London Energy and Weather

Scott Wigle

Goal

- Predict energy use based on weather and block in the London area. I will do an initial analysis on a few of the metrics to get averages and see if there is correlation between metrics.

Data

- The data in the set is electrical metering data from blocks located in London. Each block is subdivided by the meters located on the block. The energy data has both daily and half hour incremented data for the years 2011 to 2014. This energy data has the the mean, median, min, max, and standard deviation, see Figure 1. The weather data is daily weather reports from 2011-2014 with many parameters, see Figure 2. These two sets will be merged together for analysis and building of regression model.

```
smw@Artanis-ub:~/Documents/galvanize/capstone_zone/capstone_1$ ipython
Python 3.6.5 |Anaconda, Inc.| (default, Apr 29 2018, 16:14:56)
Type 'copyright', 'credits' or 'license' for more information
IPython 6.4.0 -- An enhanced Interactive Python. Type '?' for help.

In [1]: run London_meters.py

In [2]: df.head()
Out[2]:
      LCLid         day  energy_median  energy_mean  energy_max  energy_count  energy_std  energy_sum  energy_min
0  MAC000002  2012-10-12         0.1385     0.154304       0.886            46    0.196034       7.098       0.000
1  MAC000002  2012-10-13         0.1800     0.230979       0.933            48    0.192329      11.087       0.076
2  MAC000002  2012-10-14         0.1580     0.275479       1.085            48    0.274647      13.223       0.070
3  MAC000002  2012-10-15         0.1310     0.213688       1.164            48    0.224483      10.257       0.070
4  MAC000002  2012-10-16         0.1450     0.203521       0.991            48    0.184115       9.769       0.087
```

Figure 1: Daily energy data for London block 1.

```
In [4]: dfweather.head()
Out[4]:
   temperatureMax   temperatureMaxTime  windBearing                icon  dewPoint    ...      temperatureLowTime  a
0           11.96  2011-11-11 23:00:00          123                 fog      9.40    ...     2011-11-11 19:00:00
1            8.59  2011-12-11 14:00:00          198   partly-cloudy-day      4.49    ...     2011-12-12 07:00:00
2           10.33  2011-12-27 02:00:00          225   partly-cloudy-day      5.47    ...     2011-12-27 23:00:00
3            8.07  2011-12-02 23:00:00          232                wind      3.69    ...     2011-12-02 19:00:00
4            8.22  2011-12-24 23:00:00          252  partly-cloudy-night      2.79    ...     2011-12-24 19:00:00

[5 rows x 32 columns]

In [5]: dfweather.columns
Out[5]:
Index(['temperatureMax', 'temperatureMaxTime', 'windBearing', 'icon',
       'dewPoint', 'temperatureMinTime', 'cloudCover', 'windSpeed', 'pressure',
       'apparentTemperatureMinTime', 'apparentTemperatureHigh', 'precipType',
       'visibility', 'humidity', 'apparentTemperatureHighTime',
       'apparentTemperatureLow', 'apparentTemperatureMax', 'uvIndex', 'time',
       'sunsetTime', 'temperatureLow', 'temperatureMin', 'temperatureHigh',
       'sunriseTime', 'temperatureHighTime', 'uvIndexTime', 'summary',
       'temperatureLowTime', 'apparentTemperatureMin',
       'apparentTemperatureMaxTime', 'apparentTemperatureLowTime',
       'moonPhase'],
      dtype='object')
```

Figure 2: Daily weather for London.

MVP

- Create a linear regression using techniques from class to predict average energy. I want to know if it is possible to build a good model to predict the energy based off weather information. I will also be curious to see if some blocks use more energy than others. For MVP I will limit the model to only a few blocks depending on compute time. I will use all blocks if it is not to computationally heavy.

MVP +

- Use a feature ranking tool such as sklearn.feature_selection.RFE() to fid the best overall parameters to use in my model.
- Make prediction based off the meters inside each of the blocks. To do this it would sectionize out the blocks into smaller sections and increase complexity of coding and model.
- Map each block to type of block, residential, industrial, commercial.
- See if daily swings in energy use are similar on all days.

MVP ++

- Add prediction based off previous days energy data from other blocks and itself.
- Add housing data like income, age, kids in home and other fun factors that could contribute.
- I would also like to add in the estimated cost to the meter user based off the predications.