PJM Hourly Energy Consumption Exploratory Data Analysis

Moritz Kniebel
moritz.kniebel@student.uni-tuebingen.de
Matr.: 4059837

Abstract

Energy consumption data can be analyzed under numerous hypotheses. This report takes a dataset, which gives us limited information together with the energy consumption. By using analysis and visualizations, this report answers the question if we consume more energy in the climatically more extreme seasons summer and winter?

1 Introduction

Everyday, we strongly rely on the supply of energy. Its of the utmost importance for our economy as well as for our private life. From this unique dependency follows, that the energy consumption tells a lot about our habits. This report aims to understand patterns in our energy consumption and to attribute them to our behavior. An exploratory analysis of the PJM Hourly Energy Consumption Dataset will provide insights and visualization to understand our energy consumption under seasonal aspects. Intuition may tell us, that our energy consumption is higher in the winter and lower in the summer, as our lives have shifted more outside. However, this shift could also increase our energy consumption as it entails the use of other electronics. The question arises whether our energy consumption is higher in the climatically more extreme summer and winter months? The analysis shows that this is the case. Our behaviors in the summertime are surprisingly expensive in terms of energy consumption.

2 Exploratory Analysis

This section describes how the data analysis is carried out. To get a good intuition about the data, the analysis starts with the exploration of the dataset. After that, the methodology and the actual data analysis is presented with its results.
The source code of the analysis can be found under https://github.com/moreezee/DataLitTermProject.

2.1 Data Exploration

The PJM Hourly Energy Consumption Dataset contains the hourly energy output of the PJM Interconnection LLC, which is a regional transmission organization (RTO) in the United States. This analysis specifically takes the data of the American Electric Power (AEP) into account. AEP delivers energy to over 5 million citizens in 11 states. The AEP data is presented in two columns, namely *Datetime* (YYYY-MM-DD HH:MM:SS) and *Megawatt energy consumption*. The captured time interval goes from 1st of October 2004 until 3rd of August 2018 and contains 121273 entries, where each row describes the accumulated energy consumption of one hour of a specific date.
To gain intuition about the AEP data, plotting the energy consumption of one day, one week, one month and one year should give first insights and can be seen in Figure 2.1. The plots of one day, one week and one month unveil two patterns in the energy consumption. Everyday, the morning

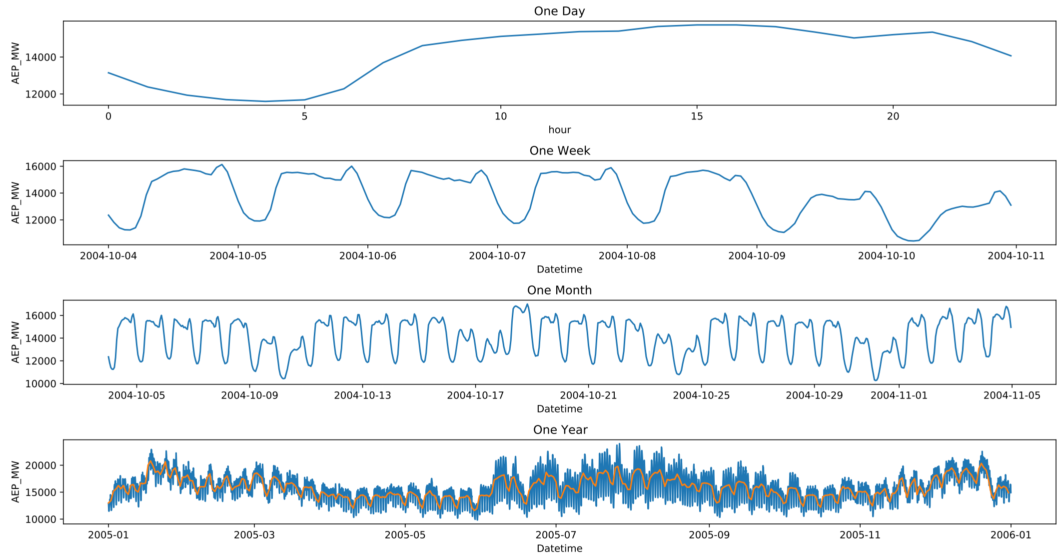

Figure 1: Visualization of the energy consumption of one day, one week, one month and one year, respectively. Patterns in the data can be clearly identified. Everyday, the energy consumption increases in the morning hours, stays relatively consistent throughout the day. In the evening hours, the energy consumption peaks and then decreases during the night time. The visualization of one month shows, that the energy consumption is lower on weekends. Further, the visualization of one year gives an idea, that the energy consumption in higher during the winter and summer months.

hours, the energy consumption begins to increase and stays relatively constant throughout the day until it encounters a peak during the evening hours. Furthermore, the energy consumption is lower on weekends. Thinking about the daily routine of an average citizen, who is resting at night, getting ready in the morning and spending time at home in the evening, the patterns found correspond to this behavior.
The one year plot also shows a rough pattern, as the energy consumption looks like a wave throughout the year with maxima in the months of winter and summer. However, a merely visual analysis is not sufficient to assert with certainty, that the energy consumption is higher in winter and summer.

2.2 Methodology

As mentioned above, it requires further analysis of the AEP data to answer the research question. This section describes the methods which will help to more information out of the data.
The AEP data only consists of the two columns *Datetime* and *Megawatt energy consumption*. As the analysis focuses on seasons, the dataset is extended with additional columns, namely date, year, month, hour, season, spring, summer, fall, winter. These columns are used to create more specific plots, which help to answer the research question visually. Further, a multi-dimensional linear regression will be performed to see how the added columns fall into importance w.r.t. the energy consumption.

2.3 Analysis

To build on the previous analysis, the next step is to take the energy consumption of the individual seasons under further investigation. Plotting the accumulated energy consumption of the whole dataset for each seasons gives a ranking of the seasonal energy consumption, as seen the upper plot in Figure 2.3. This ranking is not yet meaningful enough to answer the research question too concretely. It could be that individual years distort the overall ranking. To eliminate this uncertainty, the consequent step is to analyze a similar ranking, which shows the energy consumption for every whole year in the AEP data. This ranking is shown in the lower plot in Figure 2.3. To strengthen the previous findings, it is of interest which coefficients a multi-dimensional regression assigns to the

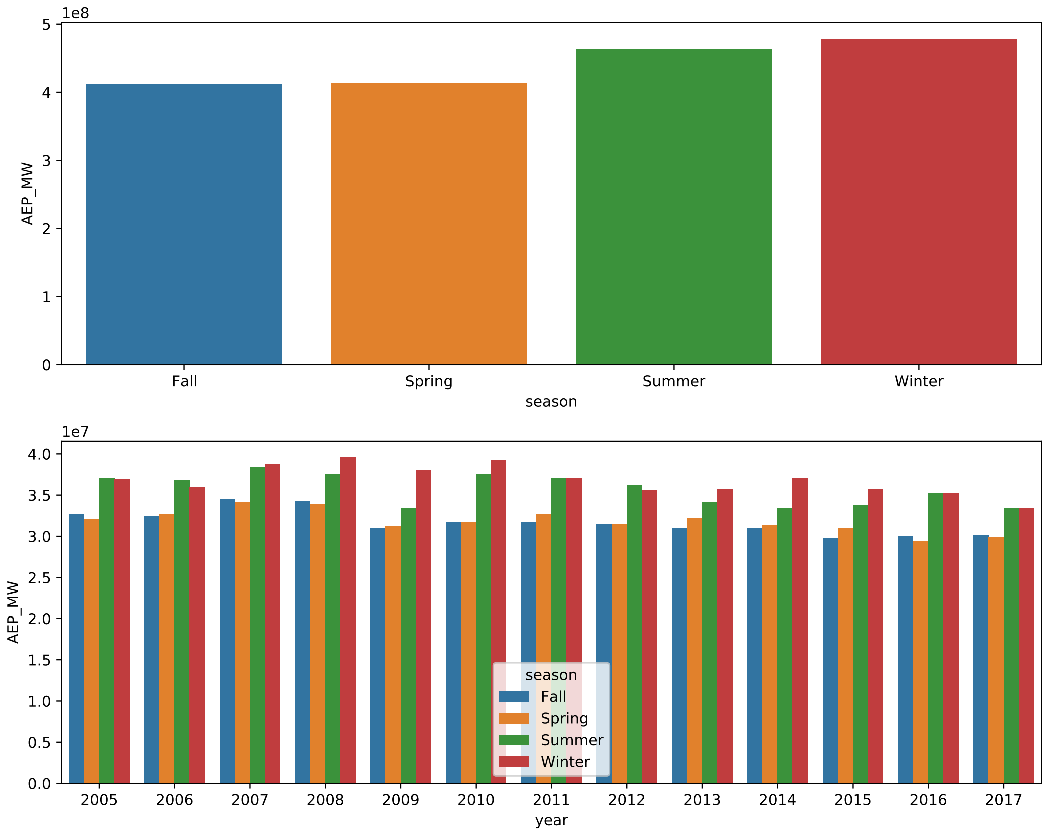

Figure 2: The upper plot shows the energy consumption of every season, accumulated over the whole dataset. This ranking shows, that winter and summer have the highest energy consumption. The lower plot shows the accumulated energy consumption of each season for every whole year in the AEP data. The ranking, that was indicated in the upper plot, can also be applied to the individual years.

individual columns of the dataset. After scaling the data, a ridge regression with factors year, month, hour, spring, summer, fall, winter yields the individual coefficients, shown in Figure 2.3.

2.4 Results

The upper plot in Figure 2.3 shows, that summer and winter register the highest energy consumption w.r.t. the whole dataset. In the lower plot in Figure 2.3, this rankings validity can be confirmed for every whole year in the dataset. Additionally, the regression gives further insights, by yielding coefficients for the factors (Figure 2.3). The values of these factors corresponds to previous findings. The hour of the day is decisive for determining energy consumption. Additionally, the seasons winter and summer have positive coefficients, while spring and fall have negative coefficients. Interpreting these coefficients, one should get higher energy consumption for a later hours of the day and during winter and summer. Likewise, one should get lower energy consumption during spring and fall. The remaining coefficients also give us hints for further analysis. However, this is out of the scope of this report and gives room for future work.
Given these results, the research question can be answered. We consume more energy in the climatically more extreme seasons winter and summer.

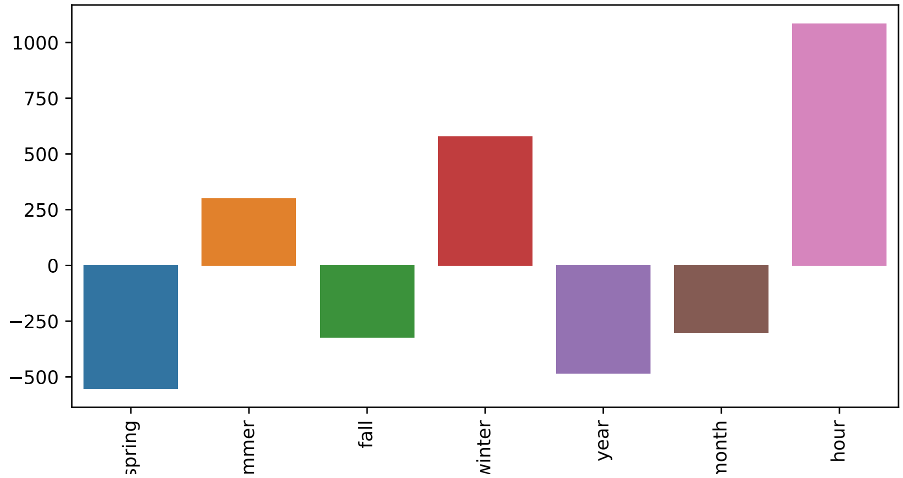

Figure 3: A ridge regression was performed on several factors of the AEP data, namely spring, summer, fall, winter, year, month and hour. The resulting coefficients are visualized in this bar-plot. One can see, that the coefficients for summer and winter are positive, while the ones for spring and fall are negative. Interpreting these coefficients tells us, that the prediction of the energy consumption is higher for summer and winter, and lower for spring and fall. The corresponds to the previous findings.

3 Conclusion

Energy consumption data can be analyzed under numerous hypotheses, as there are different influencing factors. Such factors can be demographic, geographic, economic etc.. In the case of this analysis, the data only gave us information about the energy consumption on specific dates. Nevertheless, those two columns are already enough to perform some in depth analysis of our energy consumption.