DEMAND PROFILE ANALYTICS FOR VIRTUAL POWER PLANT

INTRODUCTION

The electricity sector is changing rapidly and previously one-way market relationships between energy producers and consumers are no longer a single option. The course for decarbonisation, digitalisation and decentralisation has become a new strategic plan for the industry development. More often small and medium consumers of energy seek for greener and more profitable solutions, like solar panels, storage systems. These new market agents are called prosumers — consumers that produce energy.

Electricity is a unique product due to its physical characteristics and is traded on the so-called Day Ahead Markets - for each hour of the next day. Thus, prosumers should keep their eye on volumes they consume and produce, as renewable energy generation is intermittent and from time to time prosumers still need to buy energy from the grid. And what to do during periods where they produce more energy than you can consume or store?

One of the possible local solutions is a creation of Virtual Power Plant (VPP) - an aggregator of prosumers that regulates their relationships with each other and the energy system.

This project is inspired by the 2017 course “Qualitative Methods In Energy Economics” by Sergey Syntulsky.

BUSINESS UNDERSTANDING

Technically, VPP is an entity that optimises energy flows within group and the market given existing distribution system constraints. Energy consumption is uneven with picks during the working hours and bottoms at night. The distribution depends on the production processes of each prosumer type. For example, fridges in warehouses usually work uniformly through the day, on the opposite, office buildings need more energy from 9:00 am till 18:00 and private houses — before 9:00 and after 18:00. At the same time, power generation of solar PV systems directly connected to the level of solar radiation during the day.

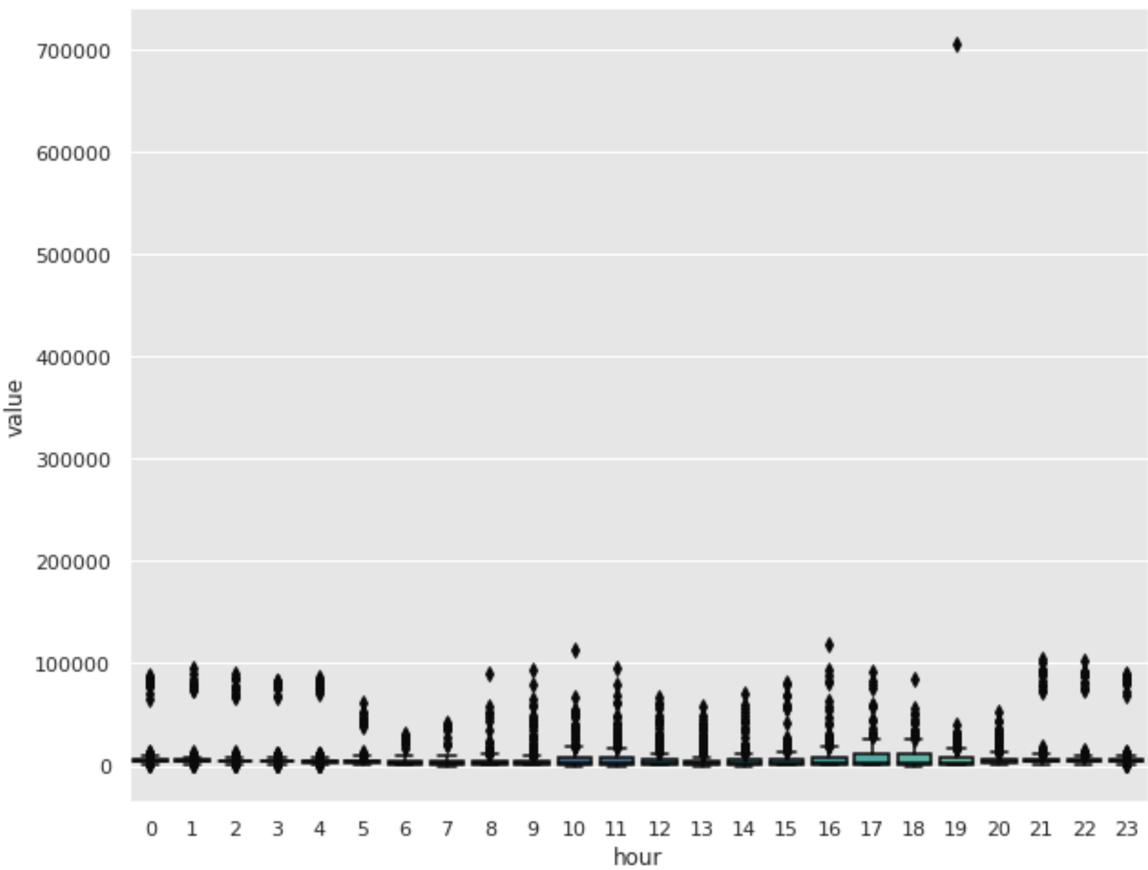
Thus, the demand profile of a set of prosumers varies through the day, the week and through the season. For VPP it is important to know typical behaviour of demand curve. This is the first question VPP developers should define for themselves after the set of prosumers is defined. However, what is the best algorithm for clustering hourly consumption data to find the most accurate estimate?

In this project we would compare three clustering methods in order to find the answer:
- K-means;
- Affinity propagation;
- HDBSCAN

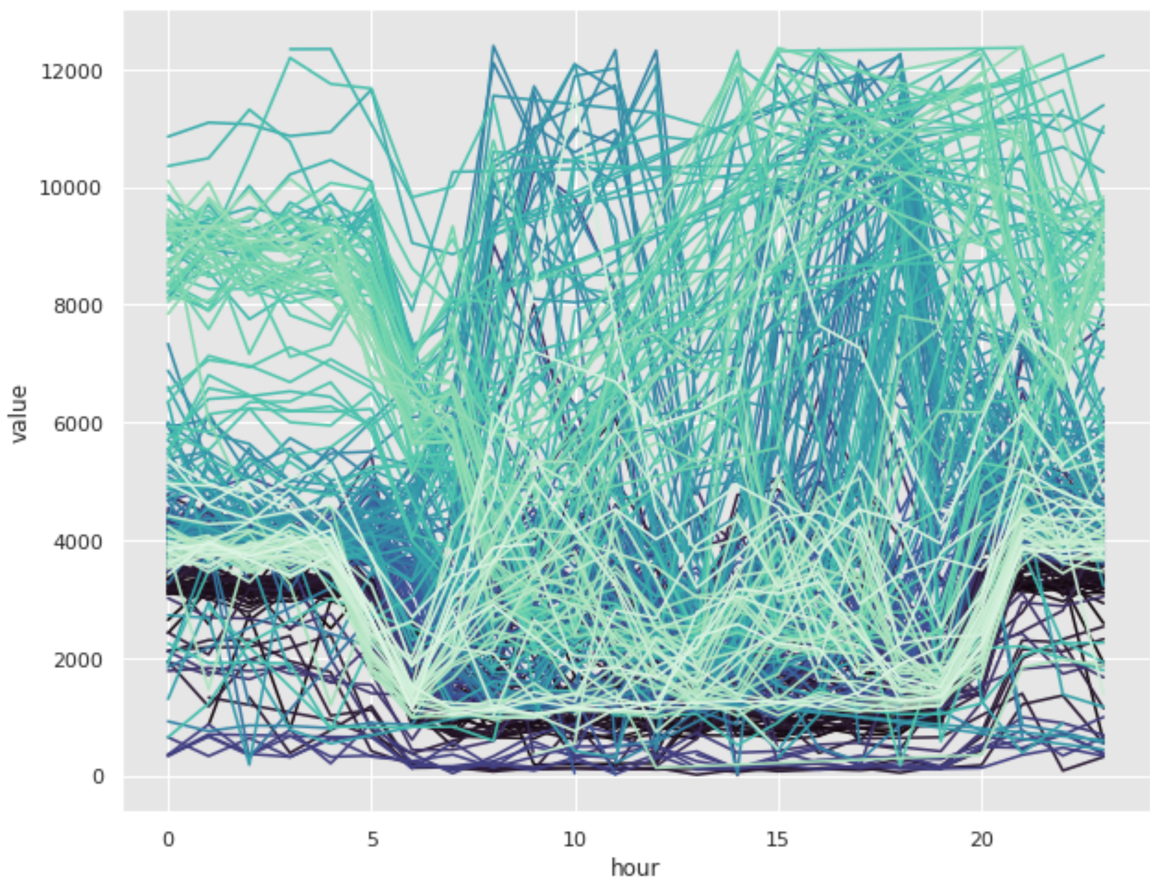
DATA

For this project the open source data from Trial Site Terni, Italy. The trial took place at a small network segment that connected prosumers with solar power plants and a hydroelectric power station. The data set provides power demand/supply profiles of customers in different energy sectors from 02/04/2014 to 21/04/2014.

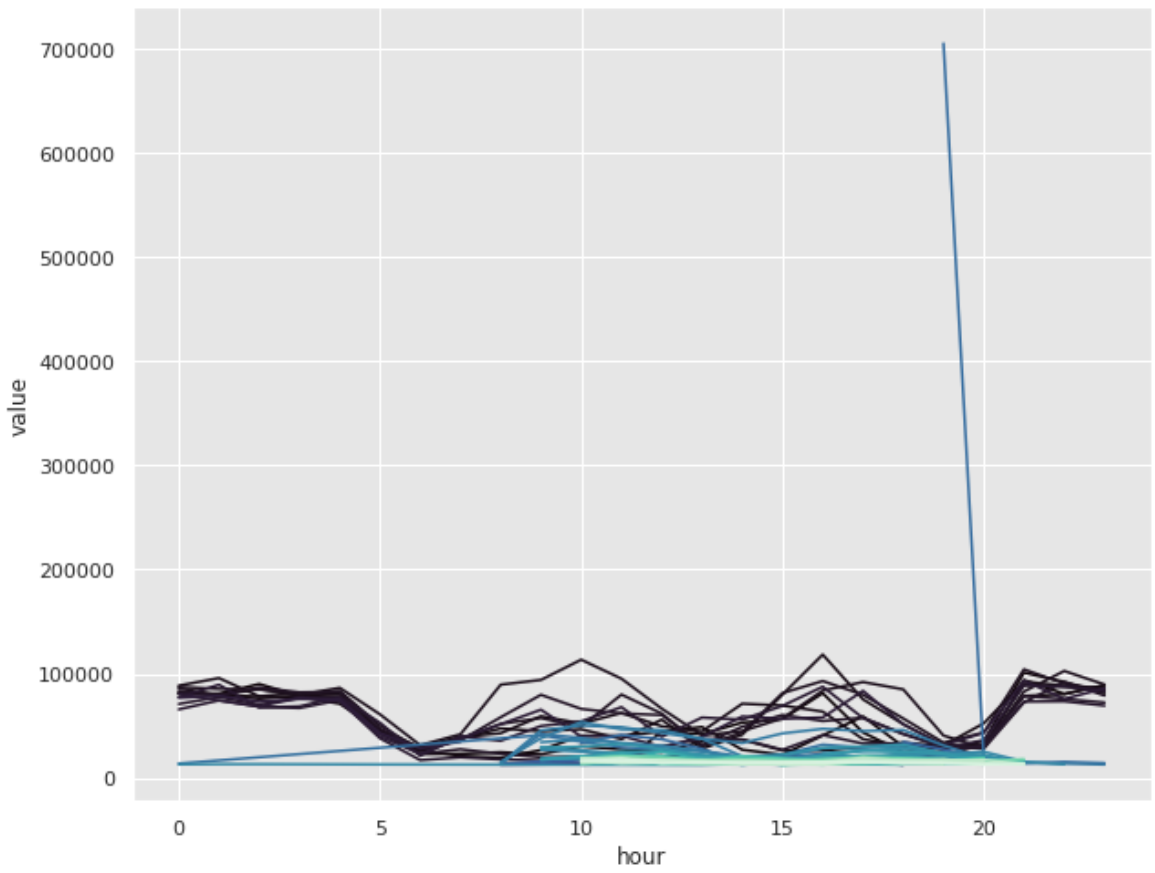
Let's look at the distribution of kW consumed by the VPP by hours



We can see that there are outliers, especially one point at 19:00. Let's remove outliers and look at the kW consumed per hour for each day available



Let’s look at the outliers now:



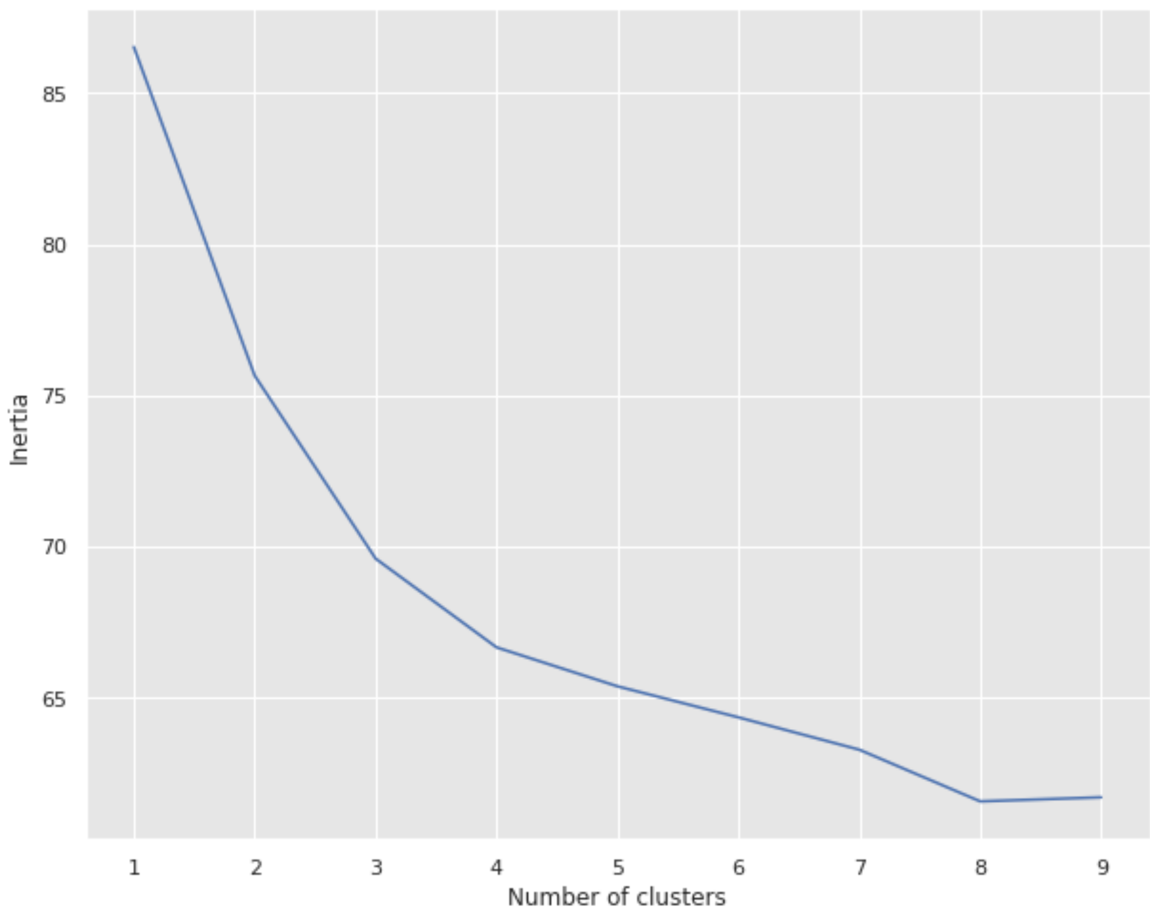
METHODOLOGY SECTION

Methodology section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why. For this project accuracy of three clustering methods would be tested. These methods are frequently used in the VPP research papers:

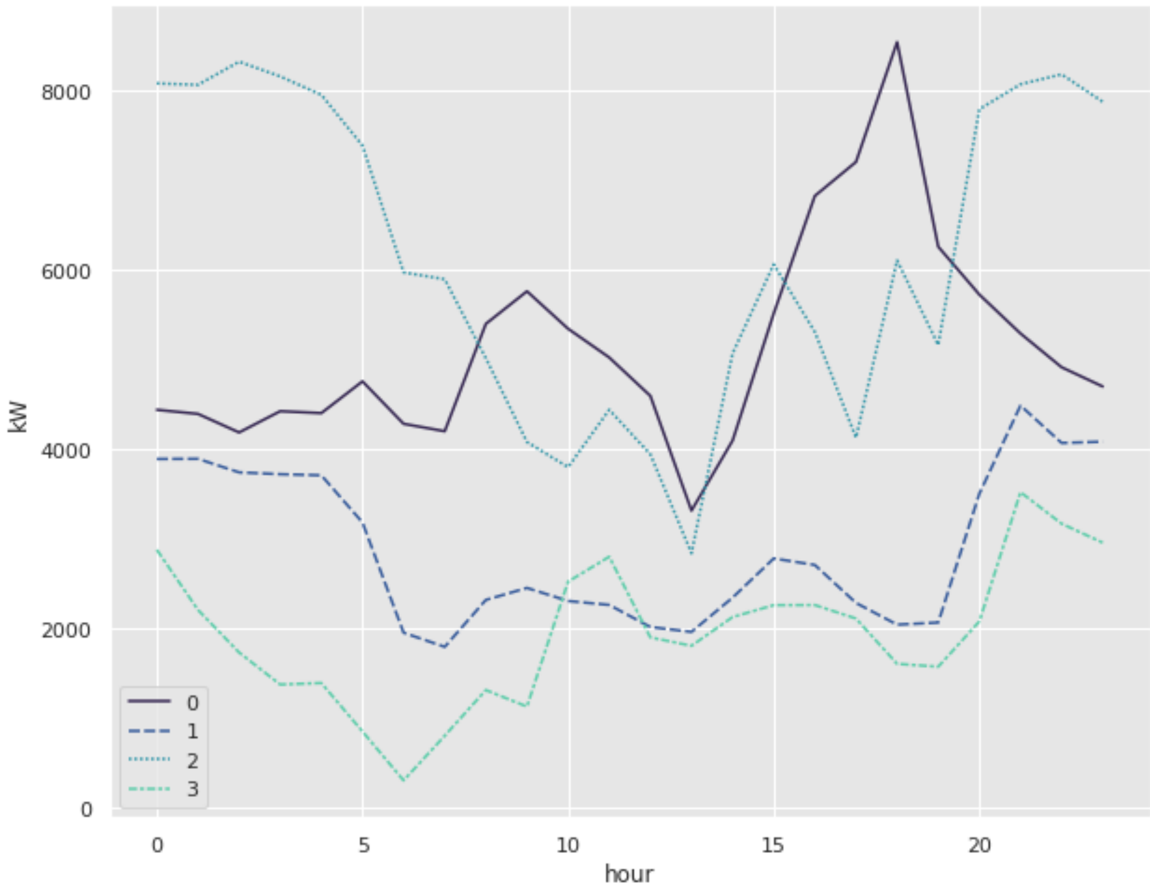
1. KMeans that minimises within cluster variances for the predefined number of clusters
2. Affinity Propagation that is based on 'communication' between nodes and doesn't require to prespecify the number of clusters before the algorithm
3. DBSCAN - density based clustering method

KMeans

As we need to define the number of clusters before using the algorithm, let's look at the sensitivity analysis of inertia to the number of clusters

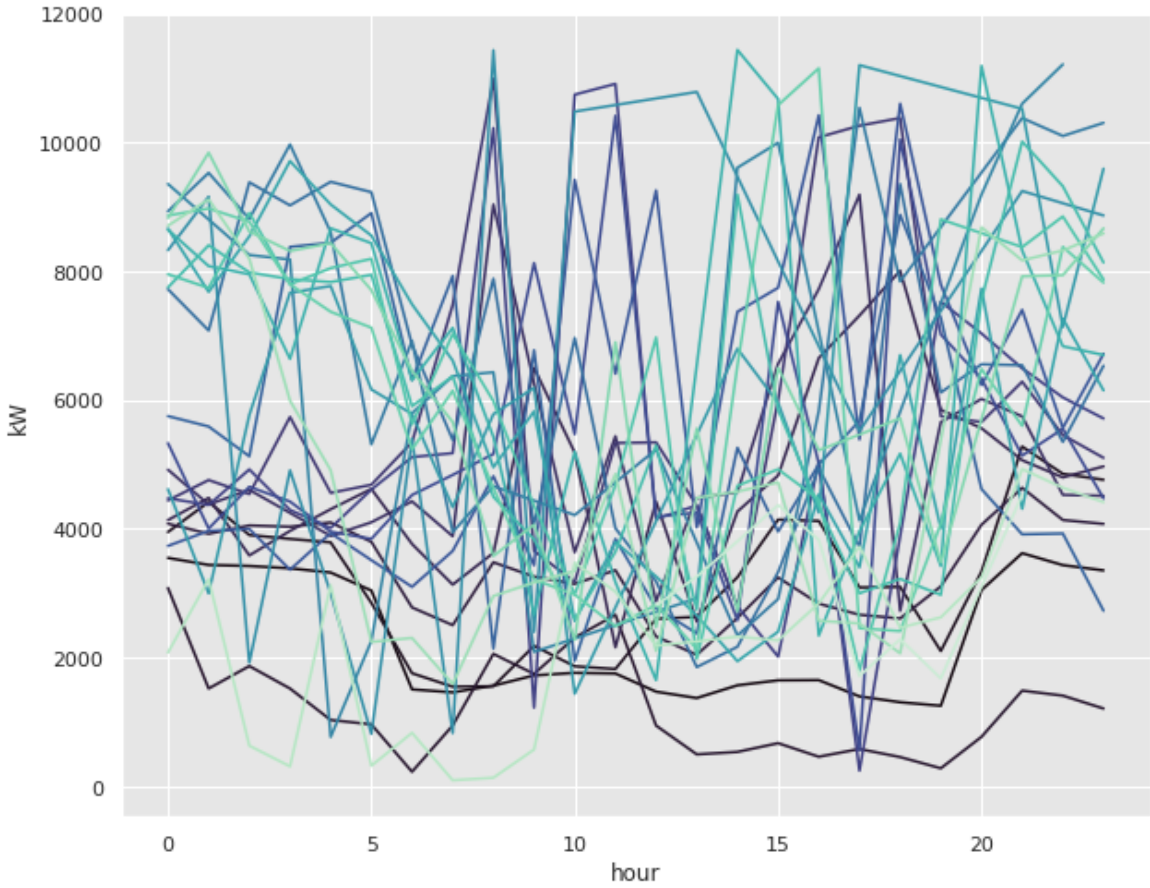


The curve flattens after the 4th cluster. So let`s define the number to 4. Let’s look at the result



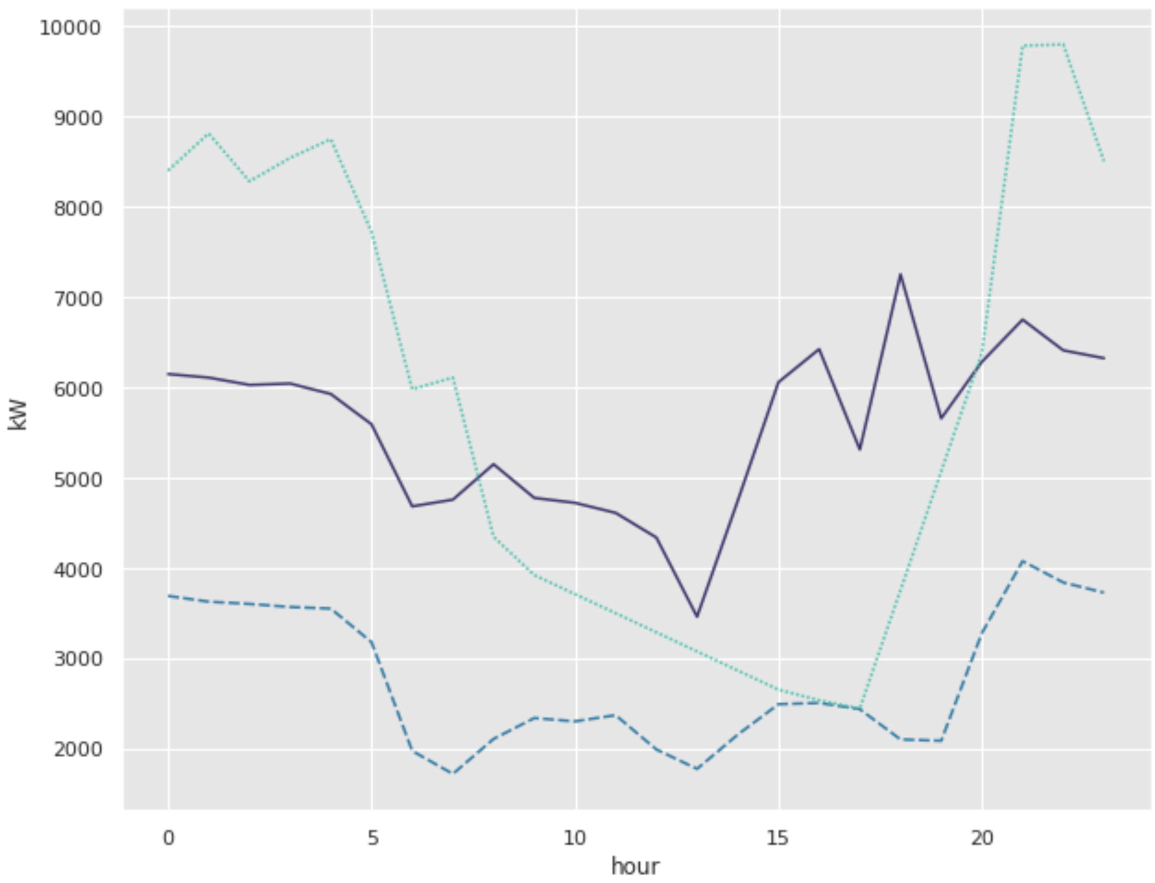
Affinity Propagation

Let’s look at the Affinity Propagation clustering. The result shows 24 clusters:



DBSCAN

For DBSCAN there are only three clusters defined as:



RESULTS

As the ground true labels of the data are uknown, according to the scikit-learn clustering performance evaluation techniques there are two metrics to use for comparing K-means, Affinity Propagation and DBSCAN clustering for the VPP demand profiling:
- Silhouette Coefficient (a higher Silhouette Coefficient score relates to a model with better defined clusters)
- Davies-Bouldin Index (a lower Davies-Bouldin index relates to a model with better separation between the clusters)

Let`s look at them:

Table 1 - Clustering performance metrics

Clustering method	Silhouette Coefficient	Davies-Bouldin Index
KMeans	0.217761	1.646941
Affinity Propagation	0.121782	1.671186
DBSCAN	0.247649	1.793448

As we can see, results are ambiguous. The Silhouette Coefficient is highest for DBSCAN, however the Davies-Bouldin Index is highest as well. However the two metricies are usually higher for density based clustering. The Silhouette Coeff is lowest for AffinityPropagation meaning the worst defined cluster in this case.

KMeans is in the middle and is considered to be a better solution for the VPP demand profiling.

DISCUSSION

The metrics show that for the given data the three clustering method under evaluation perform badly. There are possible improvements - first of all, to add additional data on weather conditions, particularly sun radiation as we expect solar panels are used by prosumers. More accurate outliers analysis may improve the performance.

CONCLUSION

The analysis of energy consumption helps to get a picture of consumers typical behaviour throught a day. The bigger the group of VPP customers the more smoothed their consumption and the better clustering performance are. Understanding the demand profiles is only the first to the VPP management system. The next step is an accurate forecasting of day-ahead values to avoid energy market fees.