Jad Mekari

# Capstone Project Report

## Energy Demand & Renewable Energy Generation

### Problem statement

As UK energy prices are increasing at an unprecedented rate, the UK's reliance on energy imports has come into question. The problem is, sustainable energy solutions are currently not fulfilling a large enough portion of the National Energy Demand. Of the wind, solar and other renewables that came on stream in 2020, nearly two-thirds – 62% – were cheaper than the cheapest new fossil fuel, according to the International Renewable Energy Agency (IRENA). Making sustainable energy solutions not only cleaner, but cheaper too.

### Project Phases:

1. The first phase of this project will be to explore both datasets, to identify meaningful insights in both energy consumption over the years, and renewable generation. This exploration will also involve feature creation, as well as a test for correlation between consumption, generation & the engineered features.

2. Using a Test-Train split of 75:25, set just before the inflection point where energy consumption dropped significantly during the Summer of 2020, we will employ a number of different modelling technquies. We will then compare these techniques' performances using the MAPE (mean absolute percentage error).

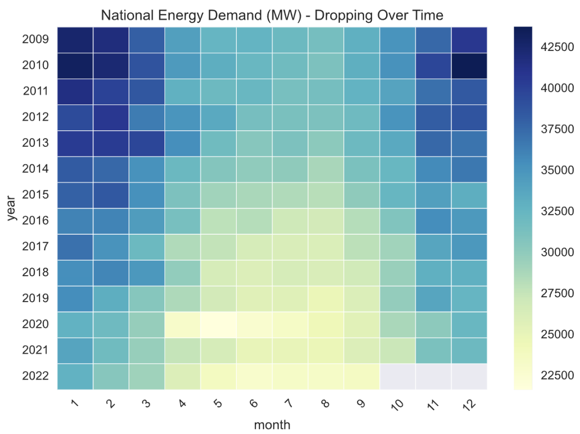
3. The aim of this phase will be to identify when in the future energy demand and renewable generation will intersect. This intersection point will (in theory) represent the beginning of a time when national energy demand can be fulfilled by renewably generated energy.

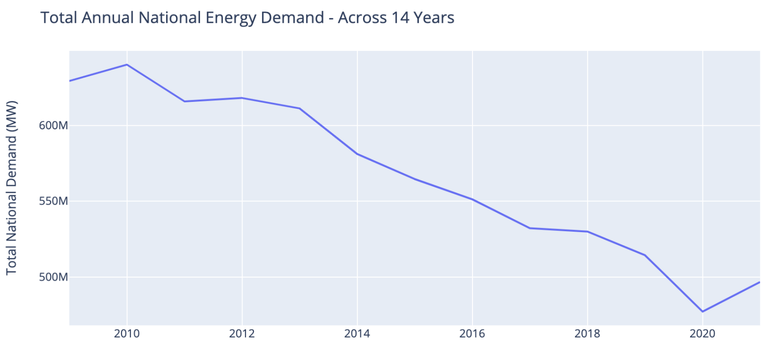
## Phase One - EDA:

### Observations

Energy Demand (Consumption):

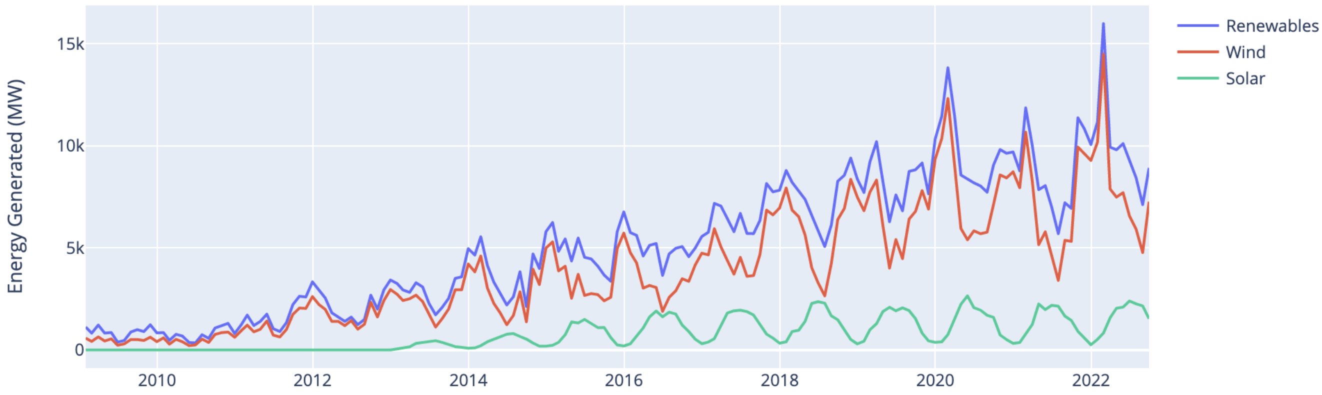
Since 2009, both the winter and summer months' energy consumption have seen a decrease, with the summer months' average consumption falling from 30,000–35,000 megawatts to 22,000–30,000 megawatts. Across all of the months in 2009–2010, usage tends to be highest in the winter months (12–1, 2) and lowest in the summer months between 5–9. This can be supported by the fact that May 2022 had the lowest usage, while December 2010 had the highest usage, which was just over double (202%) that of the former's.



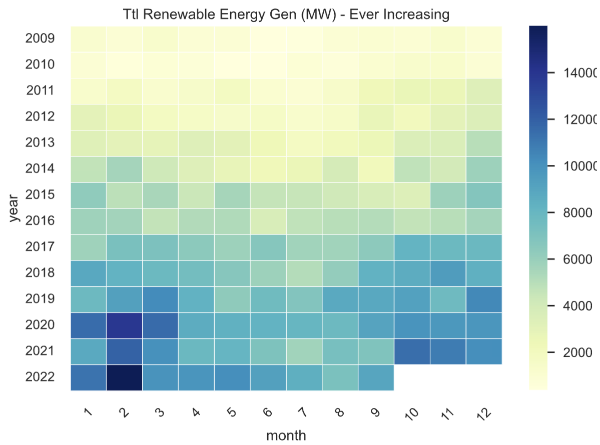




This is also supported by the fact that the sum of total energy demand aggregated yearly across the 13 years (excluding 2022 as it is incomplete) shows a steady decline in energy consumption. With 2019 - 2020 being the largest recorded YoY drop across the years at a -7% drop in total MWs.

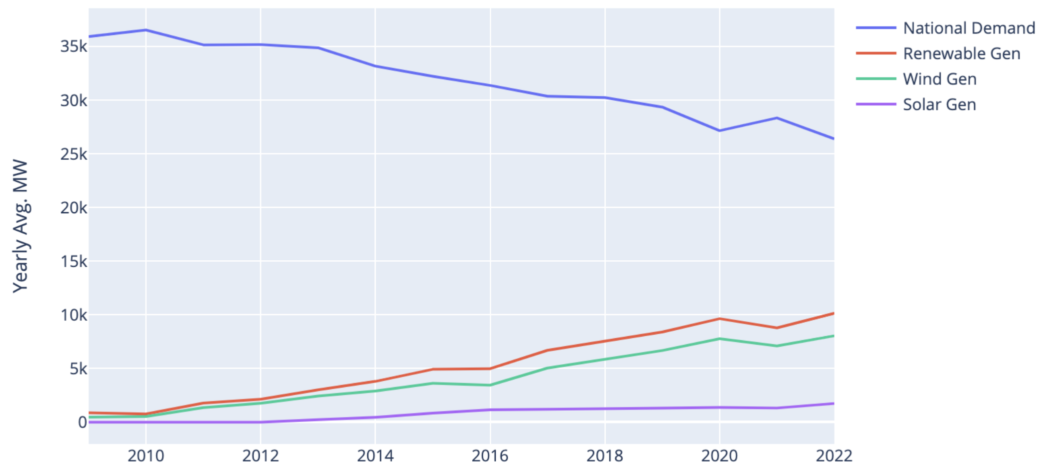
Renewable Energy Generation (Production):




- Wind makes up the majority of Renewable energy produced nationally
- Total Renewable energy generation is increasing consistently across the 14 years
- Increase is due to Wind Energy production increasing consistently and with observable heteroscedasticity i.e. the variance is increasing over time
- Solar energy has never accounted for more of the energy generation than Wind
- Solar energy and Wind energy have a clear seasonality, with both of their respective seasonal cycles peak out of sync from one another i.e. Solar energy is mostly generated in the summer when Wind energy is least generated
- Solar energy is increasing at a much slower pace than Wind energy, presumably because the UK weather is best suited for Wind Generation




Energy Demand & Renewable Generation:



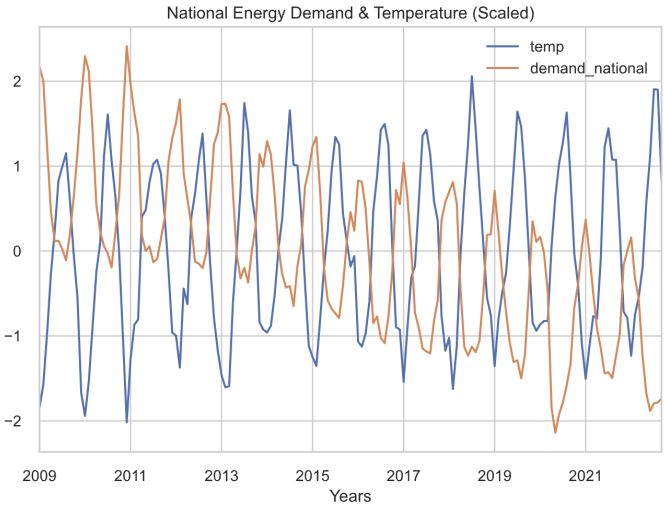

This graph illustrates the convergence between the average national energy demand and the production of renewable energy. Here, we plot the MW value on a yearly basis. Despite the fact that macro-political-technological forces influence both datasets, we can find convergence in their general trajectories. For both convergent trends, we can also see a largely linear trajectory. Thus, we may infer that a future intersection might exist. We will reach this juncture when we can provide all of the country's energy needs from renewable sources. Later on in the project, we'll try to locate this intersection.

## Correlations:

Demand & Temperature:

We were able to determine from graphing, that there was a noticeable seasonality in the demand data, with energy use declining markedly during seasons of higher temperatures. This can be attributed to several things, including:

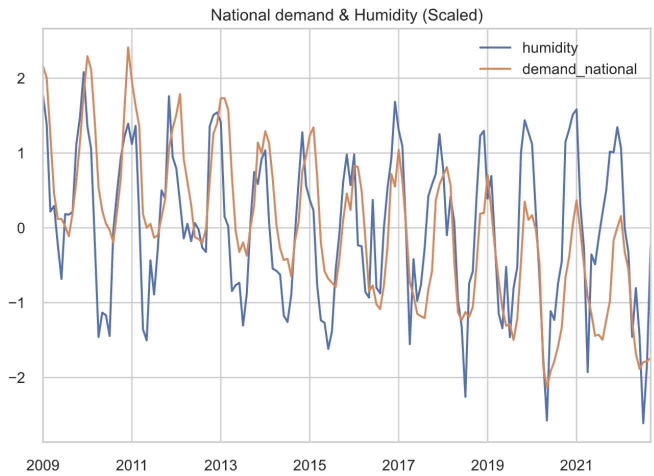
- Decreased population, which results in lower consumption since more individuals travel over the summer to coincide with school breaks
- Less requirements for electric heating
- Due to the extended daylight hours, people are often outdoors more during these months, which reduces their likelihood of home energy use.
- The prevalence of taking hot showers has decreased




Demand & Humidity:

Humidity and National Demand have a high positive correlation, one explanation for this could be that there is an increased (and decreased) amount of energy spent on cooling systems to counteract the effects of humidity on the populous as well as in industrial processing facilities that require humidity control solutions to ensure the optimal quality of the final product.

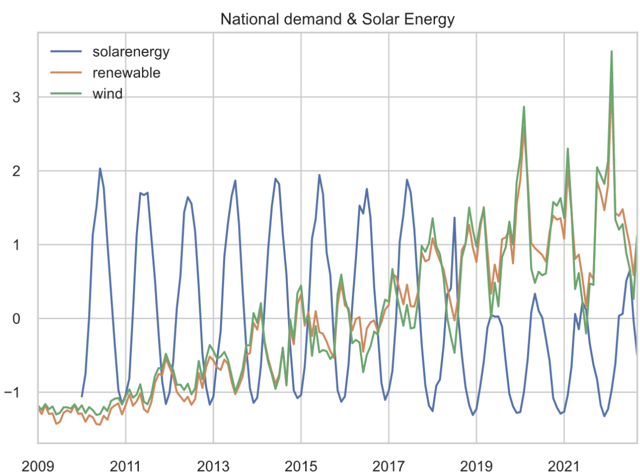
Another explanation could be the relationship between humidity and precipitation, whereby higher humidity means more rain. Rain is more prevalent in the colder months which we know are the months with the highest demand. This relationship most likely is not causal, instead it may be coincidental that both follow a similar seasonal trend due to sensitivities to time of year.




Renewables & Solar Energy:

There is a clear negative correlation between Solar Energy and Renewable Energy generation. If we look at both wind and solar, we can see that this is mainly attributed to the negative correlation between wind & solar energy. With solar energy being positively correlated to Solar Energy, offsetting its impact on Total Renewable gen.

While it is hard to explain why this isn't the case without specific domain expertise in geophysics, we are able to conclude that wind generation drops in times of high solar energy recordings.



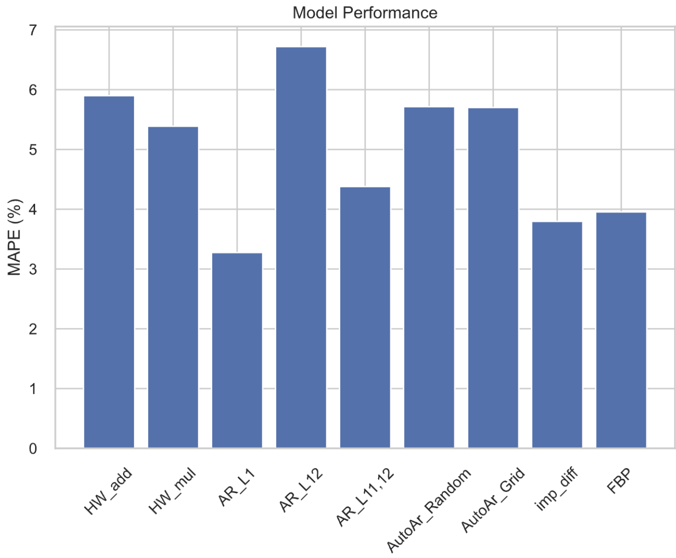
## Phase Two - Modelling Techniques:

In this phase, many different modelling techniques were employed to identify the best performer. The presence of an inflection point at the beginning of the test set, made forecasting a difficult feat. It also begs the question, is this a true reflection of the models ability to forecast into the future? Because if we are to forecast into the future, we would be training on the whole set, which would account for the inflection point.

To evaluate our models, we used MAPE (mean absolute percentage error), below we see a bar plot showing how well each of the models performed.

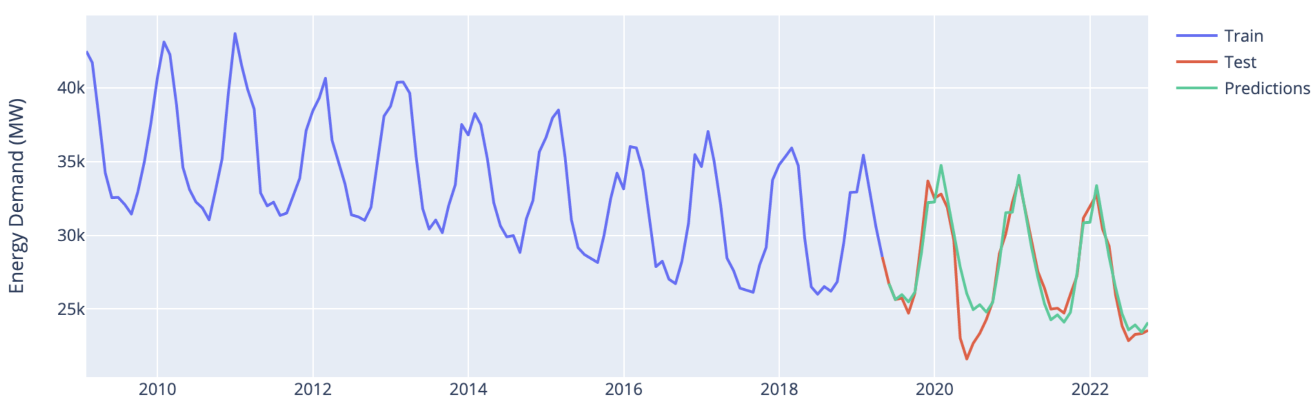
The chart to the right shows the different model techniques used during the project. Interestingly the simplest and least automated model, the AR L1 model performed the best. Second up was the SARIMAX Implicit difference model which used the parameters defined by AutoArima's Grid Search technique. Followed closely by Facebook's Prophet.

The worst performing models were the L12 AR model, followed by Holt Winters Exponential Smoothing set to 'Additive' for both the trend and seasonality. The Holt-Winters model with seasonality set to ‘Multiplicative’ didn't perform as badly, performing better than both the models produced by AutoArima set to Random Search & Grid Search.




**The best performing model, MAPE = 3.275%:**

**AR(1) - Stationarised, Tuned, Fit & Untransformed**




While the above model doesn't account for the large dip we see in 2020, it does well to match the typical seasonality and trend of the model. Interestingly the simplest model was the highest performing model.

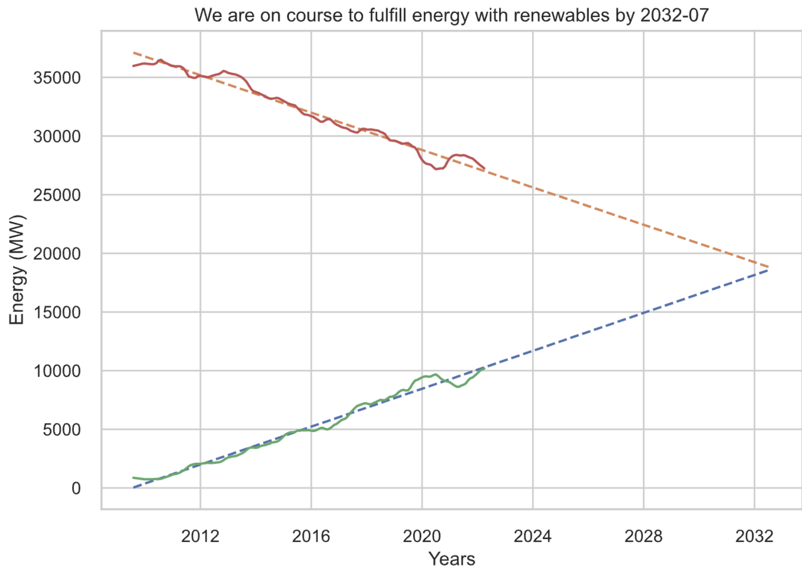
## Phase Three - When Will Renewable Energy Generation Fulfil Demand?:

We saw earlier in our EDA that Energy Consumption was dropping while Renewable Generation was increasing at a constant rate, ie. both of these datasets are converging.

As a thought experiment, let's see if we can isolate, based on the current trajectory of both datasets' trends, exactly when we will be able to fulfil all of our National Energy Demand with Renewable Energy. It should be noted that there are many external factors that will affect our ability to accurately forecast far into the future. This includes macro-economics, politics, technological advances (cost of solutions). The below does not claim to give the real answer, but will be helpful in contextualising our trajectory as a nation in becoming self-sufficient and 'green'.

For this phase of the project we won't be using our monthly averaged data because we are forecasting many years into the future. Instead by taking the annual mean for both data sets, we decompose the data in order to isolate the trend . This gives us the ability to eliminate the noise and leaves us with the overall data's trajectory, which we then solve for, using linear regression.

As per the above visualisation, we are on course to fulfil total energy demand with renewables by July 2023. That said, should take this with a pinch of salt, simplifying such a trend 10 years into the future using a linear trajectory is most likely an oversimplification of what should most likely be a polynomial trajectory into the future.




## Future Improvements:

With more time, we could improve the final prediction by adding in a margin of error, or use a stochastic method to simulate different outcomes in the convergence of the supply and demand. We also could have put more effort into building out a SARIMAX model that utilised the exogenous variables, ideally such a model would perform the best by predicting the covid summer dip in demand. More of an importance could have been put on AIC as an evaluation metric, in combination with cross validation techniques to optimise the model's parameters.