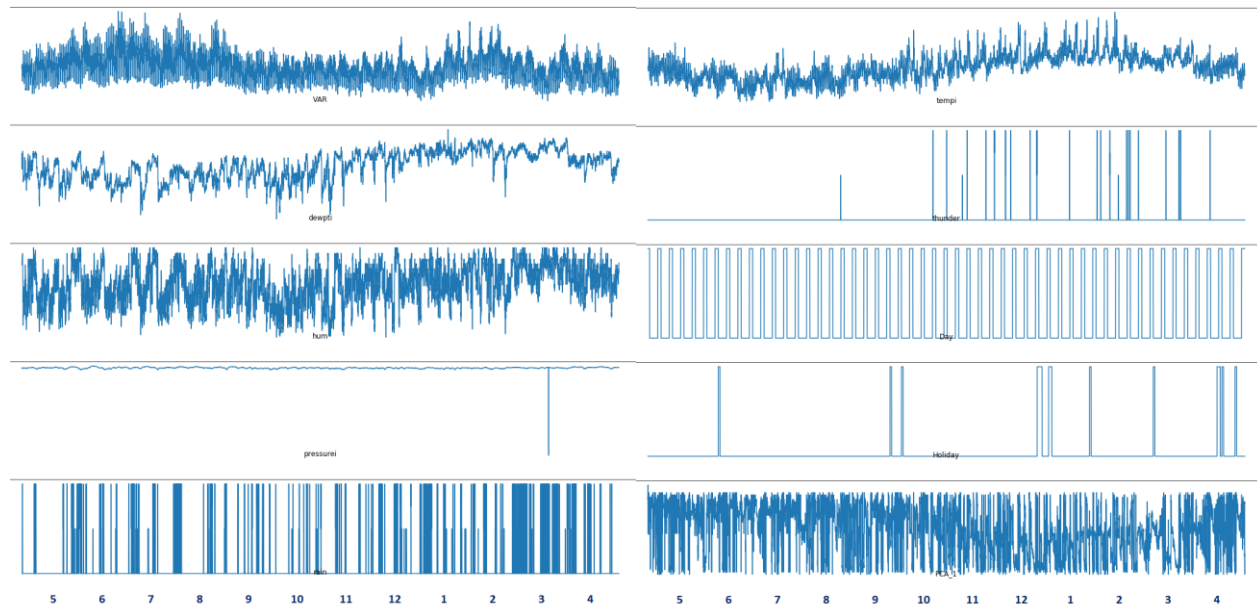


Let's have a look at the data to see

VAR

The dataset we are working with contains the variable VAR, as well as weather details .

Let's have a look at each variable individually and see if we can spot any patterns.



tempi

Let's have a look at temperature (tempi), first. Apparently, temperature tends to be the lowest in June, July and August, and the highest in January and February. This means the data was probably collected from a country in the south hemisphere, possibly Australia.

VAR

This makes sense when we look at power consumption (VAR), which is the highest between June and August – cold months during which heating is required – as well as January and February, likely due to high air conditioning usage. On the other hand, in spring and fall, when the temperatures are mild, power consumption is moderate.

hum, dewpti, thunder

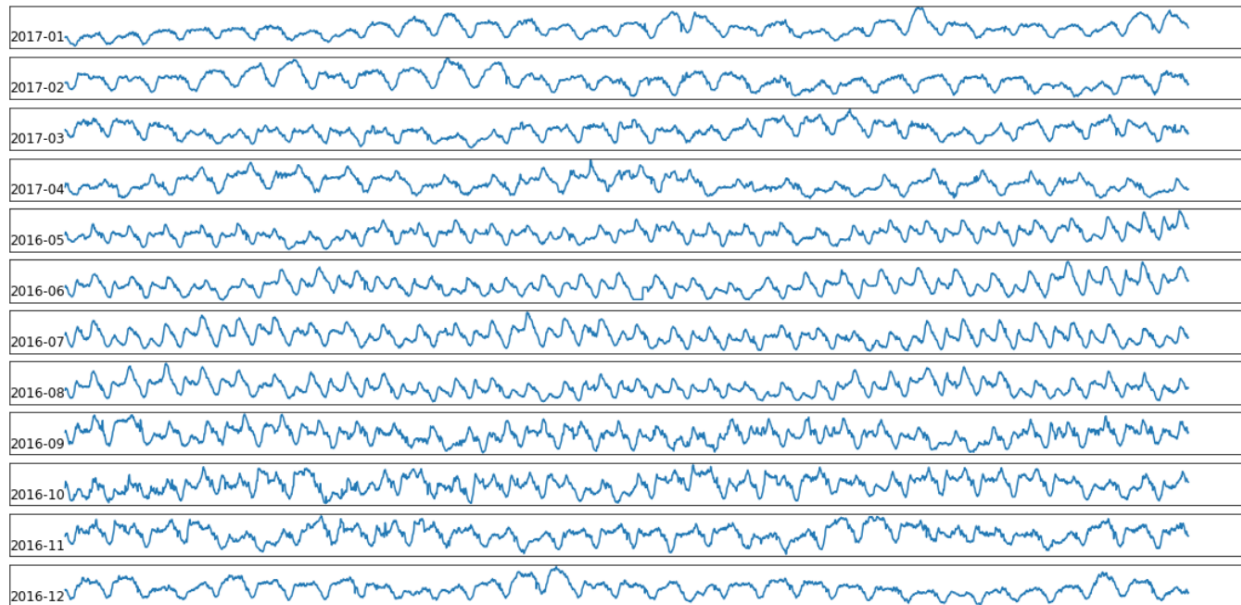
High humidity (hum) and dew point (dewpti) values as well as thunder frequency are consistent with high temperatures in warmer months, which may suggest humid subtropical climate.

pressure

There appears to be a data issue with pressure in March. A good way to deal with missing or inaccurate values is to replace them with estimates. This will require inspecting pressure data in more detail but one easy way to deal with the issue would be to use the average pressure value at the same time of the day, from a few days before and after the corrupted value occurred.

For now, let's have a closer look at power consumption (VAR), since it's the value we are modeling.

Zooming into individual months reveals a daily pattern: most likely high usage during daytime and low usage at night. Interestingly, colder months are characterized by a double daily hump. The reason is not clear at this point; one explanation could be that in colder months people consume more power in the morning, as they get ready for work (first daily peak) and in the evening, as they turn the heat up upon returning home (second daily peak). On the other hand, warmer months miss the second peak because the temperature gets lower in the evening anyway, limiting the need for air conditioning.



Zooming even closer, to weekly data, we finally observe

