**Energy Demand Forecasting – Milestone Report**

1. **Problem Statement**

**Why useful Questions?**

**For Whom?**

**Statement: *Forecasting Total Weekly Energy Demand for Spain using weather features***

The more accurate this prediction is, the better it is for humanity. The stakeholders that will directly benefit from the solution to the above statement are as follows:

1. Energy companies
2. Ministry of energy.

But, it will be indirectly beneficial to the rest of the world. How you ask?

1. If energy companies can accurately forecast energy consumption in the future, they will know exactly how much they need to produce.
2. If they produce too much, the grid can be oversupplied and fail. This is wasted energy, incurred repair costs, incurred energy which is passed on to you and me. It would be an emergency.
3. If they produce too little, there will be major power cuts. They would have to rush and pay tons of overtime costs to get power up.
4. If energy companies can forecast energy demand sufficiently, the distribution of energy between fossil fuels and renewable energy sources can be optimized and we will have a much greener planet.

A good forecast of total weekly energy demand will ease the above problems and that is the main aim of this project.

1. **Dataset description**
2. **Where did the data come from?**

This dataset was located on Kaggle. <https://www.kaggle.com/nicholasjhana/energy-consumption-generation-prices-and-weather>

The dataset has 4 years of hourly energy demand, generation by source, energy price and weather data. The energy dataset is representative of all of Spain. The weather dataset represents the 5 most populated Spanish cities. The energy data was collected from ENTSOE a public portal for Transmission Service Operator (TSO) data. Weather data was purchased by the creator of the dataset and made public on Kaggle by himself.

1. **How was the energy dataset wrangled?**

The following features were found in the energy dataset under a datetime index:

1. Energy generation data from 21 different sources.
2. Forecasted energy generation from solar and wind energy
3. Total Load data (‘total load actual’)
4. Forecasted Load data (‘total load forecast’)
5. Forecasted and actual energy price

The following features were needed from this dataset:

* Total Load data (‘total load actual’)
* Forecasted Load data(‘total load forecast’)

Our model will be aimed at predicting energy demand 1 week in advance so therefore energy price and generation will not be of use in our analysis. But prior to removing the other features, the generation data was examined and findings will be reported in the EDA section.

*Missing Data:*

The ‘total load actual’ columns seems to have 36 hours of missing data. Some of these hours are consecutive hours which if deleted can wipe out half a day worth of data. The dataset has 4 years of data and hence only 4 datapoints for each hour. It would not be suitable to delete this data. Load data is dependant on the following:

1. The hour of day
2. The day of the week
3. The month of the year

A good value to account for the above 3 above points would be to replace missing values with the average of 6 values that represent:

1. 3 weeks of future values and 3 weeks of past values
2. Same day of week (eg if the outlier is a Monday
3. On the same hour of day

Example: If the missing value is at 3pm on Monday 15th January 2015. The value would be replaced by the average of 3pm for the last 3 Monday’s and the following 3 Monday’s.

1. **How was the weather dataset wrangled?**

The following features were found in the dataset under a datetime index:

1. Maximum, minimum and average hourly temperatures in degrees Kelvin.
2. Pressure (hPa)
3. Humidity(%)
4. Wind speed(m/s)
5. Wind Direction (degrees)
6. Rain amount for the prior 1 hour and 3 hours(mm)
7. Snow amount for the prior 3 hours
8. Clouds covering the sky (%)
9. Weather description for the hour (Codes)

Based on our first EDA, the following features are needed from this dataset:

* Maximum, minimum and average hourly temperature converted to degrees Celsius as it is more relatable to everyday use.
* Pressure (hPa)
* Humidity(%)
* Wind Speed(m/s)

*Outliers:*

1. **Pressure**

The Pressure column has abnormally high maximum values which seem impossible. This is a data error.

- 1000 hPa is standard atmospheric pressure in the air.

- There were observed values 10,000 hPa and 1,000,000hPa.

As a reference, 10,000 hPA is equivalent to a Sedan car being supported on the palm of a human hand. Imagine 1,000,000 HectoPascal.

- The highest and lowest pressure recorded on earth is 1084hPa and 870hPa respectively.

Values above 1080 and below 870 shall be replaced with the mean of that exact date where values are less than 1080 hPa.

1. **Wind Speed**

The wins seed column has an abnormally maximum value which is impossible.

- Fastest wind speed ever recorded on earth is 103 m/s and a category 5 hurricane is 70m/s.

Any wind speed above 70m/s will be replaces with the mean of the day of year over 4 years.

1. **EDA findings**

This section was structured by asking a series of questions and answering them with visuals. This section aims on answering questions related to demand and weather features only. The repository contains answers to additional interesting questions that trigger curiosity but.

**1. How did annual energy demand Change between 2015 and 2018?**

**A close up of a map

Description automatically generated**It was observed that energy demand is steadily increasing at a decreasing rate between the periods of 2015 and 2018. The rate of increase has decreased from 1.1% in 2015 to 0.7% in 2018.  **2. How does energy demand vary by month?**

We see a similar pattern for every year with a cycle of 2 highs and 2 lows. Highs seem to occur during January(Winter peak) and June/July (Summer peak) with lows during April and October(Spring and Fall).

**A close up of a map

Description automatically generated3. How does energy demand change across the week?**[**¶**](https://render.githubusercontent.com/view/ipynb?commit=93ccd6aa75bfe5f56c2624cae4aeaa284b2a9ef6&enc_url=68747470733a2f2f7261772e67697468756275736572636f6e74656e742e636f6d2f792d6661777a792f656e657267792d64656d616e642d666f7265636173742f393363636436616137356266653566353663323632346361653461656161323834623261396566362f6e6f7465626f6f6b732f312d6564612e6970796e62&nwo=y-fawzy%2Fenergy-demand-forecast&path=notebooks%2F1-eda.ipynb&repository_id=227481134&repository_type=Repository#3.-How-does-energy-demand-change-across-the-week?)

We see a similar pattern during the 4 seasons.

Energy demand peaks Wednesdays, decreasing all the way to Sunday with a significant decrease on Saturday (Weekend). Then a sudden increase on Monday(start of work week).

We do see a sharp decrease during winter on Thursdays and increasing to Friday but it unknown why this occurs.

**A close up of a mans face

Description automatically generated**  
**4. How does intraday hourly demand change?**

**Summary?**

**Visuals and stats to support?**