



## DAV Hackathon

390 minutes

The DAV hackathon will be an introduction to viewing and exploring data using Jupyter notebooks.

## BACKGROUND

Energy use is at an all time high - more than 20,000 TWh (20 000 000 000 000 000 Wh) of electricity is consumed each year. Supplying this demand where and when it is needed is a huge challenge, and most countries have vast, complicated electrical grids with many different power stations, transformers and energy storage systems working together to try to deliver power where needed.

One of the biggest challenges is that one can't simply start up power stations when needed - it takes time to get the turbines turning or the coal fires blazing, and renewable sources like wind don't easily capitulate to the will of man. So it helps to be able to predict the demand for electricity at a given time - allowing power companies to use their resources efficiently. They must take account of all sorts of factors - even what shows are on TV! ([https://en.wikipedia.org/wiki/TV\\_pickup](https://en.wikipedia.org/wiki/TV_pickup) - the British power grid has to deal with a huge extra spike when an Eastenders episode finished and thousands of kettles are turned on and thousands of toilets are flushed).

Finding patterns in how electricity is consumed is useful, then, in making sure that the world keeps running smoothly. In recent years, utilities have been moving away from simply looking at the overall trends and making a best guess from there to looking at details of use in individual houses, to better understand how the energy is used. In this hackathon, we'll be using some data from a single home, outfitted with electricity meters that take readings once every minute. The goal is to find trends, visualize usage patterns and help the power company understand its customer.

## THE DATA

We will be looking at a subset of the household power consumption dataset, available here:

<https://archive.ics.uci.edu/ml/datasets/individual+household+electric+power+consumption#>

The original dataset contained 47 months worth of data - we've cut it down to 8 weeks to make it easier to handle. In you're interested, the notebook titled 'data\_prep.ipynb' shows the steps taken to slice the data, deal with some missing



values, convert columns to the appropriate types and save it for easy access during this hackathon.

The data is loaded into a Pandas dataframe. As well as columns for the date and time, we have the following (from the description of the data available at the link above):

1. `global_active_power`: household global minute-averaged active power (in kilowatt)
2. `global_reactive_power`: household global minute-averaged reactive power (in kilowatt)
3. `voltage`: minute-averaged voltage (in volt)
4. `global_intensity`: household global minute-averaged current intensity (in ampere)
5. `sub_metering_1`: energy sub-metering No. 1 (in watt-hour of active energy). It corresponds to the kitchen, containing mainly a dishwasher, an oven and a microwave (hot plates are not electric but gas powered).
6. `sub_metering_2`: energy sub-metering No. 2 (in watt-hour of active energy). It corresponds to the laundry room, containing a washing-machine, a tumble-drier, a refrigerator and a light.
7. `sub_metering_3`: energy sub-metering No. 3 (in watt-hour of active energy). It corresponds to an electric water-heater and an air-conditioner.

Explaining active vs reactive power is a job for the engineers. For the purposes of this hackathon, treat 'Global\_active\_power' as a measure of the power used. An additional metric that would be of interest to power companies is a quantity known as *power factor* - the ratio between active and reactive power. If you're looking for something extra to examine, you could add this as an extra column and investigate how it varies.

'Global\_intensity' describes the current entering the house. This is of less interest than the power drawn - it can be safely ignored.

'Voltage' simply shows the voltage supplied to the house. We tend to think of this as a constant 240 volts, but in reality it fluctuates based on a variety of factors.

There are three sub-meter readings, each corresponding to a section of the house. These should be very interesting, allowing us to see **how** the power is being used. The units are not kW, meaning some conversion will be necessary before you can compare these readings to the global power usage figure.



## THE CHALLENGE

You have been presented with a jupyter notebook with the data loaded and ready to go. The dataset contains 8 weeks worth of readings, taken every minute - more info in that in the section on the data. As it is, this data isn't of much use to anyone - what human could make sense of what is essentially a spreadsheet with 80640 rows? Your task is to sift through this data, looking for patterns and creating visualizations that will help us see the big picture at a glance, and let us draw insights as to how electricity is used.

By the end of this hackathon, the goal is to be able to tell a story about this data. Show us what you can find while answering the following questions, and think about how to illustrate your answers with visualizations

- 1. What does the usage pattern look like?**
- 2. How much electricity should the power company be ready to deliver?  
(Don't spend much time on this one)**
- 3. How would you summarize all of this data?**

## FIRST STEPS

You may be itching to jump straight into your own analysis, but to get you started and give you some ideas, try the following before moving on to your own work:

- Read the description of the data thoroughly - make sure you have a good understanding of what each column represents, and think about what might be important.
- Plot the column 'Global\_active\_power' as a line graph
- Select the first week of data, and plot voltage for just that week.
- Find the maximum power ('Global\_active\_power'), and when it occurred

If these give you trouble, ask for help - it's important to make sure you can handle these sorts of tasks before trying to answer the bigger questions.

From here, it's up to you what you try next. We've given you some questions and hints to think about, but hopefully you also have ideas to try. Let's get hacking!

## QUESTIONS TO THINK ABOUT

You may have your own ideas as to how to answer those big questions, and by all means explore those avenues. If you're unsure where to go next, or are looking for



some hints, here are some other questions to consider which will help you on your way:

### **What does the usage pattern look like?**

What do we mean by usage pattern? Is there a daily trend in how much power is used? Weekly? Monthly? Are any days or time periods different from the others? Are any of the sub-metered areas used in a different pattern to the main house?

### **How much electricity should the power company be ready to deliver?**

If the house tried to use more power than that available it could cause all sorts of problems. But there may be times when usage is guaranteed to remain low, in which case the power company could conceivably scale down power production or lower the price to tempt companies into operating during that time slot. Are there times when usage reliably stays below some threshold?

### **How would you summarize all of this data?**

There is a lot going on in this data. How would you explain it to your manager, or to a technician at the power plant? Think about: Is an average sufficient to describe each of these quantities? Perhaps an average, with a max and min as well? But what about things that vary over time - knowing that the average power use is 3kW is useful, but could leave room for a nasty surprise when it spikes to 9kW and then falls to <1kW an hour later. How can you show things like how the voltage fluctuates in a simple and easy to read plot?

### **General hints**

There are many different avenues to explore - try out different ideas and then go deeper into ones that seem interesting. There is much to be gleaned from paying close attention to things like the sub-meter readings that could tell you more about how the power is being used. Think also about information beyond the simple sensor readings. We're dealing here with human beings, so human factors will play a role - sleep, work, leisure, holidays, unpredictable bladders, midnight snacks... who knows what will contribute to the ways we use electricity?