Applied Data Scientist – Market Analysis GEMS Romania : Use Case

Today you are a Market Analyst for GEMS Romania and have to make a presentation for an Engie audience who is not necessarily an expert in Data Science and in all the concepts discussed. The entsoe-dataset is made up of data (UTC timezone) publicly accessible on entsoe transparency (ENTSO-E Transparency Platform (entsoe.eu)). The solar installed-capacity dataset is composed of data from the Hungarian TSO (2020 to 2023) and scenarios made by our team for 2024 and 2025.

A : Data manipulation / resampling

1) Using the entsoe dataset provided, reconstruct the "Power Generation Stack" for Romania and Hungary from 2017 to today. The data must be aggregated weekly. What observations can you make?

2) Reconstruct the average Load (demand) in Romania for each day of the week (Monday to Sunday) and hour (0 to 23) on 2023 data only. What observations can you make? (if you wish you can switch the data to another time zone more representative for Romania)

B : Romanian Load Forecast

3) Compute a 2-year ahead forecast for the Romanian Load and back-test your model for the previous 2 years. Please support your work with adequate data visualization and some comments about the performances/results of your model.

4) In reality, producing forecasts for this type of series is more complex, many parameters can be taken into account and fundamental models are sometimes necessary. Through your knowledge of the sector or your research, share some ideas/parameters/features that could be used in addition to historical data to predict this Load? Add those as limitations in your presentation.

C : Hungarian Solar Generation Forecast

5) The Load Factor corresponds to the expected generation of an asset or technology in relation to an installed capacity, we also speak of capacity factor (https://en.wikipedia.org/wiki/Capacity_factor). For solar, we can for instance build a Load Factor Matrix which gives the load factor for each hour and each month. You find an example below.

Build, *from the observed solar generation data from 2021 to 2023 and the monthly installed capacity in Hungary from 2021 to 2023*, an observed load factor matrix for Hungary. What are limitations of this matrix made over 3 years of observed data?
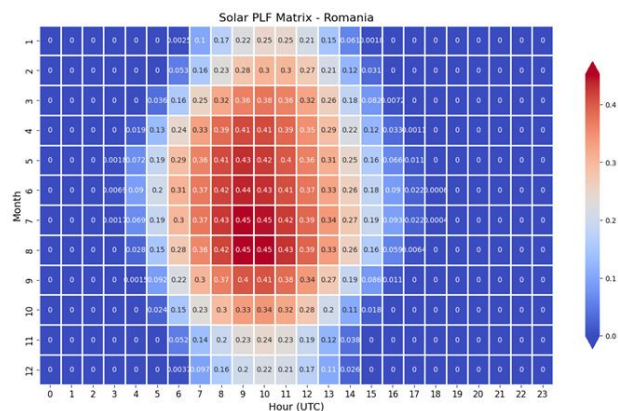
6) Using your matrix and the installed capacity projection scenarios for 2024 and 2025, forecast hourly solar generation in Hungary over these two years.

You must give us your notebook and a presentation (pdf, ppt…) with your findings and suitable data visualization of no more than 8 slides. Please send us your presentation and notebook at least 24 hours before the live presentation. The notebook needs to be playable. In case of any questions or remarks please contact zacharie.abid@engie.com