Energy dataset

Daniel García Hernández

April 4, 2020

This time the data we will be using comes form the Spanish TSO (Transmission system operator) Red Eléctrica de España, and their public website. You can get data using the API **link**.

The data consists of several aggregated timeseries. Each one of the timeseries takes into account one type of energy production like nuclear, solar, hydro, etc, plus the total country-level energy demand.

Questions:

1. **Get the shape of the dataframe**
2. **Create a list called "features" containing all the column names for those columns that are not "datetime"**
3. **How many NaN values are there in the dataset?**
4. **Convert the "datetime" column into a Series containing datetime objects**
5. **Create the following columns: year, month, day, hour, weekday. Use pandas methods and functions.**
6. **Create a "weekend" column containing True if day is Sat/Sun or False if day is Mon/Tue/...**
7. **Calculate the average values per weekday for the columns in "features" list**
8. **Create a function that returns the season out of a datetime object**
9. **Calculate average demand and spot price for each season. Use that function to create a "season" column**
10. **Why are there so many NaN values in the "solar" column?**
11. **Calculate the total income for each type of power generation and total costs for consumption**

12. **Create two plots: hourly average power demand and hourly average spot price**

13. **Create a categorical variable called "wind_category" using the "wind" column. Use three categories: low, mid, high.**

14. **Which is the power generation type that has generated the most energy? And the least?**