**Energy Consumption Prediction**

**Description**

* This project used a [Kaggle dataset](https://www.kaggle.com/robikscube/hourly-energy-consumption) of AEP’s eastern grid’s hourly energy consumption from December 2004 – December 2018. This was combined with a [dataset](https://www.kaggle.com/selfishgene/historical-hourly-weather-data) that contained hourly temperature information. Using this information, the algorithm is able to predict the energy consumption of any given hour with an R^2 of 0.77.

**Methods**

* After examining a map of AEP’s eastern grid, and seeing which cities were available in the temperature dataset, Indianapolis was elected to be used as the baseline temperature
* A close up of a map

  Description automatically generated
* The temperature dataset only contained info from 2013 – 2017, so this was the time frame used
* The dataset was sliced into 5 different features
  + Hour of the Day
  + Day of the Week
  + Month
  + Year
  + Temperature
* Initially a linear regression model was attempted, but it then became clear most of these features had “bends” and couldn’t be accurately modeled linearly
* I found a python library [pyGAM](https://pygam.readthedocs.io/en/latest/notebooks/tour_of_pygam.html) that allows one to build custom models based on different qualities of the features.
* In this case a spline regression general additive model was used for our features which does an excellent job of accounting for small twists in the data
  + This essentially applies a different function to each feature, rather than a linear or polynomial slope.
* A random 80% of the data was used to train the model, and the other 20% was used to test the model’s predictive power

**Visuals**

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

**Statistical Outputs**

A screenshot of a cell phone

Description automatically generated

As you can see, each feature has a P-value under 0.001. This isn’t surprising when you see how well the generated formula fits the features.

**Sources**

The initial data was pulled from [Rob Mulla’s Kaggle](https://www.kaggle.com/robikscube) as well as [David Beniaguev’s page](https://www.kaggle.com/selfishgene). Also, pyGAM has done extensive work with generalized additive models, which you can read more about [here.](https://pygam.readthedocs.io/en/latest/notebooks/tour_of_pygam.html)