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.

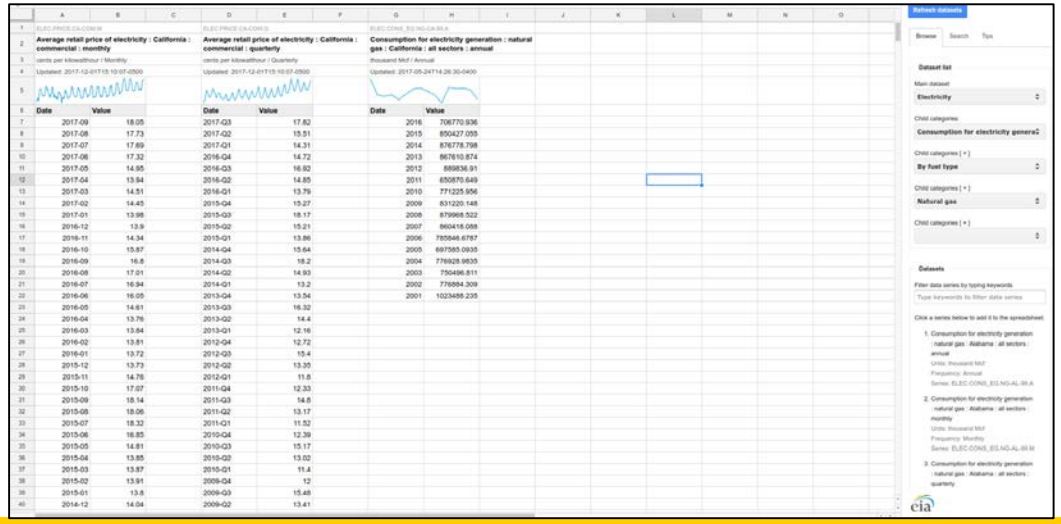
Data Collection and Cleaning

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.



Correlation Analysis

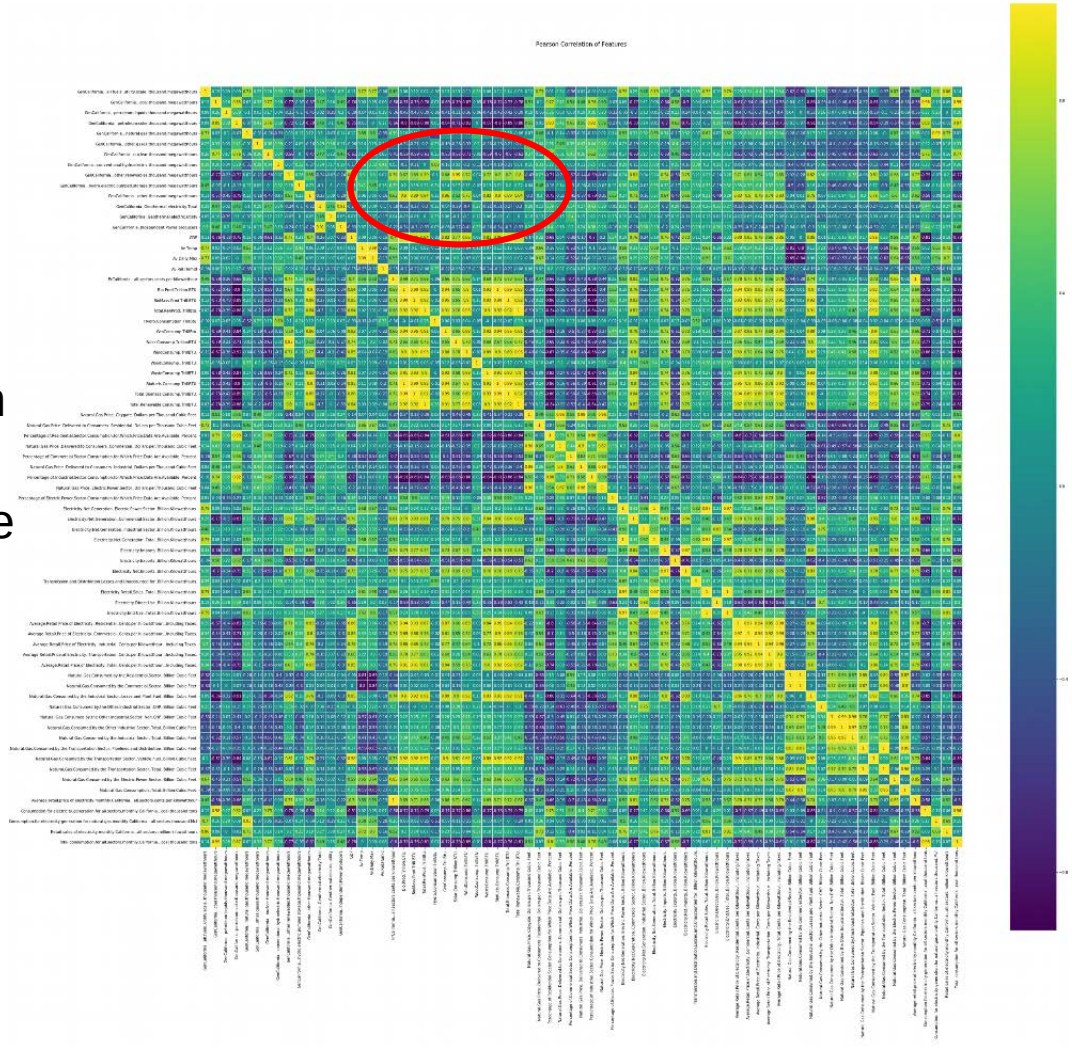
For feature selection


Pearson Correlation of Features

GDP

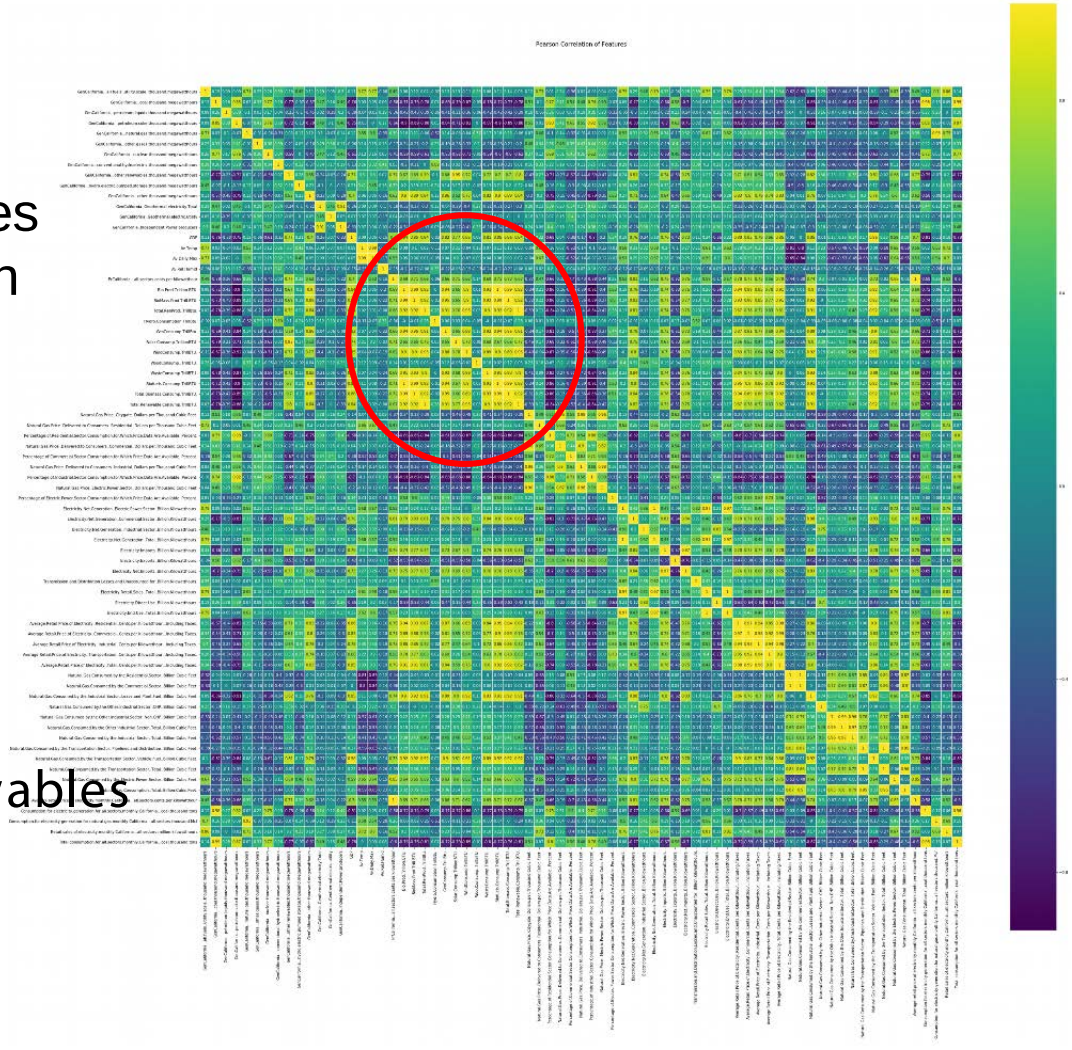
Renewables Consumption

...and Average Retail Electricity Price



Renewables Production

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



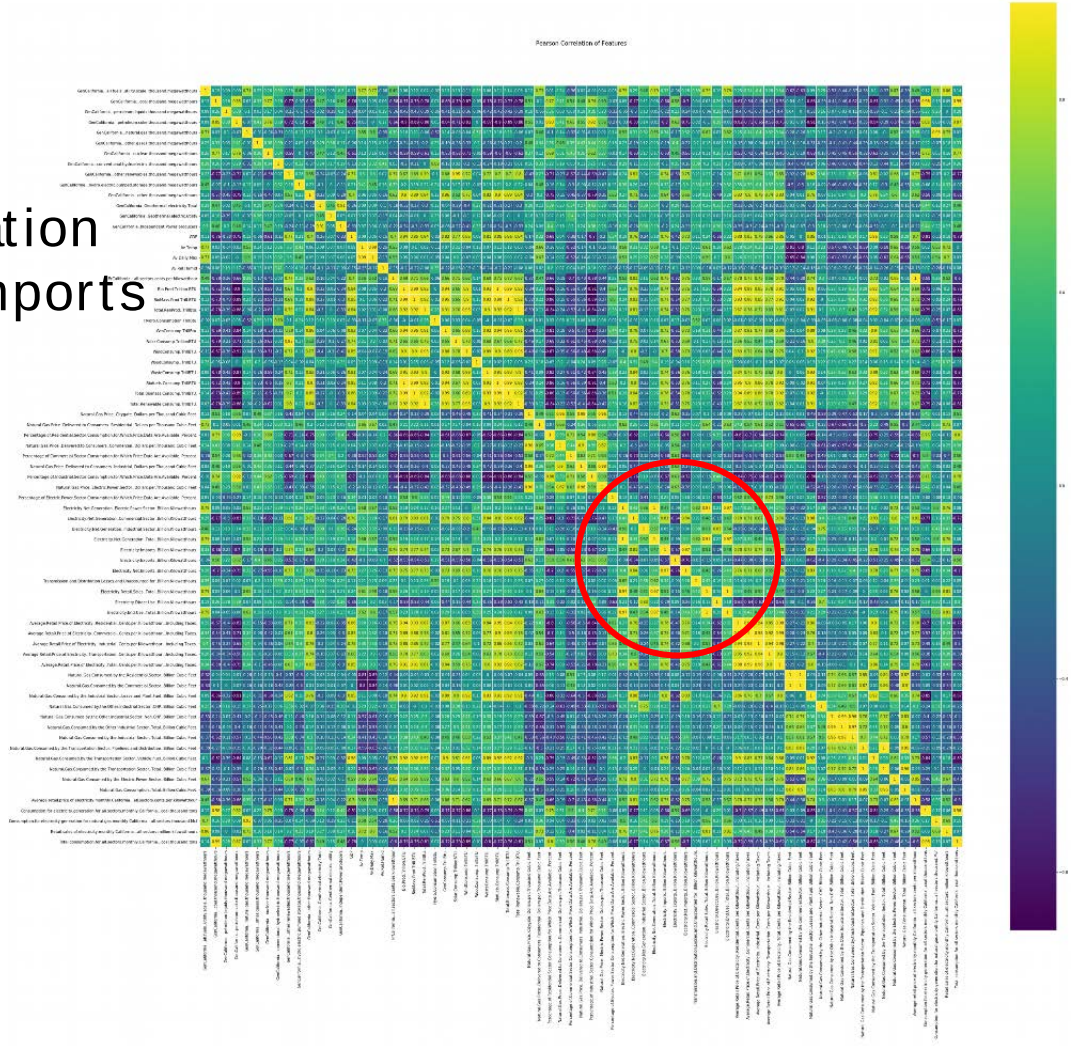
Natural Gas Price

Electricity Price

Different Sectors

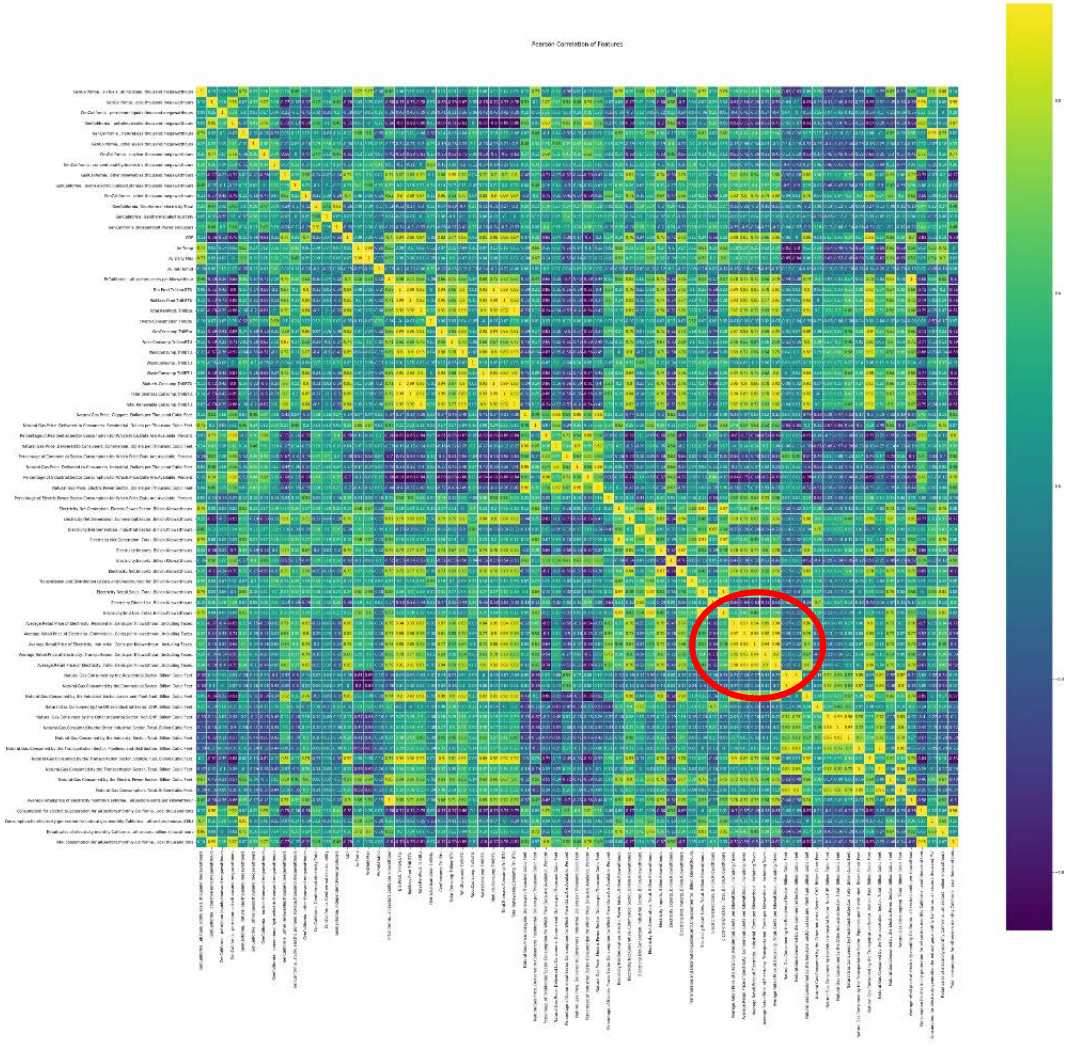
- Residential
- Commercial
- Inudstrial

Net Generation Exports/ Imports in different sectors

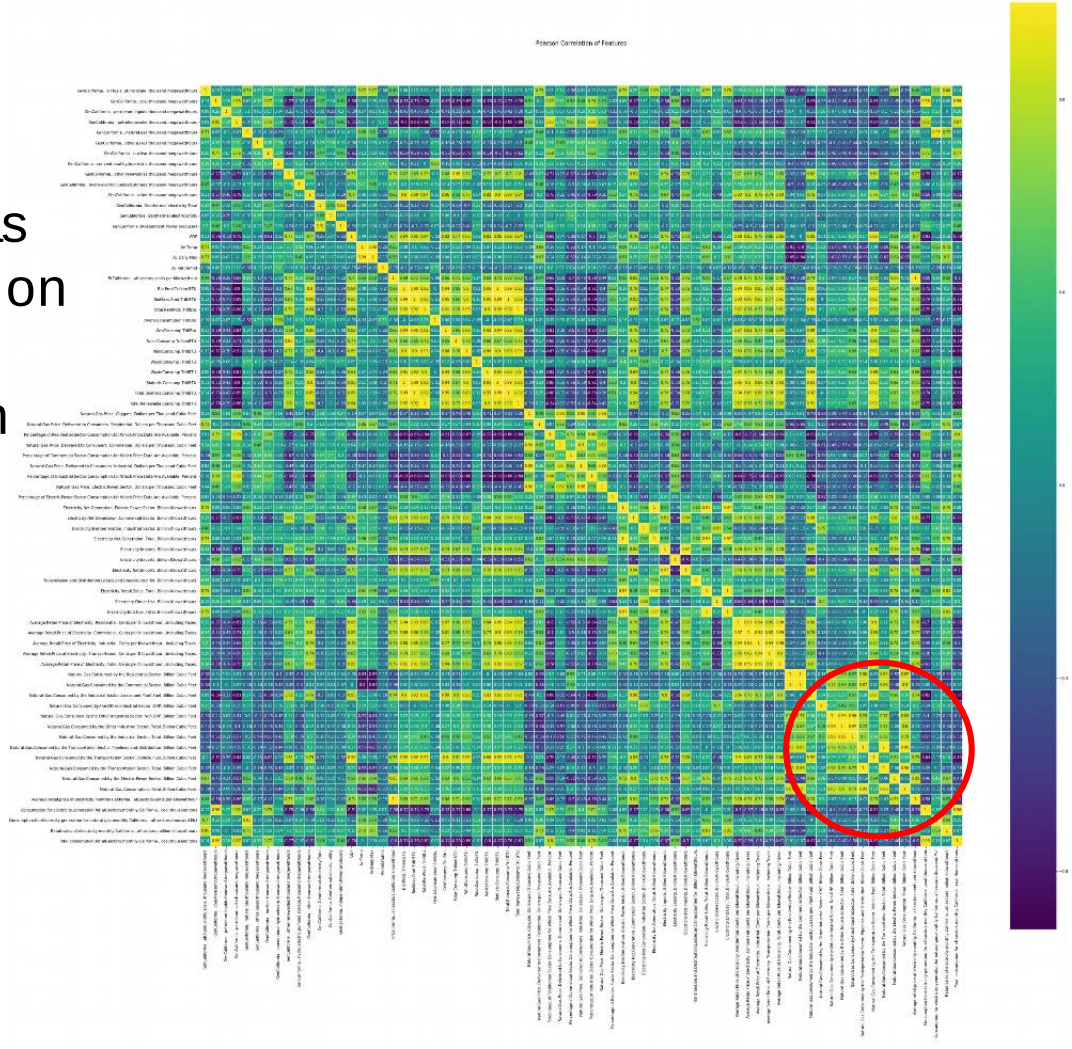


Average Electricity Retail Price

Different Sectors/ Units



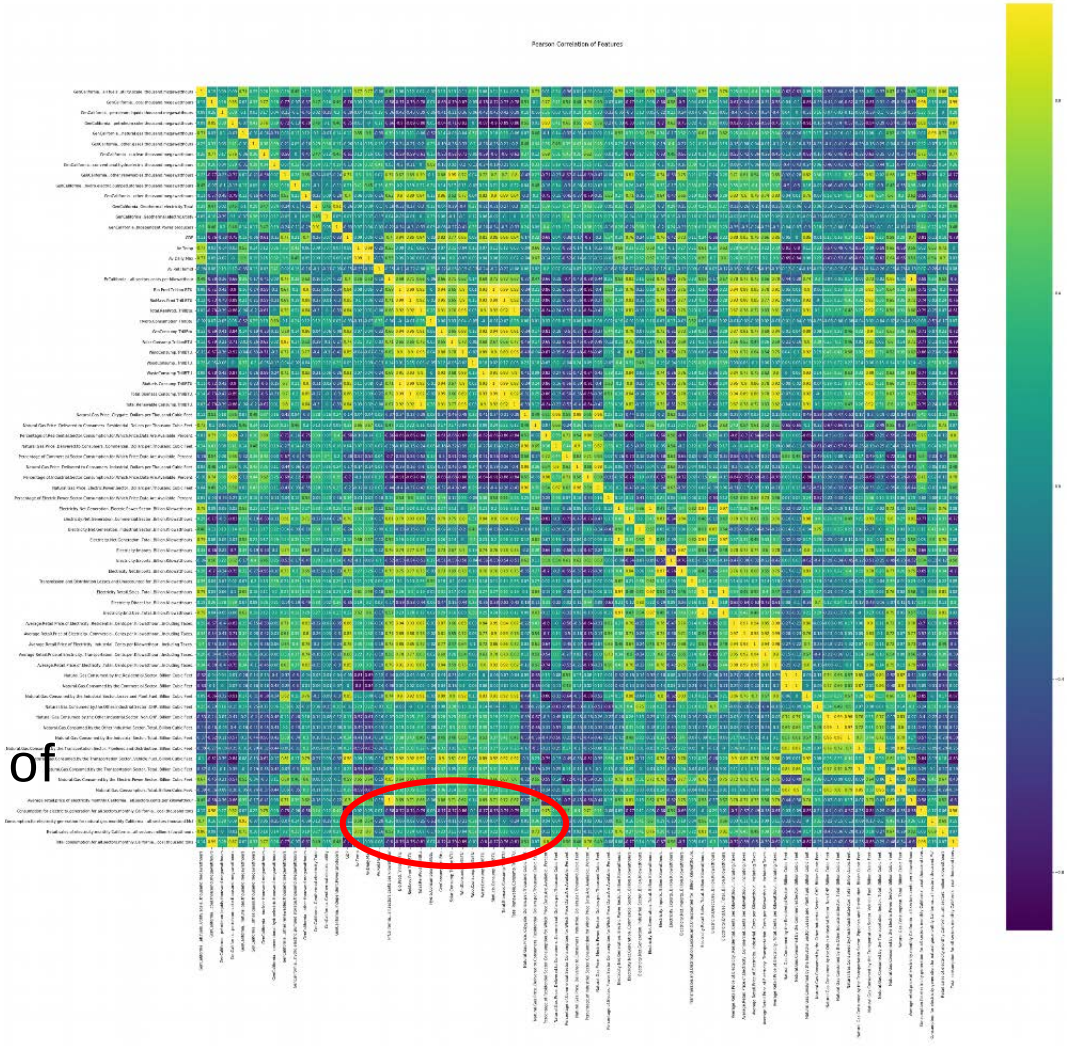
Natural Gas Consumption and Production by sector.



Petroleum Coke and Coal Generation

…And

Average Monthly Retail Price of Electricity



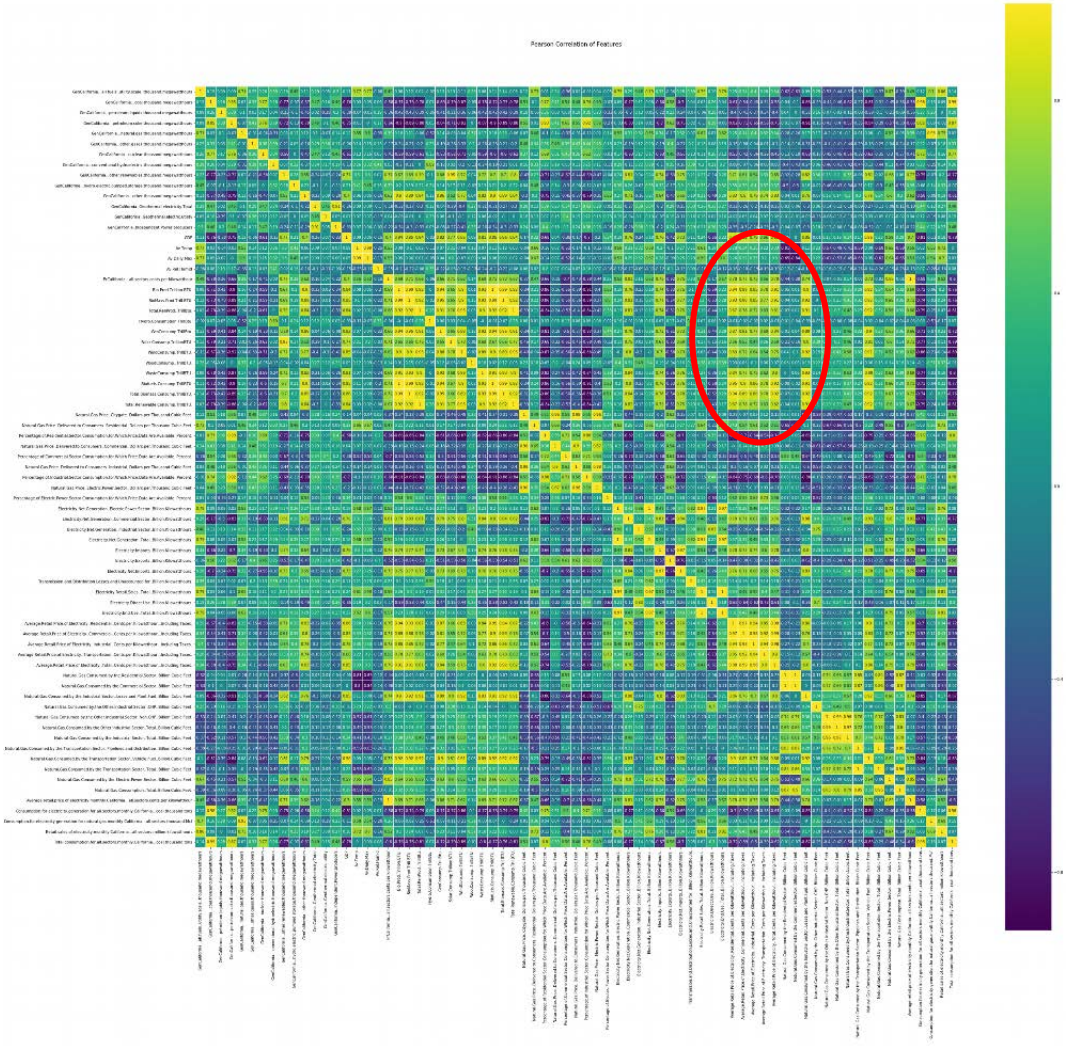
Renewables Production and Consumption

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

And...

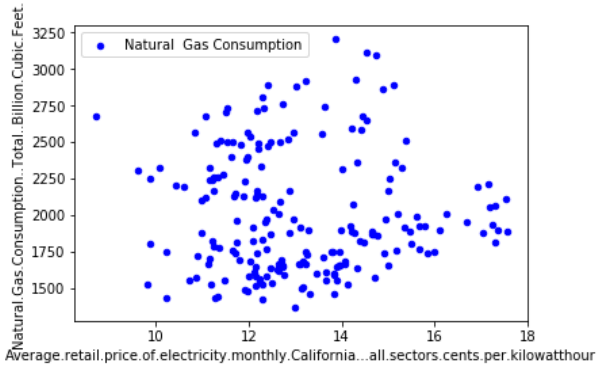
- Average Monthly Retail Electricity Price in different sectors.

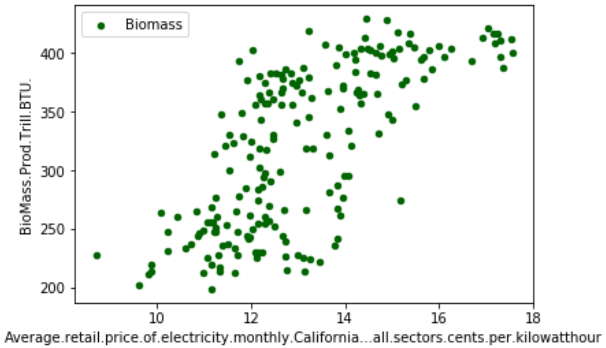
- Electricity Import Price

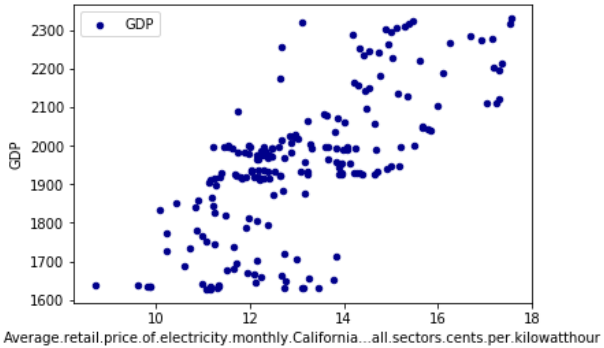


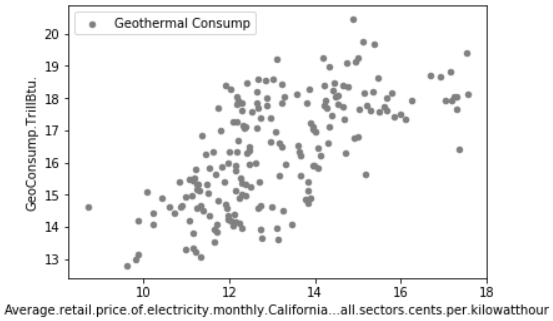
Correlations Summary (Features)

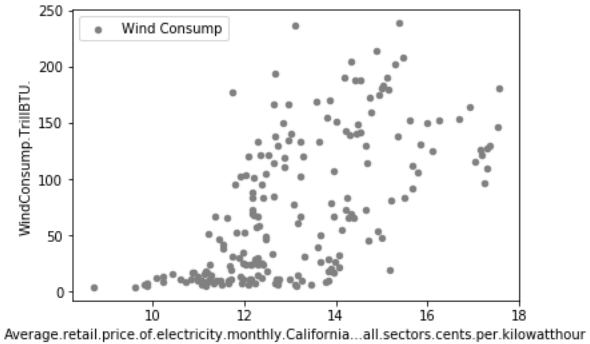
- GDP (.7)
- Natural Gas Consumed by Industrial Sector (.74)
- Natural Gas Consumed by Electric Power Sector (.85)
- Net Electricity Imports (.75)
- Net Electricity Generation (.81)
- Renewables (.77)
  - Solar Consumption (.71)
  - Geothermal Consumption (.65)
  - Wind Consumption (.62)
  - Biofuels Production (.72)

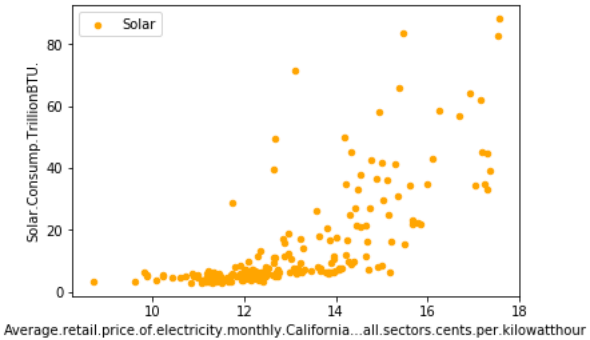

Natural Gas Consumption
Natural.Gas.Consumption...Total..Billion.Cubic.Feet.
Average.retail.price.of.electricity.monthly.California...all.sectors.cents.per.kilowatthour

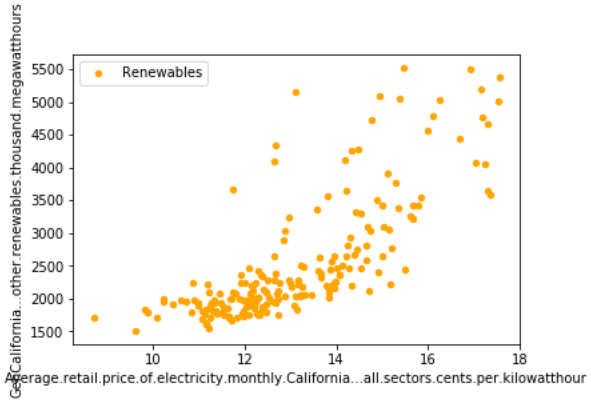

Biomass
BioMass.Prod.Trill.BTU.
Average.retail.price.of.electricity.monthly.California...all.sectors.cents.per.kilowatthour


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

Prediction

Time Series

Prediction with Time Series

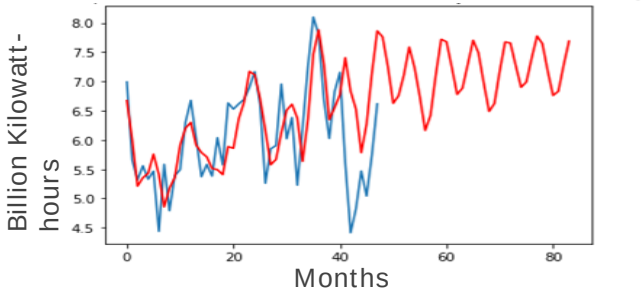
- Time Series Prediction (Autoregressive Integrated Moving Average or ARIMA) for 3 years in advance - train 2000-2014, test 2014-2017, predict 2017-2020.

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

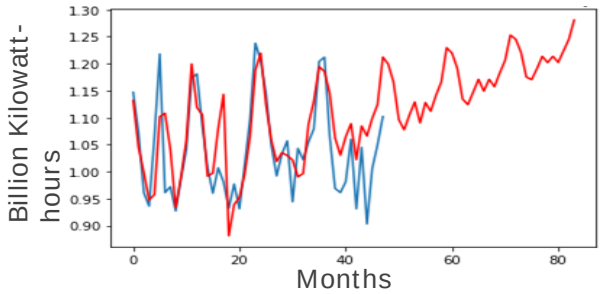
Time Series Results

- 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.

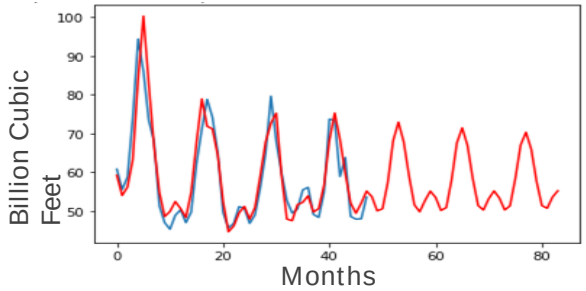
Electricity Imports (3 Years)



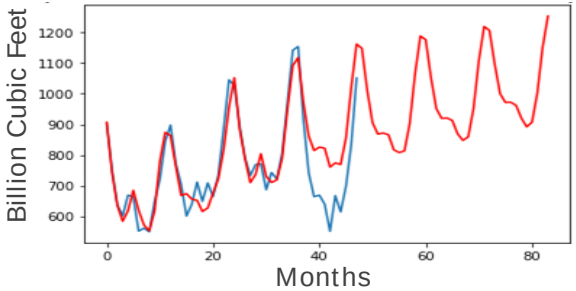
Net Commercial Generation (3 Years)

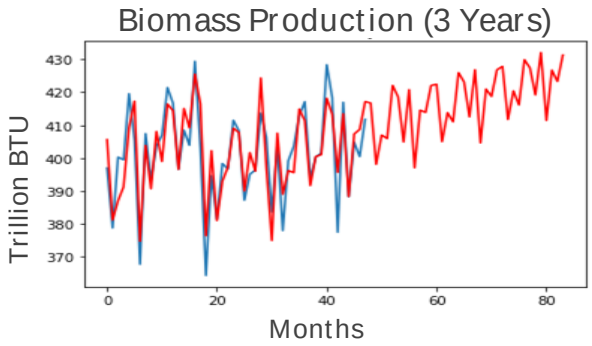


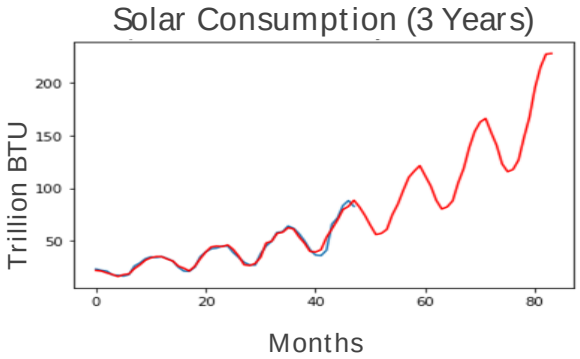
Natural Gas Consumed by Industrial Sector (3 Years)

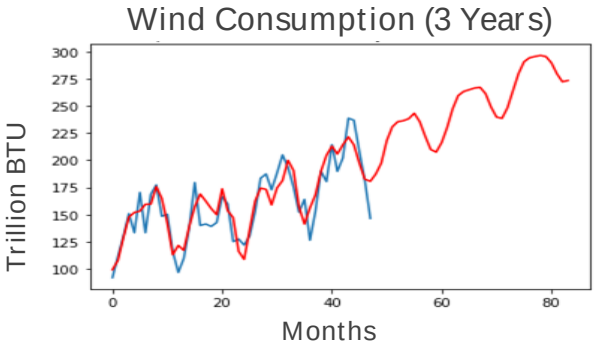


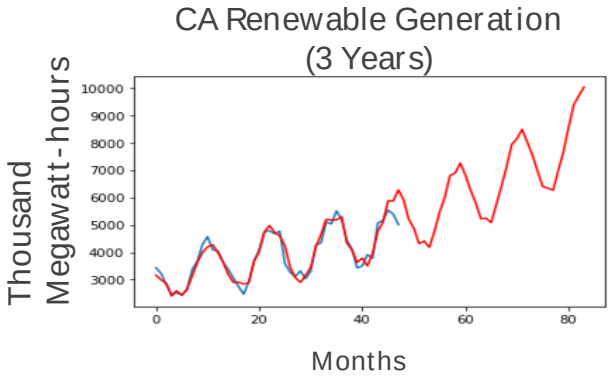
Natural Gas Consumed by Electric Power Sector (3 Years)



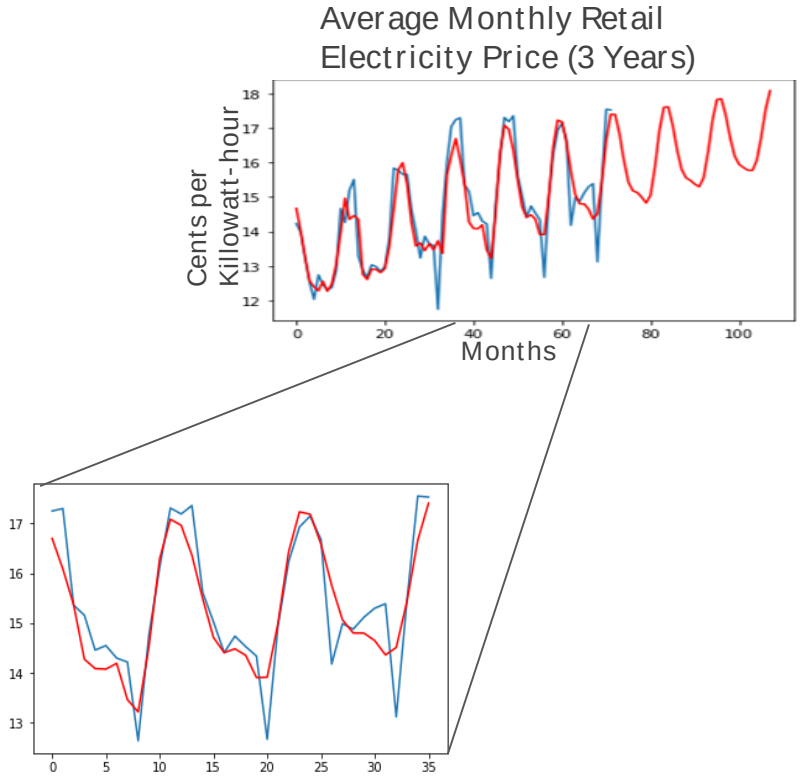

Biomass Production (3 Years)
Trillion BTU
Months

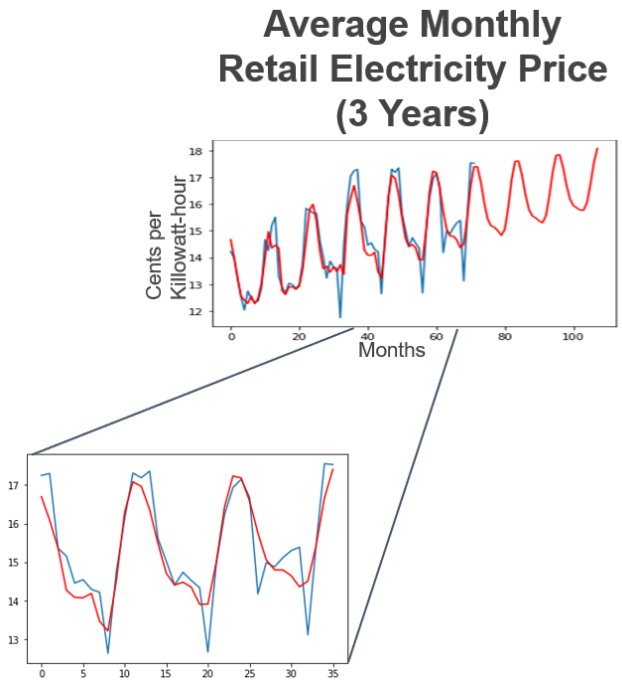

Solar Consumption (3 Years)
Trillion BTU
Months

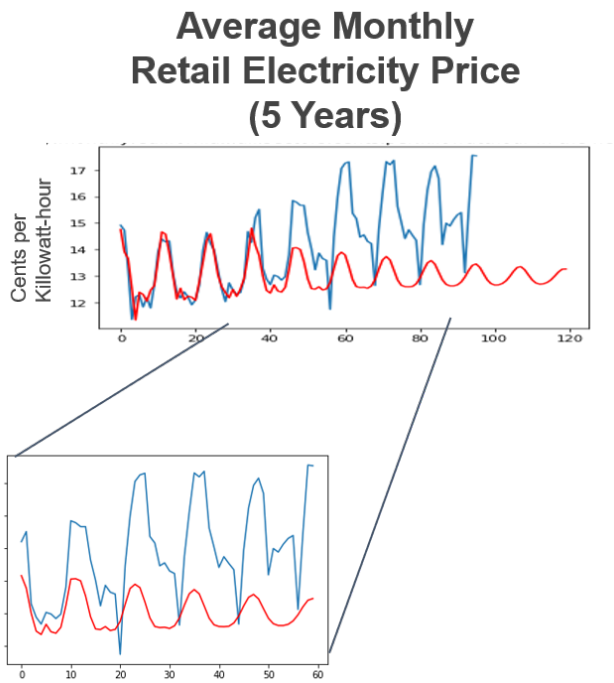

Wind Consumption (3 Years)
Trillion BTU
Months


CA Renewable Generation
(3 Years)
Thousand
Megawatt-hours
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
```







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

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

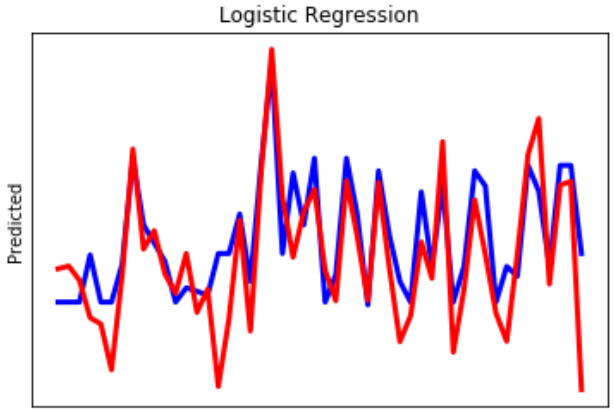
Hand-Picked Features Model

Use Time Series to predict value of features in the future.

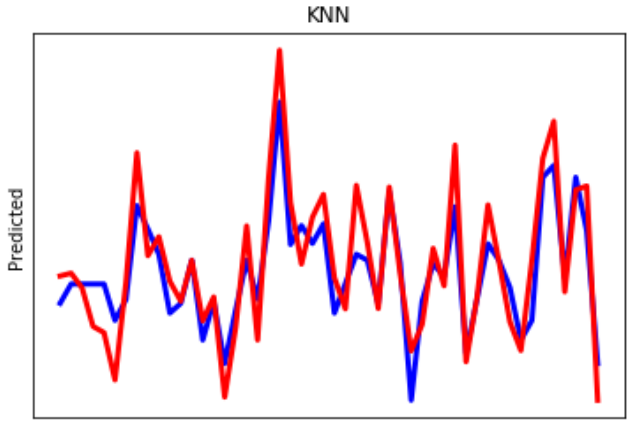
Train a model with the highest-correlated features.

Apply coefficients of tuned algorithm as weights to projected features.

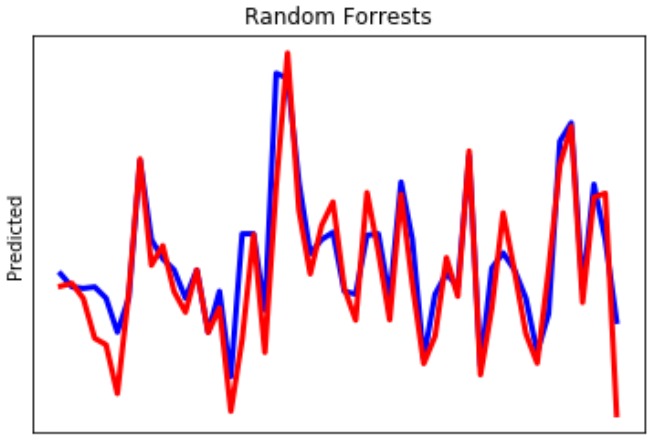
Logistic Regression



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.

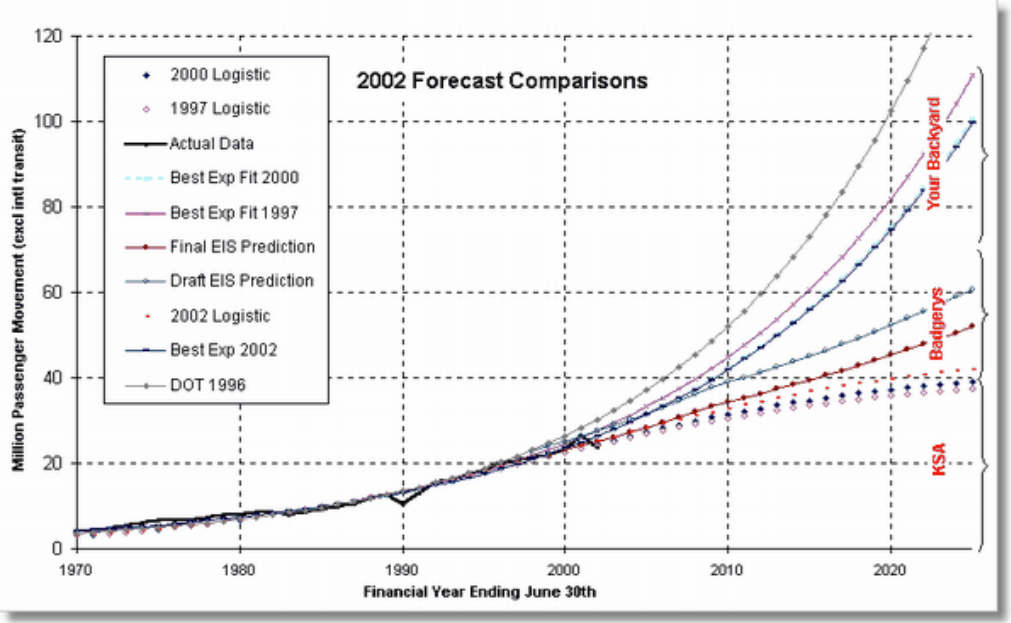
Conclusions

So many factors:

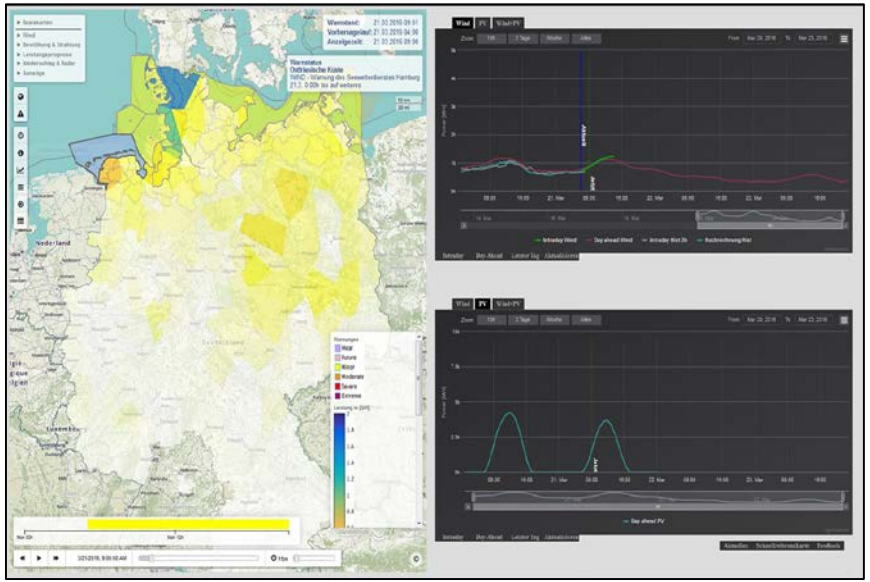
- Political Climate
- Economic Climate
- Corporate Decisions
- Weather and Climate Patterns (next week at best)
- Usage Patterns (next week at best)
- OPEC
- Foreign Policy

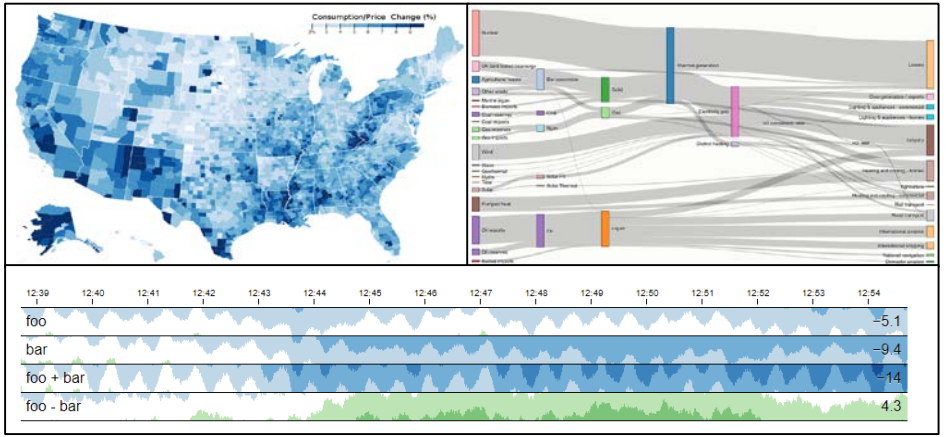
One really hard aspect of this problem (and one of the main reasons nobody has really solved this problem) is that these factors pretty much cannot be predicted, and, unfortunately, they affect prices pretty drastically.

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.

Thanks!

Any questions?