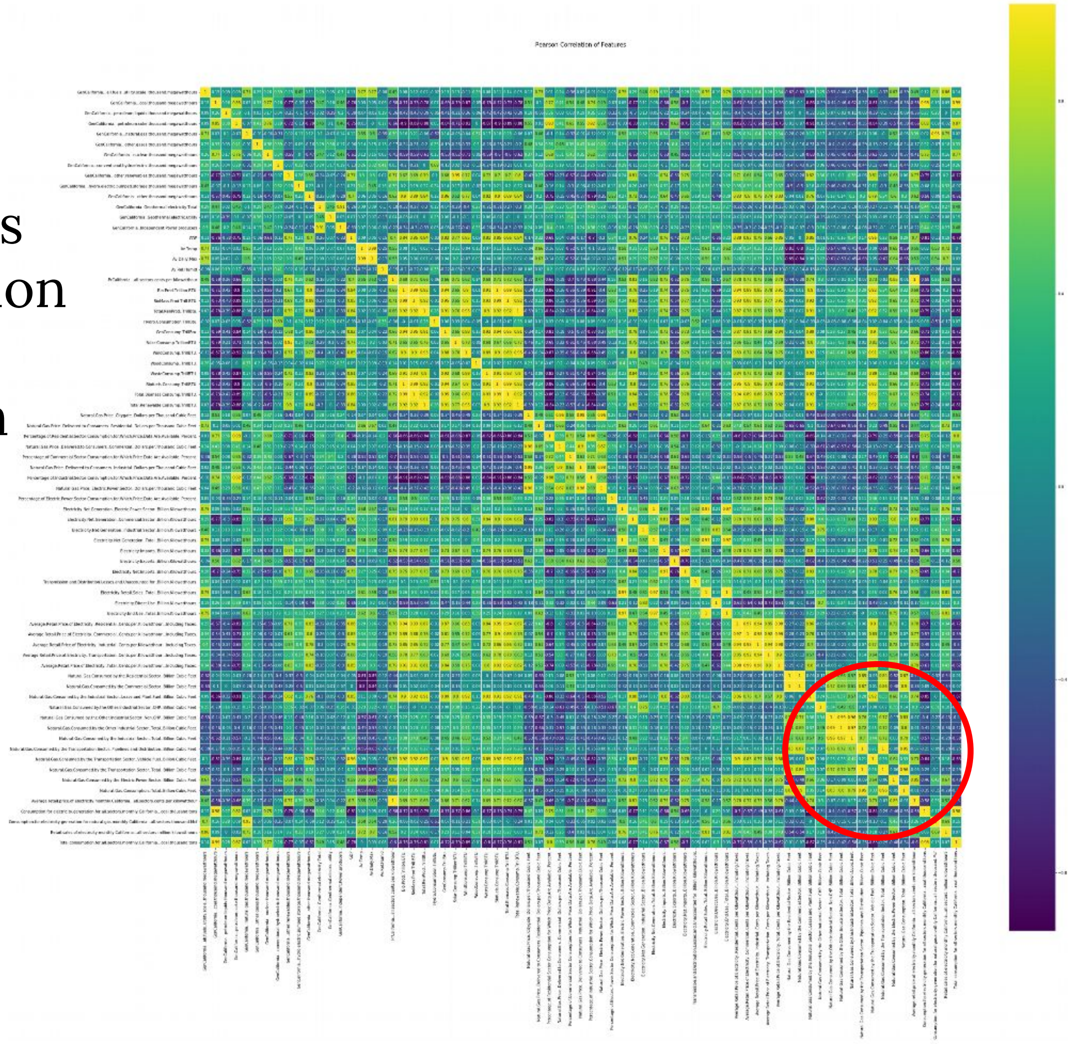


Average Electricity Retail Price

Different Sectors/ Units



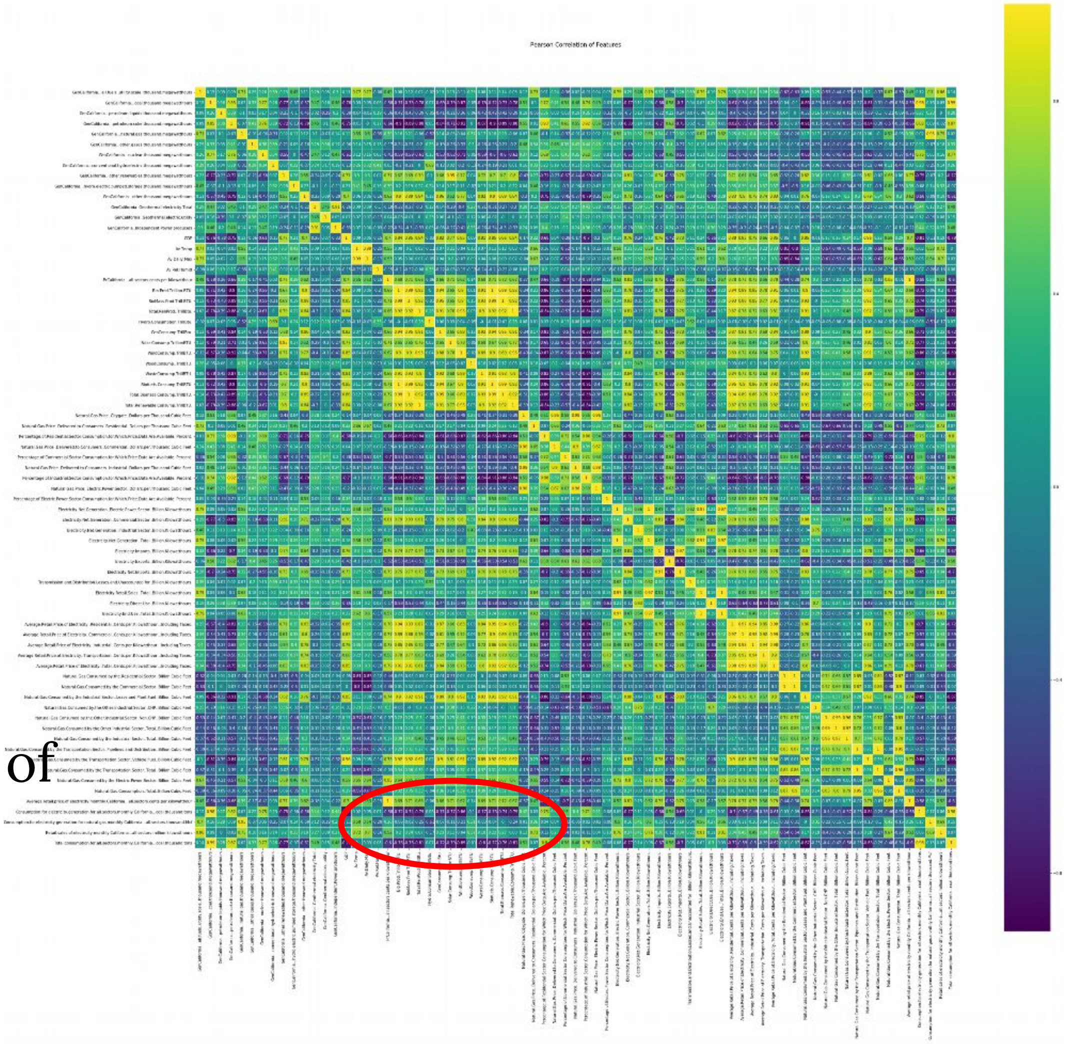
Natural Gas Consumption and Production by sector.



Petroleum
Coke
and Coal
Generation

... And

Average
Monthly
Retail Price of
Electricity



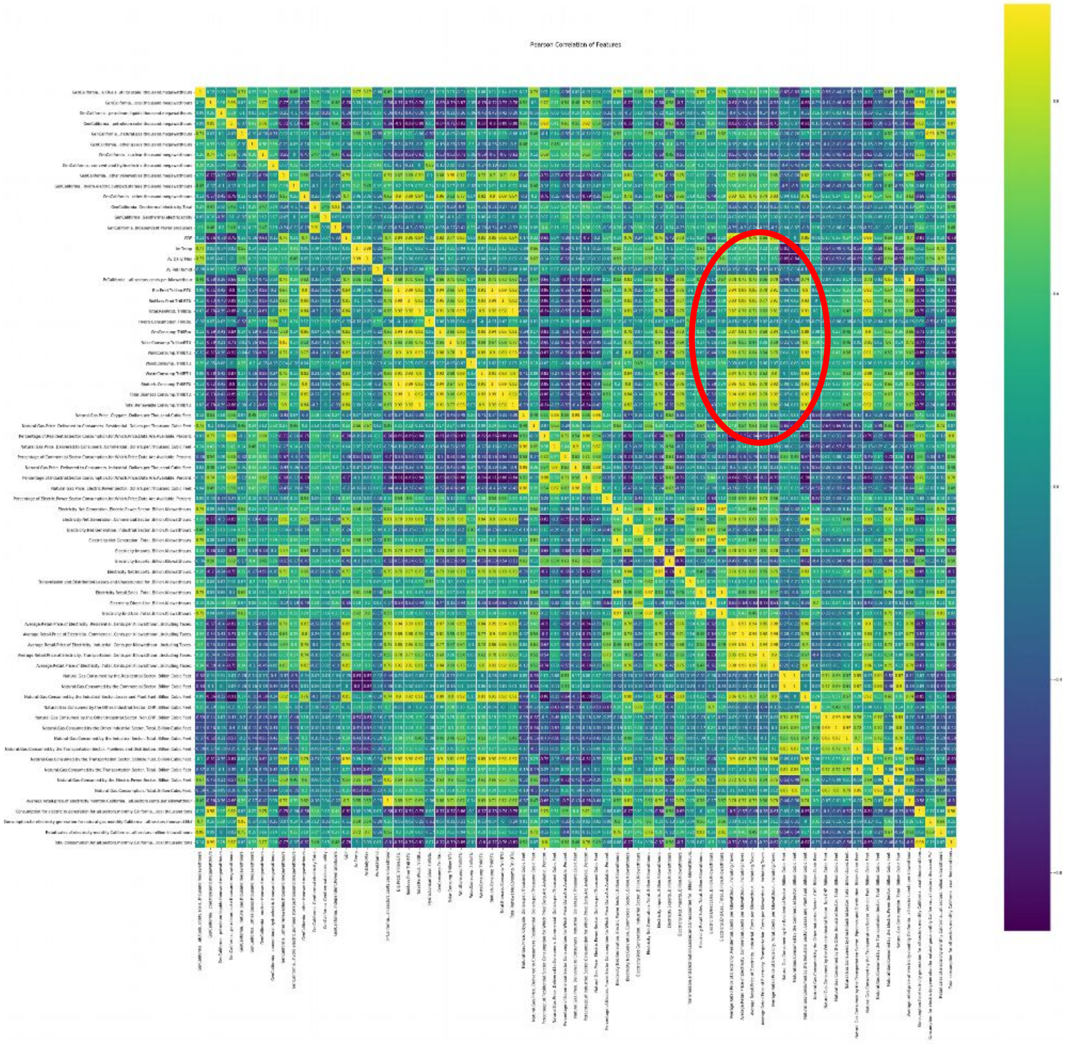
Renewables Production and Consumption

- Bio
- Geo
- Solar
- Wind
- Hydro
- Waste
- Total Renewables

And...

- Average Monthly Retail Electricity Price in different sectors.

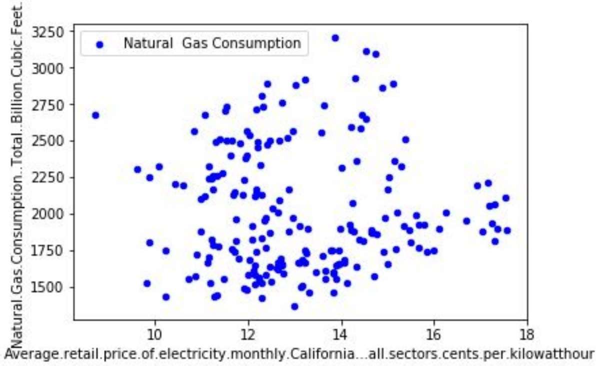
- Electricity Import Price

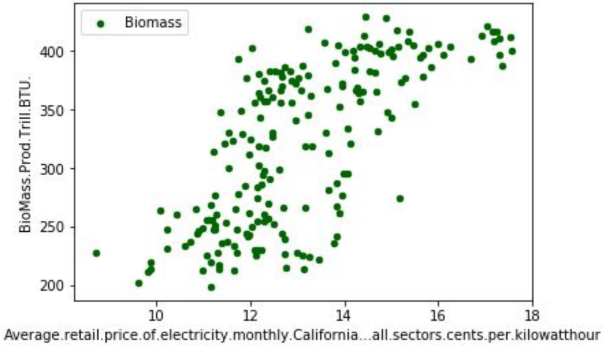


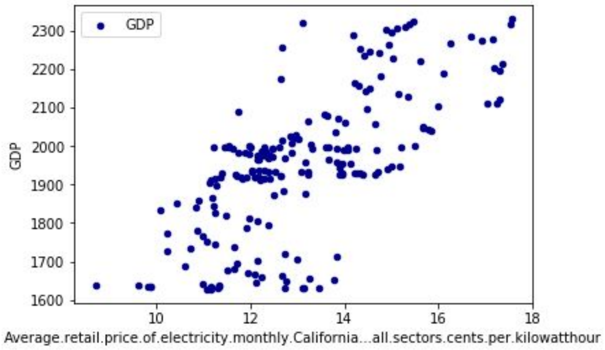
Correlations Summary (Features)

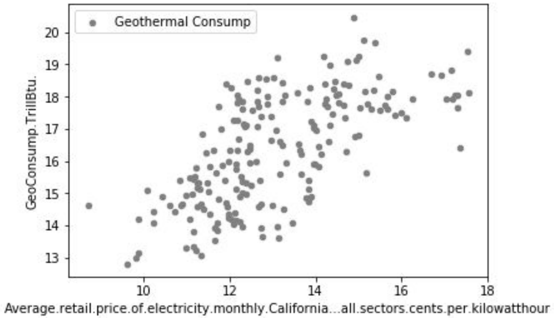
- GDP (.7)
- Natural Gas Consumed by Industrial Sector (.74)
- Natural Gas Consumed by Electric Power Sector (.85)
- Net Electricity Imports (.75)
- Net Electricity Generation (.81)

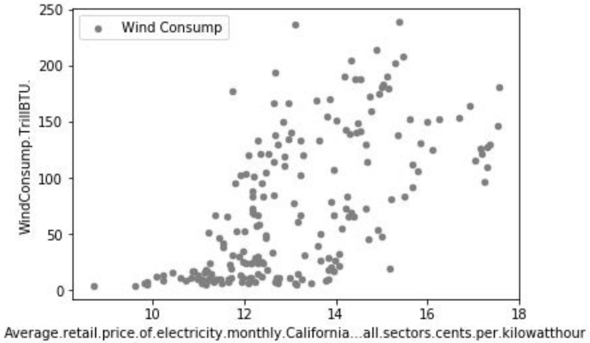
- Renewables (.77)
  - Solar Consumption (.71)
  - Geothermal Consumption (.65)
  - Wind Consumption (.62)
  - Biofuels Production (.72)

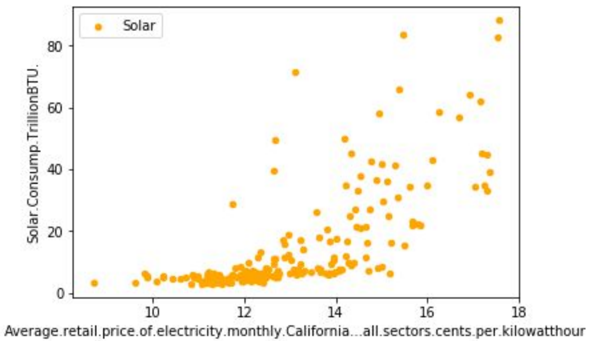

Natural Gas Consumption
Natural.Gas.Consumption..Total..Billion.Cubic.Feet.
Average.retail.price.of.electricity.monthly.California...all.sectors.cents.per.kilowatthour

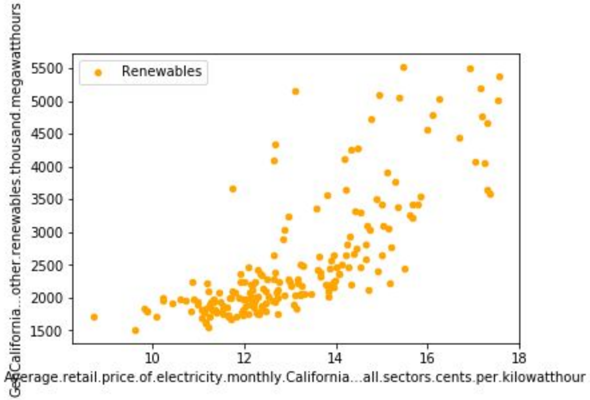

Biomass
BioMass.Prod.Trill.BTU.
Average.retail.price.of.electricity.monthly.California...all.sectors.cents.per.kilowatthour


GDP
GDP
Average.retail.price.of.electricity.monthly.California...all.sectors.cents.per.kilowatthour


Geothermal Consump
GeoConsump.TrillBtu.
Average.retail.price.of.electricity.monthly.California...all.sectors.cents.per.kilowatthour


Wind Consump
WindConsump.TrillBTU.
Average.retail.price.of.electricity.monthly.California...all.sectors.cents.per.kilowatthour


Solar
Solar.Consump.TrillionBTU.
Average.retail.price.of.electricity.monthly.California...all.sectors.cents.per.kilowatthour


Renewables
Gen.California...other.renewables.thousand.megawatthours
Average.retail.price.of.electricity.monthly.California...all.sectors.cents.per.kilowatthour

Prediction

Time Series

Prediction with Time Series

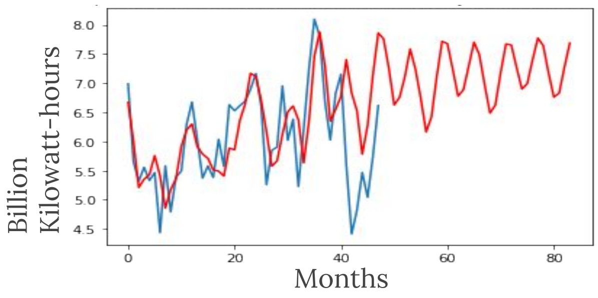
- Time Series Prediction (Autoregressive Integrated Moving Average or ARIMA) for 3 years in advance – train 2000–2014, test 2014–2017, predict 2017–2020.

- … for 5 years in advance – train 2000–2012, test 2012–2017, predict 2017–2022.

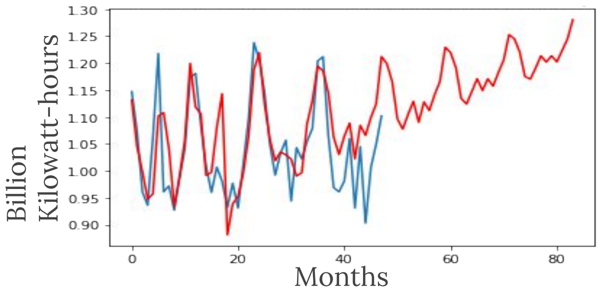
Time Series Results

- Time Series Prediction for 3 years performed superbly.
  - Considered Cross-validation, but discovered that doesn’t work well with time series because it misses trends and captures other erroneous trends. Time series depends on history.
- Time Series for 5 years performed slightly worse, but still good.

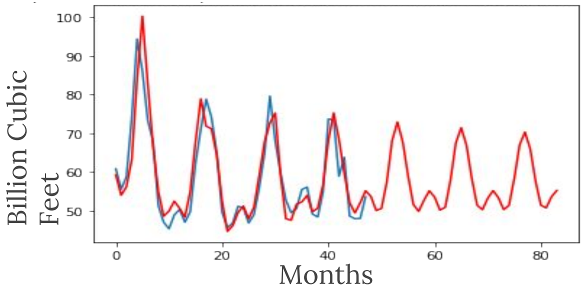
Electricity Imports (3 Years)



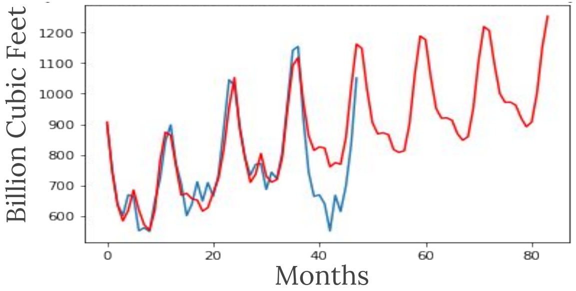
Net Commercial Generation (3 Years)

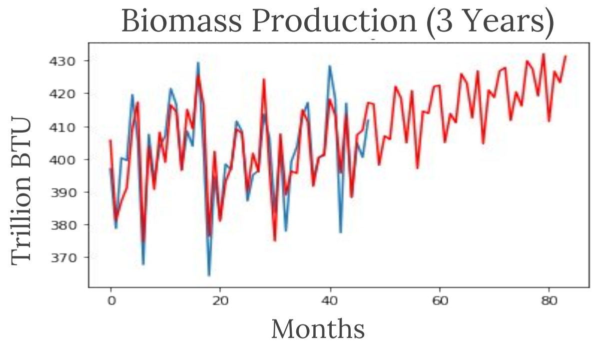


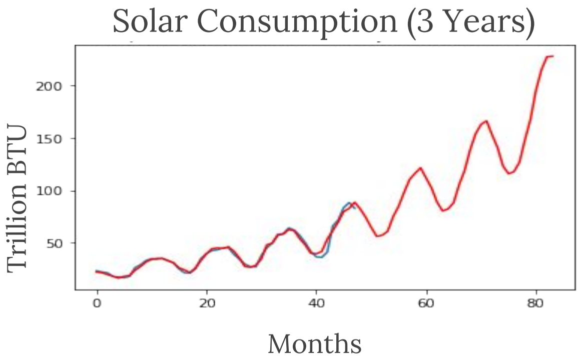
Natural Gas Consumed by Industrial Sector (3 Years)

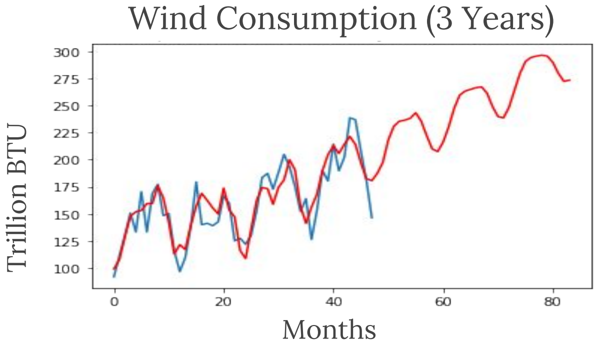


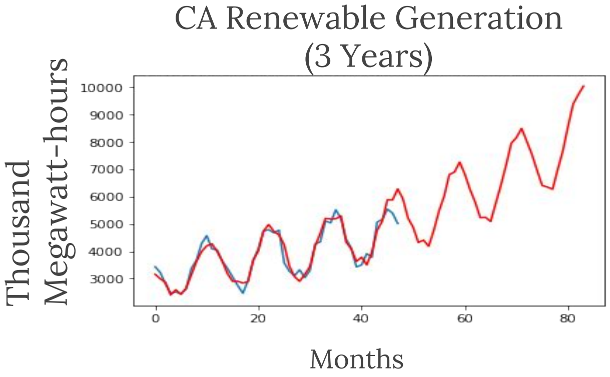
Natural Gas Consumed by Electric Power Sector (3 Years)



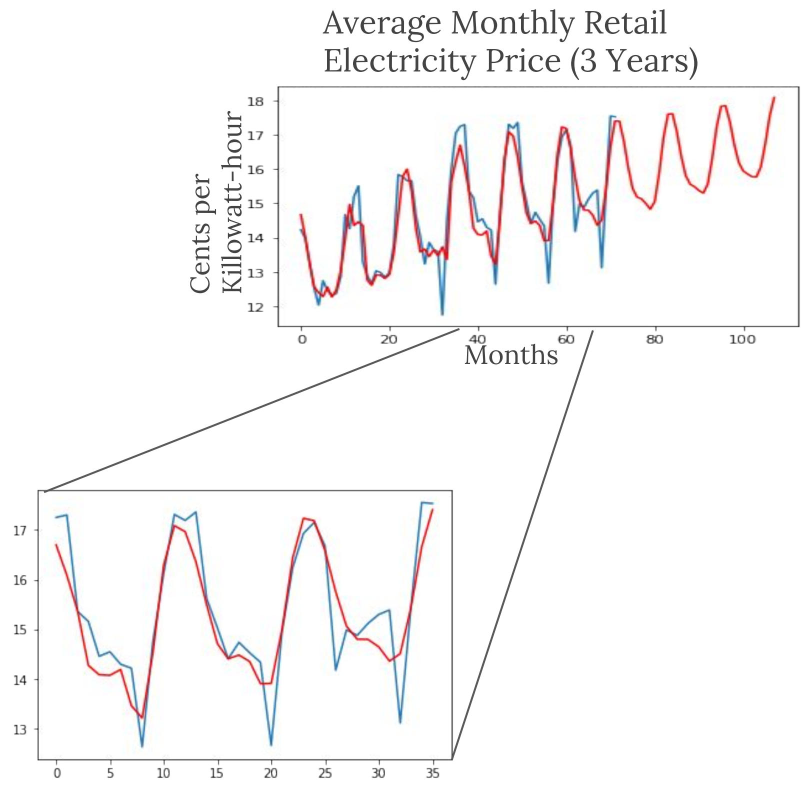

Biomass Production (3 Years)
Trillion BTU
430
420
410
400
390
380
370
0
20
40
60
80
Months


Solar Consumption (3 Years)
Trillion BTU
200
150
100
50
0
20
40
60
80
Months

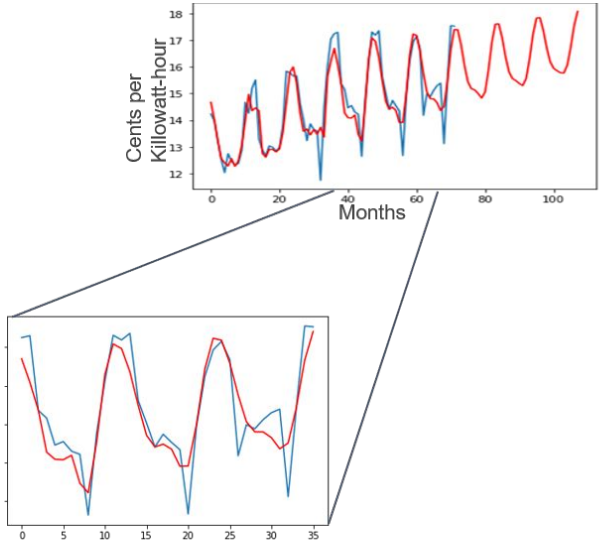

Wind Consumption (3 Years)
Trillion BTU
300
275
250
225
200
175
150
125
100
0
20
40
60
80
Months


CA Renewable Generation
(3 Years)
Thousand
Megawatt-hours
10000
9000
8000
7000
6000
5000
4000
3000
0
20
40
60
80
Months

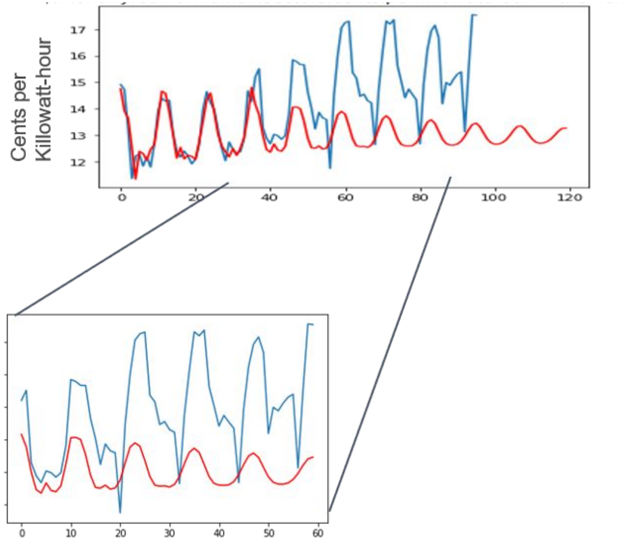
```
Lag: 14
Coefficients: [-0.06925843  0.62417522  0.04481464  0.01982984 -0.02458993  0.02896534
 -0.03363928  0.08095022 -0.07569835  0.03176049 -0.02809795  0.29576012
  0.33499937 -0.07588766 -0.21199173]
predicted=16.698876, expected=17.250000
predicted=16.086452, expected=17.300000
predicted=15.367021, expected=15.360000
predicted=14.276157, expected=15.160000
predicted=14.089079, expected=14.460000
predicted=14.078031, expected=14.550000
predicted=14.193885, expected=14.300000
predicted=13.466611, expected=14.220000
predicted=13.219550, expected=12.640000
predicted=14.530995, expected=14.760000
predicted=16.291482, expected=16.120000
predicted=17.084774, expected=17.310000
predicted=16.963351, expected=17.190000
predicted=16.362001, expected=17.360000
predicted=15.494480, expected=15.620000
predicted=14.713708, expected=15.030000
predicted=14.410178, expected=14.410000
predicted=14.484953, expected=14.740000
predicted=14.354295, expected=14.530000
predicted=13.910401, expected=14.340000
predicted=13.914972, expected=12.670000
predicted=14.993065, expected=14.940000
predicted=16.436993, expected=16.240000
predicted=17.233379, expected=16.930000
predicted=17.184642, expected=17.150000
predicted=16.586172, expected=16.680000
predicted=15.756495, expected=14.180000
predicted=15.066525, expected=14.990000
predicted=14.802596, expected=14.880000
predicted=14.799852, expected=15.120000
predicted=14.650228, expected=15.300000
predicted=14.362594, expected=15.390000
predicted=14.512454, expected=13.120000
predicted=15.441955, expected=15.460000
predicted=16.661402, expected=17.550000
predicted=17.405589, expected=17.530000
Test MSE: 0.387
```



Average Monthly Retail Electricity Price (3 Years)



Average Monthly Retail Electricity Price (5 Years)



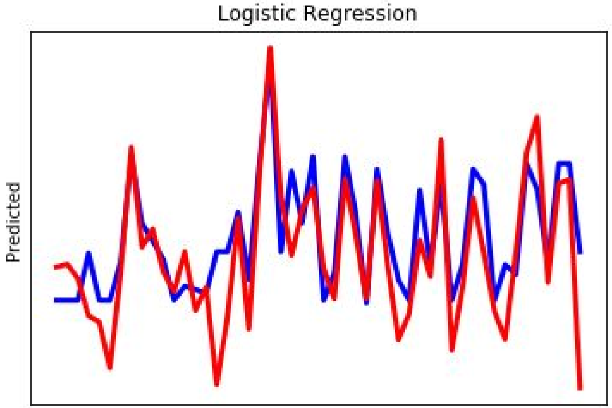
MSE (Mean Square Error)=**0.38**

MSE (Mean Square Error)=**4.972**

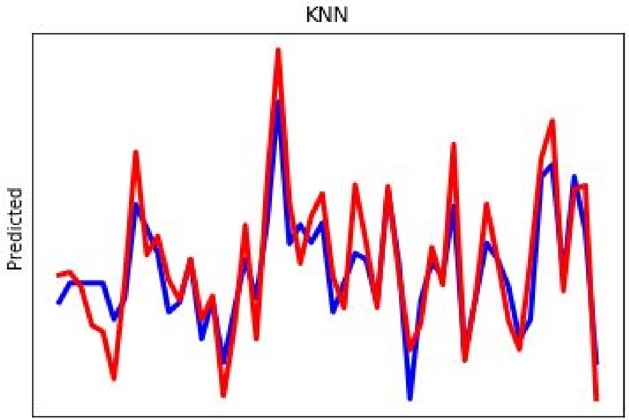
Hand-Picked Features Model

1. Use Time Series to predict value of features in the future.
2. Train a model with the highest-correlated features.
3. Apply coefficients of tuned algorithm as weights to projected features.

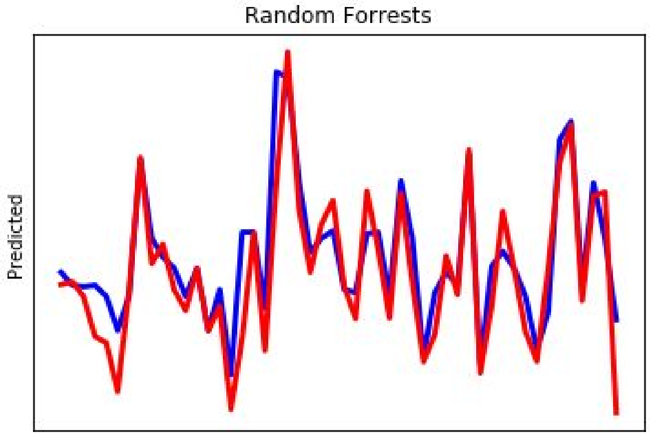
Logistic Regression



KNN



Random Forrests



Conclusions

3-Year Prediction

Time series model performed superbly.

Will investigate trained model with projected feature data next.

5-Year Prediction

Time series model did not perform great, but tuned model with time series predicted feature data performed much better.

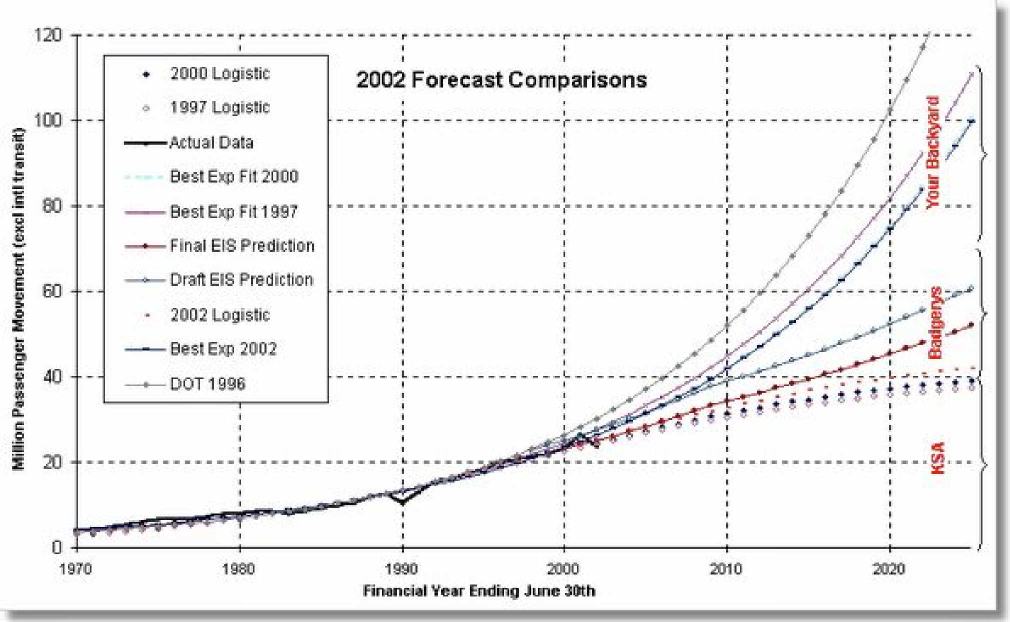
Conclusions

So many factors:

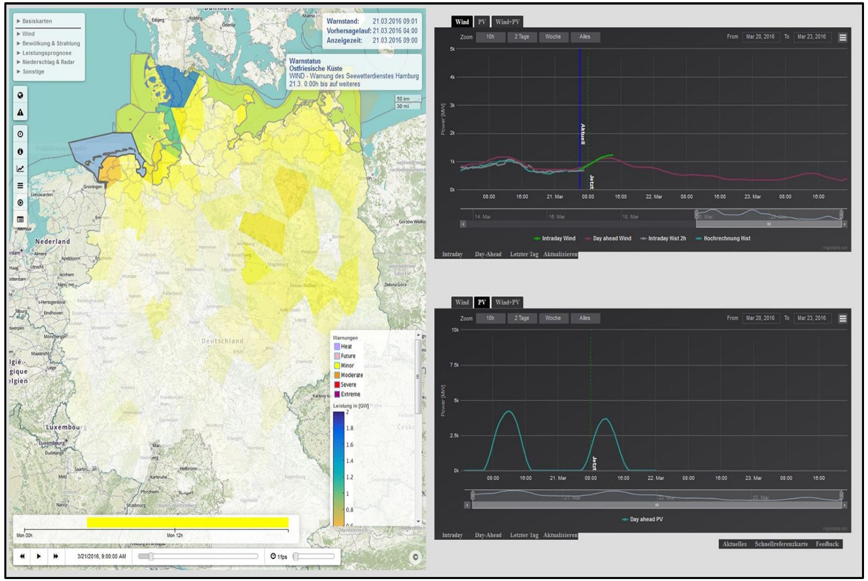
- Political Climate
- Economic Climate
- Corporate Decisions
- Weather and Climate Patterns (next week at best)
- Usage Patterns (next week at best)
- OPEC
- Foreign Policy

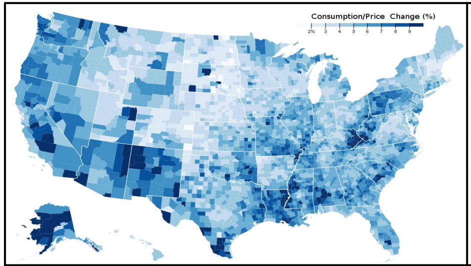
One really hard aspect of this problem (and one of the main reasons nobody has really solved this problem) is that these factors pretty much cannot be predicted, and, unfortunately, they affect prices pretty drastically.

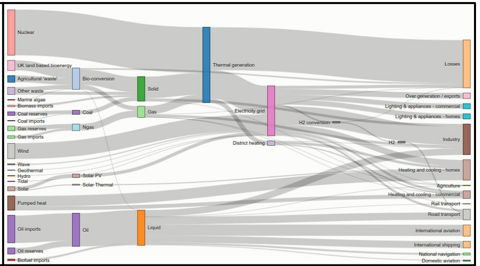
Conclusions

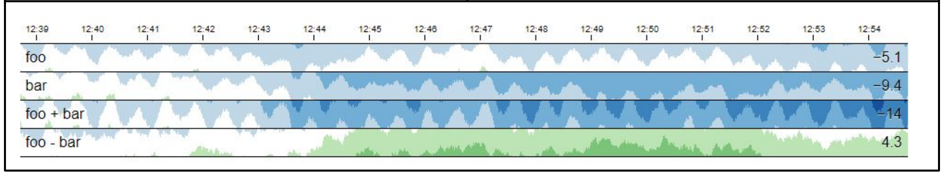


Perturbance Flexible + Confidence Intervals









Terminal

Visualization

Technical Components

Interface
Interactive Interface
Online Learning
Perturbance-Flexible
Front End
User Domain
Knowledge Input
Real Time Data
Visualization

Prediction
Linear Regression
Time Series
Random Forest
Regressor
Vector autoregressive
model (ARIMA)
Linear Dynamic Systems
Gaussian State Space
Models

Techniques
Synthetic Data
Augmentation
Automated Database
Query and processing
(Pipelining)
Online Learning
Cross Validation

Next Steps

Try a new fuzzy model for more precise time series prediction (maybe tweak Random Forest)

Provide perturbance flexibility with confidence intervals (and a way to quantify perturbations).

Develop User Interface, reason about user domain knowledge input, look into online learning with constant feedback into pipeline, and find more data.

Reference

[1] A. Liaw and M. Wiener (2002). **Classification and Regression by randomForest**. R News 2(3), 18--22.

[2] Jason Brownlee (January, 2th, 2017) **Autoregression Models for Time Series Forecasting With Python** https://machinelearningmastery.com/autoregression-models-time-series-forecasting-python/

[3] Statsmodels http://www.statsmodels.org/stable/index.html

Thanks!

Any questions ?