Networking Tour of Data Science

Project Proposal

Where to live in the Netherlands in order to reduce electricity bills?

Group 16 - Jelena Simeunovic, Andre Clerc, Jules Afresne, Aurélien Kinet

Abrupt climate changes and pollution due to usage of fossil fuels have resulted in European Union initiative known as Vision 2050. The EU has objective to lead in addressing the global climate change challenge. Transfer from fossil to renewable energy source is just one of its goals. The other one is related to reduction of electricity consumption. In order to create incentive for people to reduce electricity consumption, especially in peak hours, governments will probably introduce different prices, according to energy consumption. Different areas of country will therefore have different price rates for electricity, depending on consumption. One of the question that we will try to answer is: Where should you live in order to reduce your electricity bills and live "greener"?

We will focus on one country (the Netherlands) in order to analyse data and narrow down the solution. Data set that we are going to use is Kaggle's dataset on electricity consumption for past 7 years. We will first create graph where nodes are going to be zip codes, and use time series data from each node as attributes.

In order to explore data, we will first build one graph where the physical distance between nodes is going to be an indication of similarity (we have longitude and latitude of each node). After this we will analyse type of graph as well as degree distribution. Then we will explore how many connected components are there, as well as sparsity patterns and try to find well separated clusters. Afterwards, we will compare several different techniques for dimensionality reduction. We will not use all the nodes that are known, by masking them and "hiding its data" and therefore we will have some "unknown" data consumption at certain nodes. Then, based on smoothness between neighboring nodes, we will try to discover the electricity consumption at that point of unknown data and compare the differences with ground truth.

Then we will build different graph, where measure of similarity will be based on Pearson correlation between data lying on the nodes. Here we can actually create edges between nodes with signals which have similar curves. We expect residential areas to be connected together and on the other hand industrial areas. For both graphs we will use graph filters and evaluate which gives better results, closer to the ground truth. However, there always might be factory close to residential area that can affect our smoothness and we will try to find this anomalies, based on variation of signal. We will perform this using ARMA filtering.

At the end of this part, we will try to understand where residential areas are. Furthermore, we will try to discover where the electricity consumption is the lowest and how to select where to live if we want to save some money and live greener.