**Capstone 2 Project Proposal: Electricity Demand and Generation Forecasting**

**Project goal:** forecast the hourly total load (demand) for one year, and possibly the renewable electricity production (from solar and wind) also, based on weather data and three previous years of hourly generation data for all of Spain.

**Context:** Accurate power generation forecasts, both short term and long term, are critical to a green energy future. Accurate forecasts enable renewable generators to correctly price their energy and maximize contribution from renewable sources (while minimizing carbon sources), and are important to operators for system stability and planning.

**Criteria for Success**: This model will be successful if the forecasted total load, forecasted onshore wind generation, and forecasted solar generation are close (within approximately 10%) to the actual values for these variables. We can also consider it successful if the predicted total load is more accurate than the standard system operator prediction contained in the dataset.

**Scope of Solution Space:** This initiative will forecast total demand (load), solar production, and wind production on an hourly basis based on the previous three years of hourly load and production data, and on weather data (hourly for three previous years and for forecast window).

Electricity demand forecasting is often done using time series forecasting when dealing with data containing only time and demand variables. Electricity demand is strongly dependent on exogenous variables like temperature, humidity, etc. Thus, with weather data available, regression may also be a relevant model choice. It would be desirable to try out and compare several types of models potentially including, but not limited to: ARIMA, SARIMAX, Random Forest, XGBoost. It may be necessary to perform a Fourier Transform or other transformations to reduce the seasonality (yearly, weekly, daily) of the time series data.

**Constraints within solution space:** Three years of training data may not be enough to accurately predict the targets, especially if one year was anomalous with regards to weather or demand.

**Stakeholders to provide key insight:** The stakeholders for such a real world project would be the system operator, and wind and solar generators.

**Key data sources:** The weather data that will be included in the model includes hourly: time, temperature, pressure, humidity, wind speed/direction, and weather descriptor (clouds, few clouds, clear, etc…). The hourly energy production data includes time, generation by method (approximately 21 different methods), day ahead solar generation forecast, day ahead wind generation forecast, total load forecast, total load actual, day ahead price per MWH, and actual price per MWH. All data is sourced from a [Kaggle dataset](https://www.kaggle.com/datasets/nicholasjhana/energy-consumption-generation-prices-and-weather) that was sourced from [Open Weather](https://openweathermap.org/api), [ENTSOE transparency platform](https://transparency.entsoe.eu/dashboard/show), and [Red Electrica](https://www.esios.ree.es/en/market-and-prices).