Siddharth Arora **Trinity 2023**

Spark Grid Case

Since the 1980s, Spark Grid (SG) has held a competitive position in the energy sector in Great Britain. SG has a huge customer base, and it operates a vast power transmission network across the whole country. However, more recently, SG has suffered severe financial setbacks due to poor decision-making, and the future of SG looks ominous.

In the past, due to their strong position in the market, the haughty senior management at SG completely disregarded any advice from their predictive analytics team, and instead made operational decisions based on subjective judgment. The policy at SG was simple – generate excess electricity using fossil fuels to prevent blackouts and sell the surplus electricity at a loss to avoid costs associated with battery storage. With the adaptation of stricter climate mitigation guidelines, higher penetration of renewables like solar and wind into the energy grid, and increased competition, SG had no option but to make informed decisions.

In the last bid to save Spark Grid and reignite its once bright legacy, the company's top management was forced to resign, and Saria Marapova, one of the top professionals in the energy sector, was hired as the CEO. Saria soon realized that the company's operational decision-making was gobbledygook (to put it politely). She sent an email calling for an urgent meeting on Monday at 9:00 am with the analytics team, who had become bored of playing Tetris on their computers for years. The prospect of being a part of a potential revival of the company ignited a new excitement in the analytics team, which was led by Nonas Jeubauer.

At 9:00 am on Monday, the exuberant analytics team at Spark met for a huddle in their office. Nonas gave an inspirational speech, and like warriors on their way to the battlefield, the team made its way to the conference room to meet Saria. Morning Saria, it is a pleasure to meet you, said Nonas, in a slightly squeaky voice. Likewise, Nonas, it is great to meet you and your team, and next time, please be on time. The team felt slightly disoriented; no CEO had spoken to them in ages.

After an awkward silence, Saria took charge of the meeting and said – "as you all know, we are on the brink of collapse. Our competitors use statistical models and machine learning for predictive and descriptive purposes. I want you to tell me what model/s you have developed". Nonas replied – "it is encouraging that you believe in using forecasts for making decisions". Saria looked baffled, she thought, who does not. Nonas continued – "we have developed a state-of-the-art linear regression model with multiple variables using all the historical data. We have a strong hunch that the adjusted R^2 from our model is impressively high. In fact, to be specific, I think it may be around 0.9 give or take". Saria thought - did the head of our predictive analytics team just say, that his modelling is based on a 'strong hunch', and in which multiverse was that statement specific!

Nonas continued – "to accommodate the effects of weather on electricity demand, we use historical 'temperature', 'wind speed', and 'cloud cover'. To model any trend and yearly seasonality, we use a variable for 'day of the sample' and 'day of the year', respectively. Since demand exhibits a weekly seasonality, we also have dummy variables to accommodate the 'day of the week' effect during the modelling. Moreover, special days (such as public holidays, e.g., New Year's Day) and proximity days (days around special days) witness a lower demand compared to normal days. To tackle this holiday effect, we use dummy variables for 'special days' and 'proximity days'". In a confident tone, Nonas said – "we have delivered, our model

is reliable, and there is no further scope for improvement". Meff Joore, who was a senior data scientist in the team, interjected at this point and said – "well, it is perhaps worth noting that demand has a lagged effect, whereby daily demand observed in the last seven days may be correlated with today's demand, but we have not investigated using lagged dependent variables in our modelling yet". Meff continued – "we have also not investigated any variable transformations to potential incorporate nonlinear effects". Nonas did not look pleased at this point. Saria took an analytics elective during her MBA, so she asked – "are all variables significant, how good is the model fit, and how do the residuals look?" Unfortunately, she did not get a proper response.

After the presentation, Saria thanked the team and went to her office to reflect on the meeting. She knew that there was a need for external consultants to help with this modelling exercise, so she sent an email (more of a plea) asking for assistance.

You decide to accept Saria's request because you feel this work could save millions in operational costs for SG and reduce their carbon footprint. Saria/SG are requesting you to:

1. Provide any insights into the relationships between variables. See provided dataset - GBElectricityAndWeatherData.xlsx or GBElectricityAndWeatherData.rds

2. Build a linear regression model of the following form using the entire dataset that as currently used at SG:

$$Demand = a + bTemperature + cWindSpeed + dCloudCover + eSpecialDay + fProximityDay + gDayOfSample + hDayOfYear + iMon + jTue + kWed + lThu + mFri + nSat + error$$

3. Develop a model that can generate forecasts for electricity demand, whereby your model overcomes the shortcomings in the approach adopted by Spark Grid. Compare the performance of your model with the Spark Grid model developed in Step 2.