/ 100 Datasources
Datasources

To perform this analysis, I looked for households energy consumption data for several years. Thus, I came accross a data for several households in Konstanz, Germany. The data used include consumption of households devices such as dishwasher, freezer, heat_pump, washing_machine, circulation_pump, refrigerator, and electric vehicle (ev). Energy production of households photovoltaic systems (pv), the import and export of energy is included as well. Since data cumulatively sums up over time and I wanted to have plain kWh-values per time interval, a few data manipulation steps were necessary (see Prototype 1 and 2).
The data is available in a simple csv format on https://data.open-power-system-data.org /household_data/2020-04-15/ (https://data.open-power-system-data.org/household_data /2020-04-15/). A backup of the original data and the prepared data are available on https://web.tresorit.com/l/tQeWY#E7GFEo5HKmOZA0gLNxUf_g (https://web.tresorit.com /l/tQeWY#E7GFEo5HKmOZA0gLNxUf_g).

Since this analysis is not a big data problem, Hadoop is kind of an overkill for such an analysis. To answer this question the data could be aggregated over days or weeks and be analysed on a single local computer with Pandas.

The data provide us information about the long-term energy consumption of household devices as well as of industrial devices. However, I focus on the energy consumption of households.

Description of the dataset:

This data package contains measured time series data for several small businesses and residential households relevant for household- or low-voltage-level power system modeling. The data includes solar power generation as well as electricity consumption (load) in a resolution up to single device consumption. The starting point for the time series, as well as data quality, varies between households, with gaps spanning from a few minutes to entire days. All measurement devices provided cumulative energy consumption/generation over time. Hence overall energy consumption/generation is retained, in case of data gaps due to communication problems. Measurements were conducted 1-minute intervals, with all data made available in an interpolated, uniform and regular time interval. All data gaps are either interpolated linearly, or filled with data of prior days. Additionally, data in 15 and 60-minute resolution is provided for compatibility with other time series data. Data processing is conducted in Jupyter Notebooks/Python/pandas. (https://data.open-power-system-data.org /household_data/ (https://data.open-power-system-data.org /household_data/))

.../ 100 Datasources