Electric Load Profiling for Apartments

Since 2010 many electric companies around the United States have begun to install smart meters on each location they service. Prior to smart meter installation companies would send an employee to gather usage information from meters at each service address. Often times the companies would only take an reading every third month and calculate usage estimates for the remaining months. This resulted in inaccurate billing statements, which was a source of frustration for many consumers. With the installation of smart meters electric companies are able to track energy consumption on an hourly basis resulting in precise billing statements. Additionally, utility companies now have access to a wealth of data that can provide insights on how to better manage their resources, such as the ability to better predict how much energy needs to be produce to meet the demand of a specific area.

To make these predictions, energy companies must use the data obtained to make a load profile. Load profiles can encompass as little as a single customer upto a entire power plant. Through analyzing these profiles at the different levels the company will benefit from improved energy production forecast that will minimize overhead cost, result in a smaller carbon footprint and ultimately credit goodwill with the customers through bill reduction.

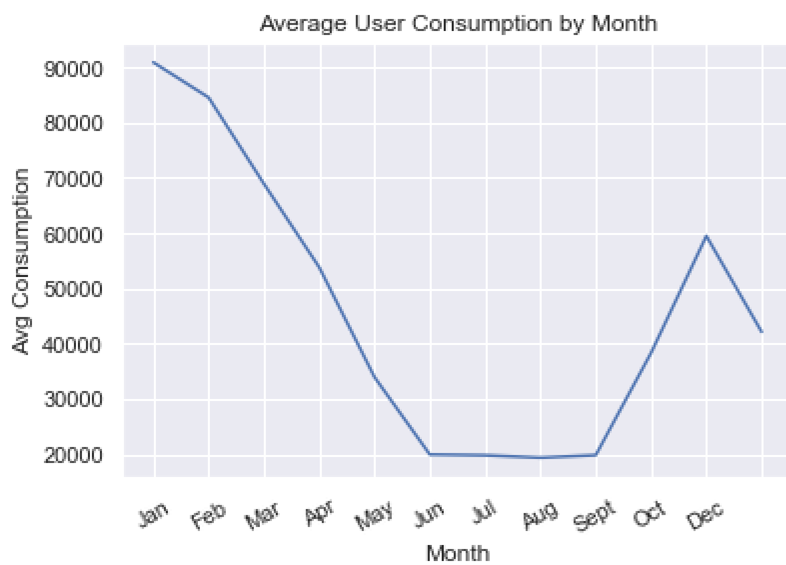
As a proof of concept am using data obtained from the University of Massachusetts' Smart* Data Set for Sustainability, which is available at http://traces.cs.umass.edu/index.php/Smart/Smart. The dataset contains energy usage from 114 single-family apartments collected over a two year period in 15 minute intervals.

For the scope of my project I am using data collected from January 1, 2016 to December 31, 2016. I downloaded both the apartment and weather data from the website. Upon unzipping the files I created a Jupyter notebook where I imported the pandas, datetime and glob modules.

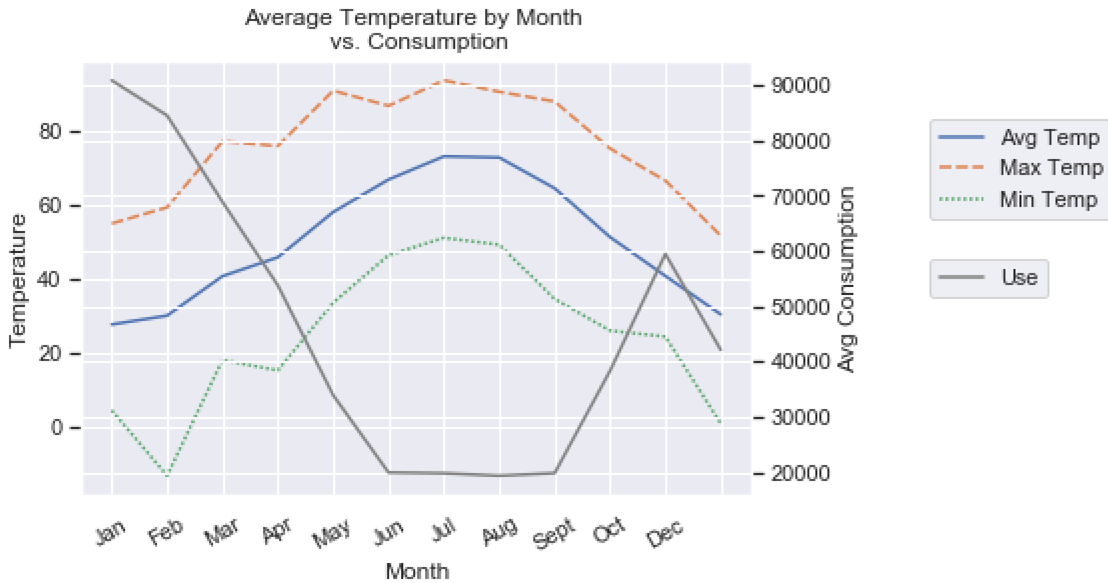
I used the glob module to obtain a list of the apartment data from 2015. I then looped through these files reading in each file, which contain reads collected every 15 minutes for the year, into a pandas dataframe. The first column contains a date and time stamp for each read. I then converted this column to a datetime field and extracted the Date using the to_date functions of pandas. I then used the groupby function to add all readings obtained in a day together. From here I transposed the frame so the dates became columns. I then concatenated the apartments into one dataframe for all apartment data. I then used these same principles to group the readings together by month.

Upon examination of this dataframe I discovered all 114 apartments had a few days of missing readings. The missing reads were all from the beginning of the year leading me to believe the meters where not all installed by January 1, 2015. There were also 3 days at the end of the year with missing readings. Since the number of consecutive missing days was small there is a high probability that the missing days had similar usage to the days surrounding it. Thus, I used a backfill method to fill in missing values in January and a forward fill method to fill in missing

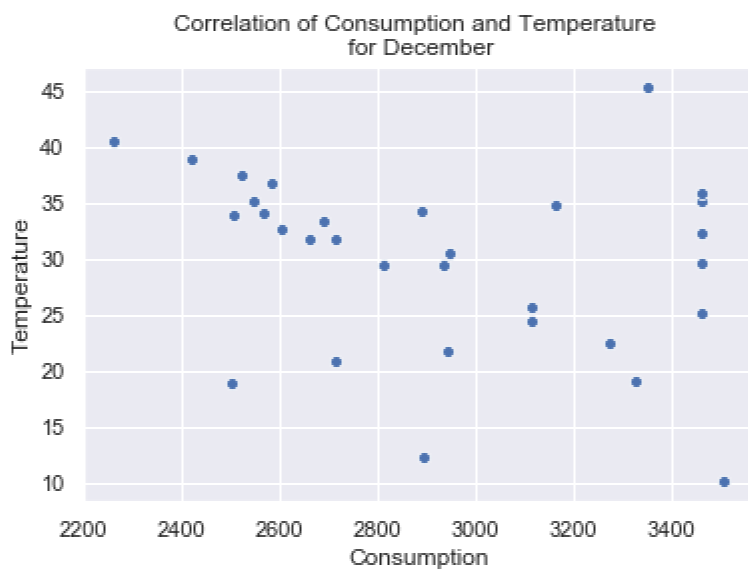
To determine the amount of energy the company needs to produce we first need to understand how much energy the consumers are demanding. By graphing the monthly average consumption for the year.



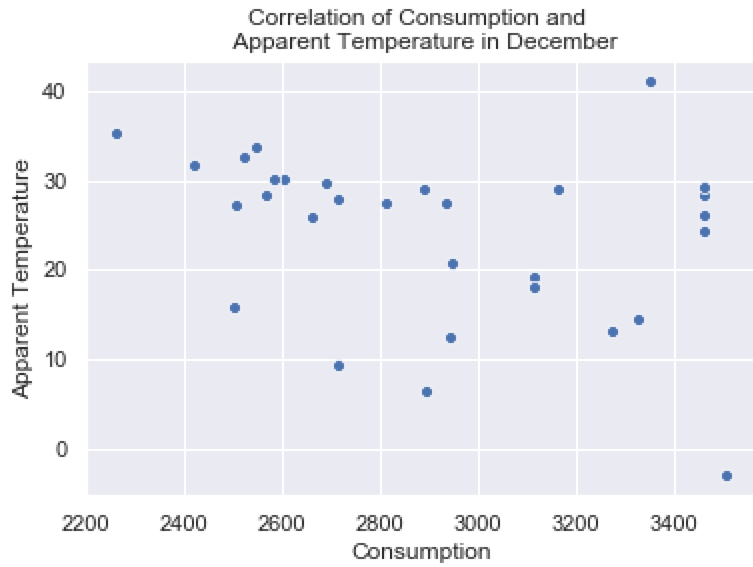
The graph follows an expected pattern for North American: more energy usage in the colder months with one exception. December shows a drop in consumption relative to October. Since temperatures tend to fluctuate as seasons change it is quite possible the drop is due to an unexpected weather pattern. However, by graphically comparing consumption and weather (as seen below) this theory is shown not to be valid.



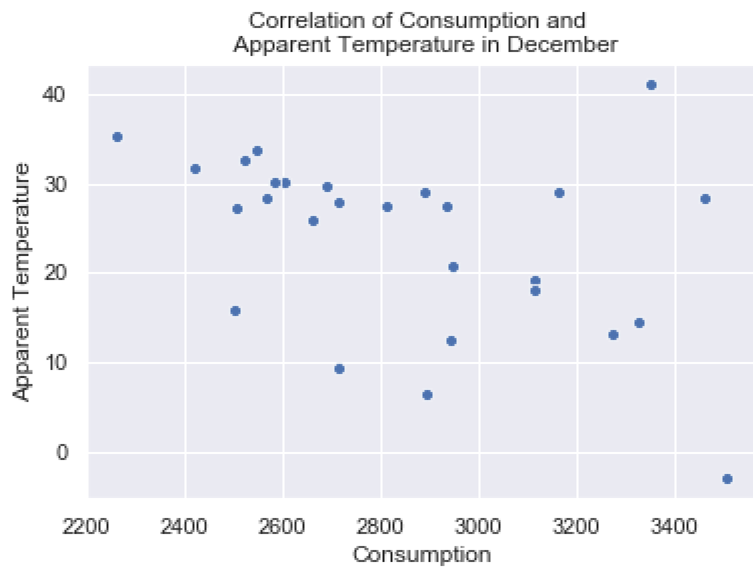
Another explanation for the consumption pattern is that there are a few consumers that are more resilient to cold weather and would use less heat. A simple correlation plot of the two variables in the month of December shows that while there are outliers some use more energy and some use less. It is difficult to decipher the correlation.



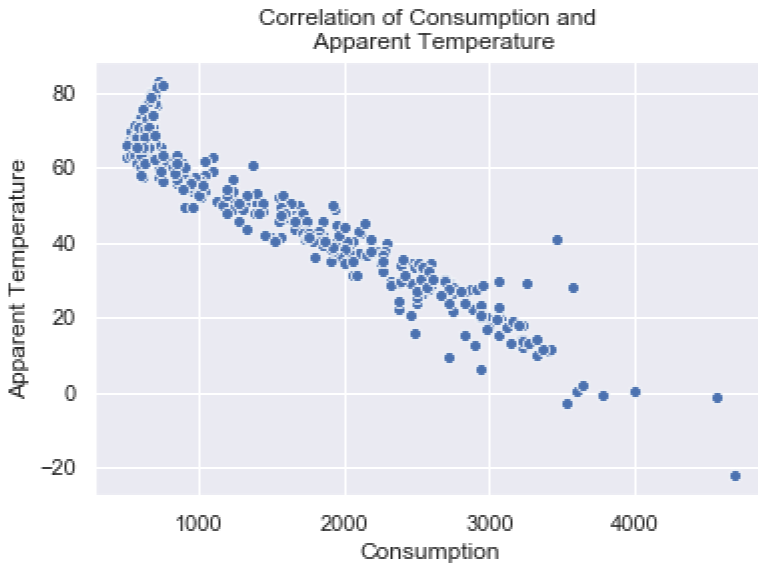
Given that temperature does not seem to show the complete picture of the data we can turn to the apparent temperature for more insight. When graphing the correlation of this to use we find the following:



The correlation appears to be greater between apparent temperature and actual temperature. There are a number of points that use over 3,400 watts of energy, upon exploring further it was found that a bias was introduced by filling the data. The last three days when graphing other days in December we find:



To verify the correlation of apparent temperature to consumption hold true throughout the year we can graph all points.



The results seem to point to a more promising predictor of consumption than actual temperature. A pearson correlation test reveals a the correlation of consumption and temperature is -0.9526 with a p-value of 8.6482 x 10^{-191}. While the correlation of consumption and apparent temperature is -0.9580 with a p-value of 5.2256 x 10^{-200}.

The next step in the project is to use machine learning to find ways to effectively predict the minimum amount of energy the company needs to produce to serve its customers.