# PREDICTING ELECTRICITY PRICES AND QUANTIFYING DEPENDENCE ON RENEWABLE ENERGIES

A report submitted in partial fulfilment of the requirements for the degree of
Master of Science: Data Science

by
Thara Pappan Selvam
52211590

Under the guidance of
Mamen Romano Blasco and Ekkehard Ullner

School of Natural and Computing Sciences
University of Aberdeen
August 11, 2023

## Abstract

This project aims to create a model that will predict energy prices and examine the dependence on renewable energies in the UK over the next ten years using historical data obtained from Elexon and National Grid. The cost of electricity does, however, fluctuate considerably daily. To address this, statistics were gathered on energy production, demand data, market index data, and system price. To find patterns and relationships between the variables, our method required studying time series data. I created a predictive model that can anticipate power costs for each half-hour period using the data from this investigation.

Two robust machine learning algorithms—Artificial Neural Networks (ANN) and Random Forest Regression—were used to produce reliable predictions. These algorithms, which use historical pricing data to train the model, can make future forecasts more accurate. The time series analysis (seasonality), correlation between the variables, and prediction of the market price in this study have broad ramifications for the energy industry. Suppliers can improve their trading and risk management methods by offering accurate estimates of electricity prices. Additionally, being aware of the relationship between renewable energy sources and electricity prices enables effective planning and integration of renewable energy, resulting in sustainable energy practices. This study concludes that the Random Forest Regression predicted a score of 0.86 R2. In therefore, suppliers and other participants in the energy industry can adopt this tool to more accurately predict costs.

# 1. Introduction

The cost of electricity in the UK is affected by various factors, such as the state of the weather, patterns of demand, and the availability of renewable energy sources. Renewable energy sources do not provide electricity at a consistent rate. Electricity prices might change depending on how much energy is produced. This research aims to create machine-learning models to forecast UK pricing for upcoming settlement periods. Informed decisions on electricity trading, demand response, and energy source planning are made possible thanks to the project's support of providers, consumers, and other participants in the energy market. An essential component of daily life, electricity is used to generate, transport, deliver, and consume energy.

Elexon gathers information on the energy generated by the generators and the actual energy used by customers after each half-hour settlement period. Then, it compares this information to the volumes that were agreed upon in the agreements made between the energy suppliers and the power-producing businesses. The imbalance is the difference between the actual and agreed-upon volumes [1].

The National Grid Electricity System Operator (ESO) uses balancing services to control these imbalances and stabilize the grid. To ensure that the energy system has the appropriate amount of electricity available to fulfil demand, these services assist in increasing or decreasing electricity generation or consumption in real-time. Elexon uses the information on imbalances to determine the price of power. The power price will increase in direct proportion to the size of the discrepancy between actual and agreed-upon volumes. This pricing structure encourages energy provider structure to forecast demand more precisely and supports a balanced and steady electricity market [2].

## 2. Background Information About Electricity Marking.

Elexon determines electricity prices in the UK, and the electricity is generated by power generation companies in the UK. Energy suppliers have to predict how much energy customers will consume in half an hour, called the settlement period. Based on the prediction, the suppliers agree to buy electricity from power generation companies. Here Elexon plays an important role. Suppliers should inform Elexon which kind of contract is made. Elexon determines how much generators have produced and how much the customer has consumed [3]

### 2.1 What Exactly Are Settlement Periods?

Electricity is traded in half-hour 'chucks' in the United Kingdom. These half-hour intervals are known as Settlement Periods. Each Settlement Day is divided into 48 Settlements Periods, with the first one corresponding to 00:00 to 00:30, the second to 00:30 to 1:00, and so on until it reaches Settlement Periods 48 (23:30 to 00:00). Settlement Periods are always in local time. Everything about trading and the generation of market price is predicated on the Settlement Period [4].

### 2.2 Calculate the Imbalance Cashflows.

- If consumers have used more electricity than was contracted, the additional electricity is sold from the system at **System Buy Price**.
- If the energy generator generates more electricity than was contracted the additional electricity is sold to the system at **System Sell Price**.[4]

## 3. Research Questions.

- Which model will give better results when predicting electricity prices?
- How will electricity prices change over the next decade?
- What is the relationship between Renewable energy and market electricity price?

# 4. Tools

## 4.1 Python

Python has become the industry’s most widely used language for data analysis thanks to its adaptability and simplicity. The abundance of its library ecosystem, which includes pandas, NumPy, and Matplotlib, enables data scientists to quickly analyze and comprehend complex information. Python's popularity is due to its versatility, simplicity of data manipulation, and capacity for illuminating visualizations. Python offers a flexible and powerful platform for data-driven projects, whether exploratory data analysis, statistical modelling, or machine learning, making it the preferred option for people and companies wanting to uncover insight information from their data.

### 4.2.1 Library

The following are the libraries I used in the project for analysis and prediction.

**Pandas**

- Most people prefer using Python's Pandas library to manipulate data because it is a powerful and well-liked tool for analysis. I used it to import my CSV data files, Excel sheets, and SQL. Also, I used this library to create a new data frame.

**NumPy**

- For handling arrays and matrices, I used NumPy in this project.

**Seaborn and Matplotlib**

- In my project, I used both libraries for data visualization to create line plots, scatter plots, bar and histograms, and heatmaps.

**DateTime**

- This library is used to handle the dates and time intervals of the settlement date in this project.

**Sklearn**

- It is a powerful machine-learning library. I applied this package to my linear Regression and Random Forest Regression models.

**Sklearn. metrics**

- This library provides a wide range of matrices for assessing the accuracy of the model's predictions. A function for calculating the difference between predicted and actual values.

## 4.2 MySQL

The various datasets were imported into a single table using the MySQL server. I constructed a table and loaded all the data into it, so it would be easy to analyze and can retrieve the desired data based on the query.

A schema with the name "electricity predict" was created. Tables were created with information on demand and energy source generation. The Load query option in the MySQL query browser was utilized to import the data.

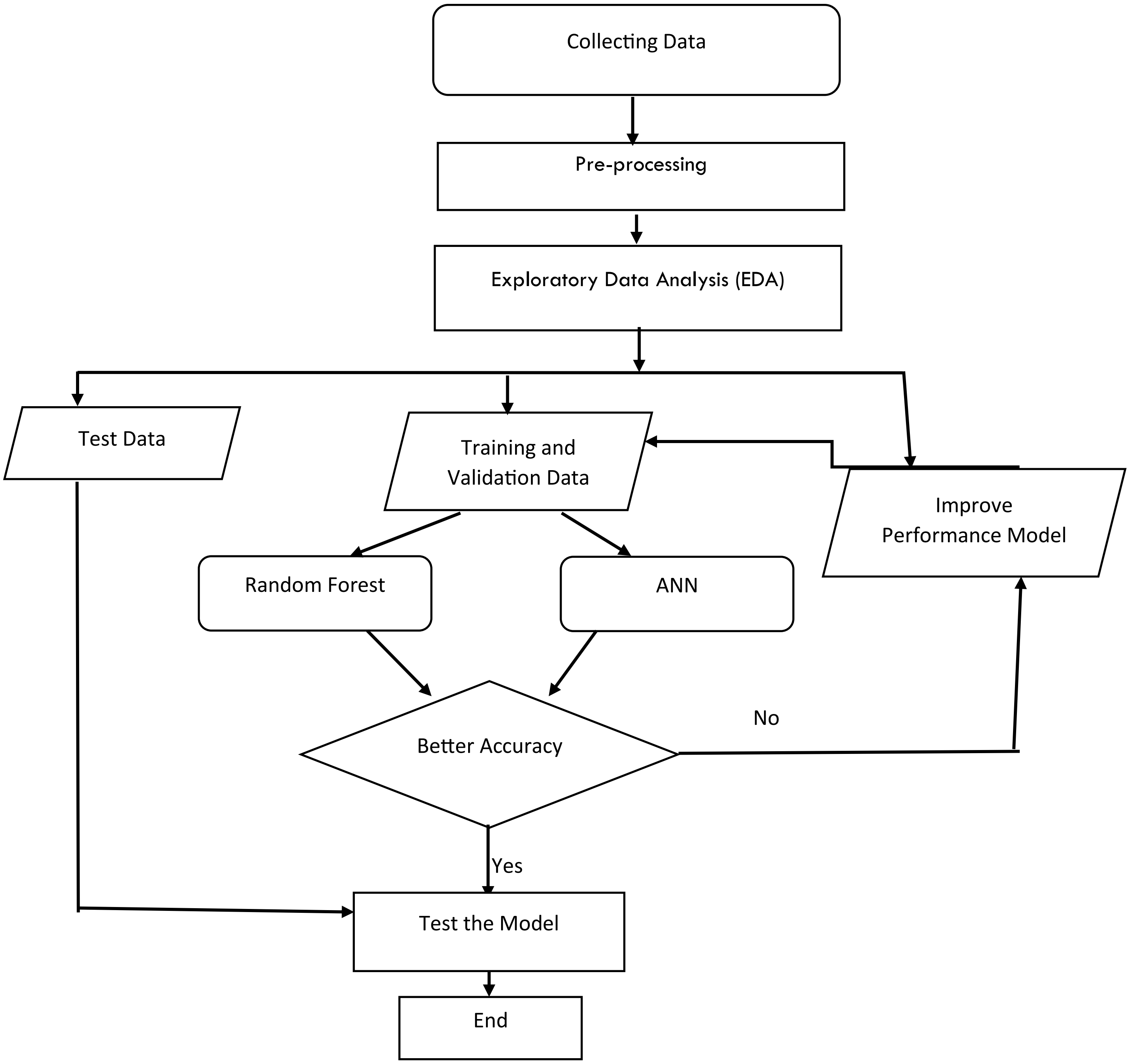
## 5. Data Access

I analysed data from Elexon and National Grid (Electricity System operator), and National Grid ESO data on demand and energy sources [2]. They offered helpful information on electricity consumption, storage capacity, and energy source for each settlement period (half an hour) in the UK.

One of the data sets used in this study came from Elexon. This service supplied Market Index Data, Best View Prices, System Prices, and Net Imbalance. Market Index Data (MID) was utilized to calculate the Reverse Price for each Settlement Period and reflect the wholesale electricity price in the short-term market. Prices for the best views are calculated every half-hour over 24 hours. Prices are quoted in pounds per megawatt-hour. The System Sell Price (SSP), System Buy Price (SBP), and Net Imbalance volume (NIV) are determined in line with the Balancing and Settlement Code. We can identify the seasonality and the influence of electricity price swings by using historical system price data [1].

## 6. Methodology

### Flow Chat



## 6.1 Exploratory Data Analysis

EDA is to analyse the data, understand the pattern, and present the data in graphical representation. For EDA representation in this project, I used Jupiter's notebook.

### 6.1.1 Data Cleaning and Pre-processing

**Handling missing values**

The most important part of pre-processing is this step it guarantees that our data is perfect for analysis. In general, we want to validate null and empty data.

**Standardizing Formats**

For analysis, I converted the data format. The DATE-TIME column in the energy source data contains time zone information, so I removed it and eliminated the Demand, Market Index, and System price data sets. I also modified the Settlement Period field from data to string. Then, in the Data Frame, I added one new column, using the lambda function, and converted each string data to a Date Time object in the **'%Y-%m-%d'** format. Again, a new Time Interval column was created for the dataset using the settlement period. The same steps were taken for all three data frames.

**Data Transformation**

Non-numerical data is called categorical data, for example (converting string to number). It is necessary to avoid any bias in the algorithms. In this project I converted three independent variables Hydro_perc, Solar_perc, and Wind_perc into numeric, as previously it was in the object, making it find the correlation before the prediction model.

### 6.1.2 Data Visualization

Data visualization is used to explore the data, patterns, and outliers. In this project, I used the 2014 to 2023 periods of data. Data on energy supplies, demand, market index, purchase, and system price were imported for the initial analysis. A box plot was used to analyse the energy source data, enabling a general comprehension of the data's distribution and detecting any potential outliers. This process was essential for comprehending the energy source dataset's variability and central tendency and provided critical data for subsequent studies.

Energy Source from 2014 to 2023 for each settlement period between 2014 – 2023, as shown in Fig 1.

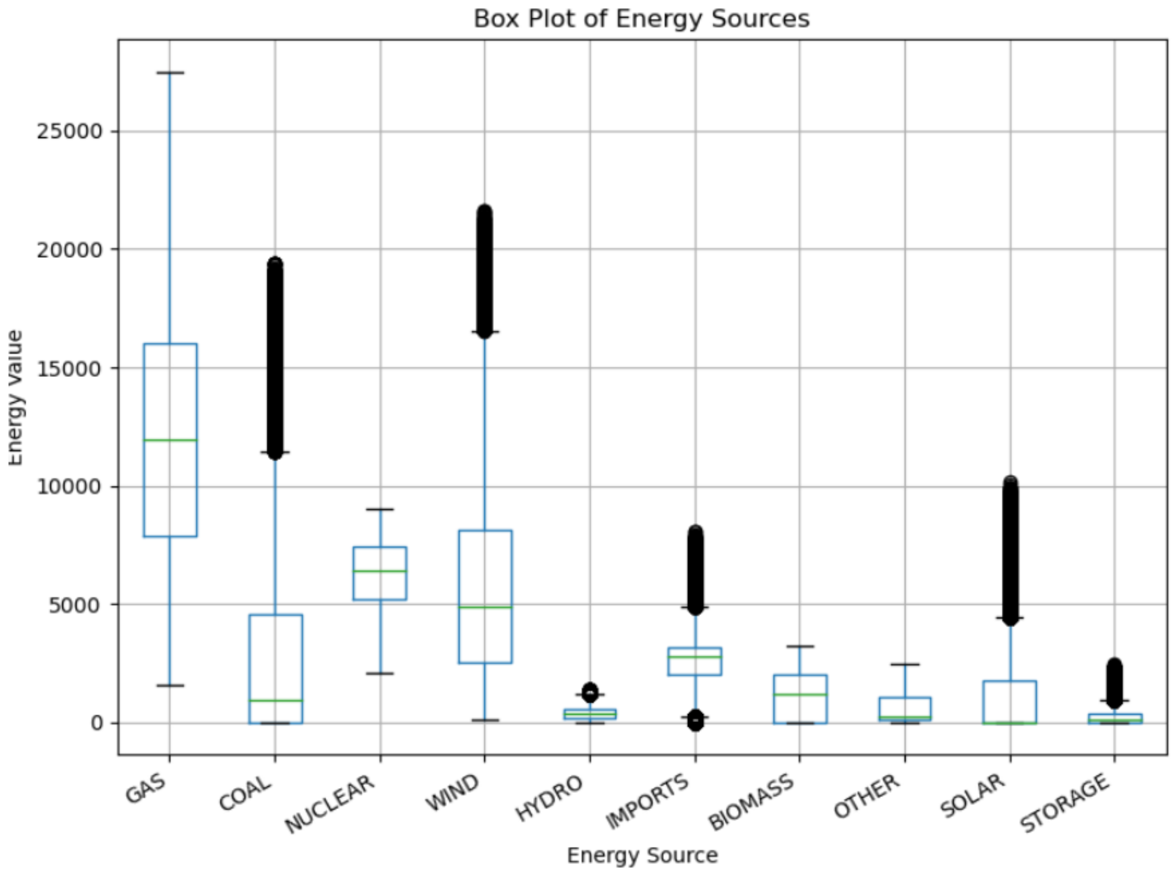



**Figure 1: Power generation in box plot of 2014 to 2023**

National Demand and Renewable Energy (Solar and Wind) Generation, to compare the trends of electricity demand and generation over the years.

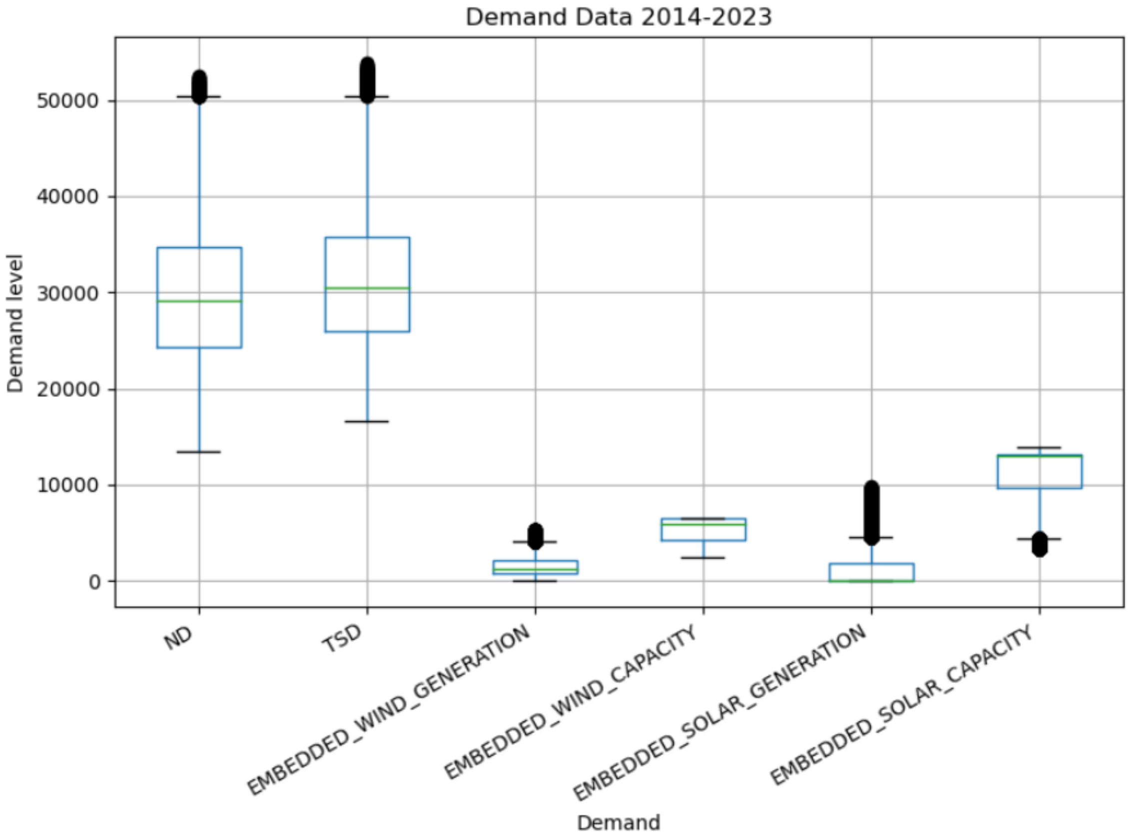



**Figure 2: Demand data box plot of 2014 to 2023**

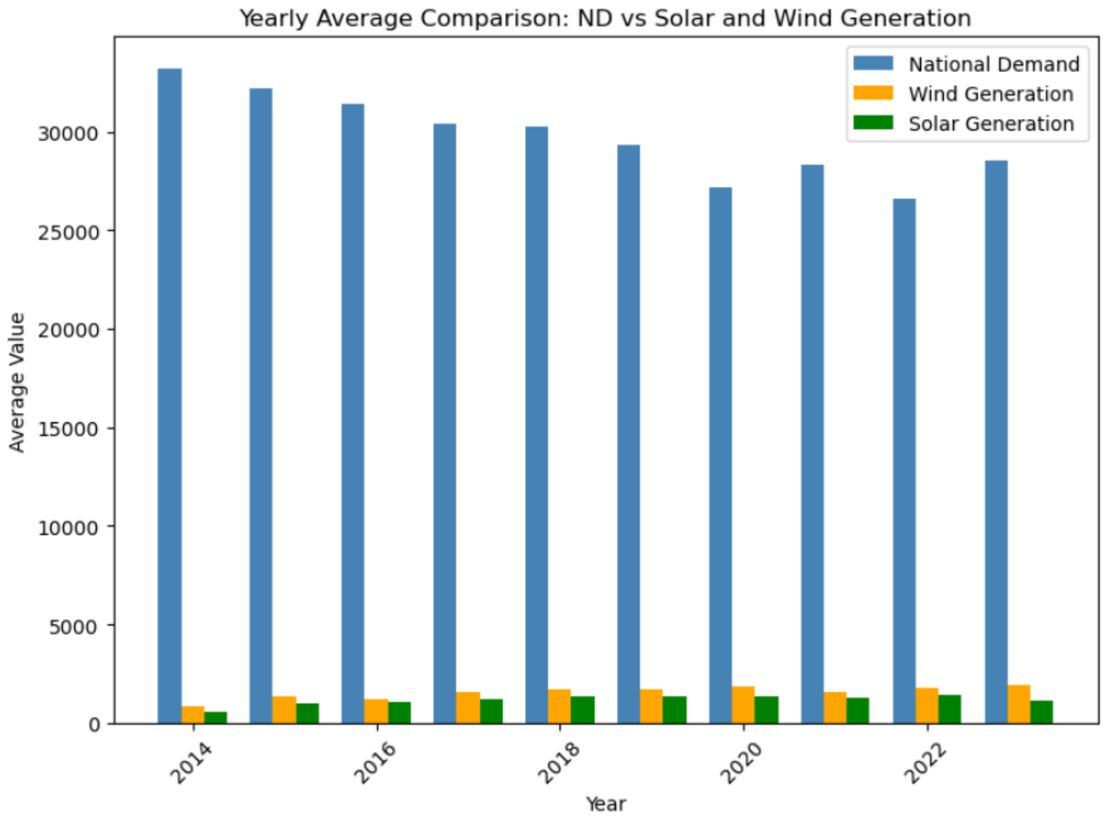



**Figure 3: Wind, Solar vs National Demand 2014 to 2023**

Renewable percentage data shows the pattern of energy generated during 2014-2023. (Renewable percentage sum of Wind, Solar, Hydro)

Year by year, the percentage of Renewable source generation is increasing. I also analyzed the data during covid, and Ukraine war period as shown in Figure 4.

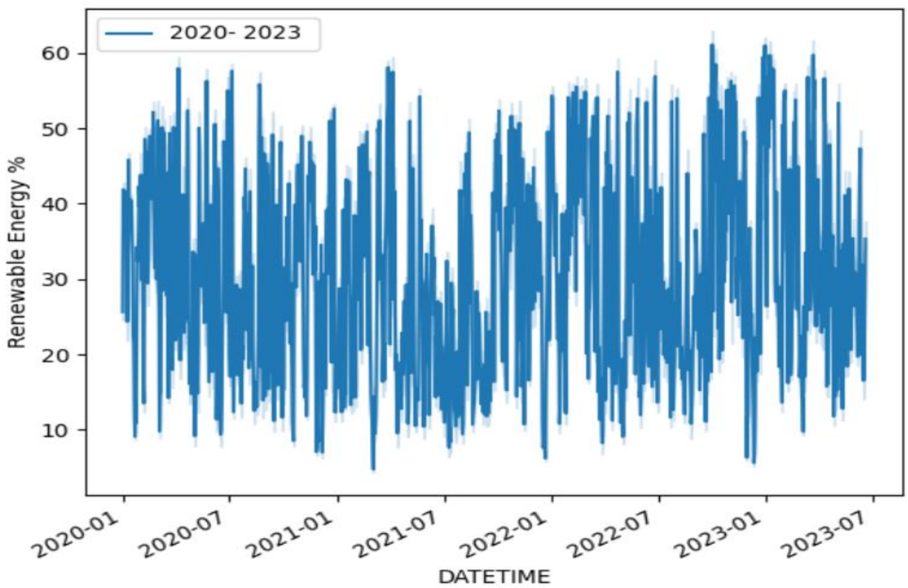



**Figure 3: A renewable percentage from 2020 to 2023**

Using a line plot, the buy and sale data from 2014 to 2023 were plotted to visualize patterns and price variations over time. Notably, differences between buy and sale prices were seen between 2014 and October 2015, while after that period both prices are the same.

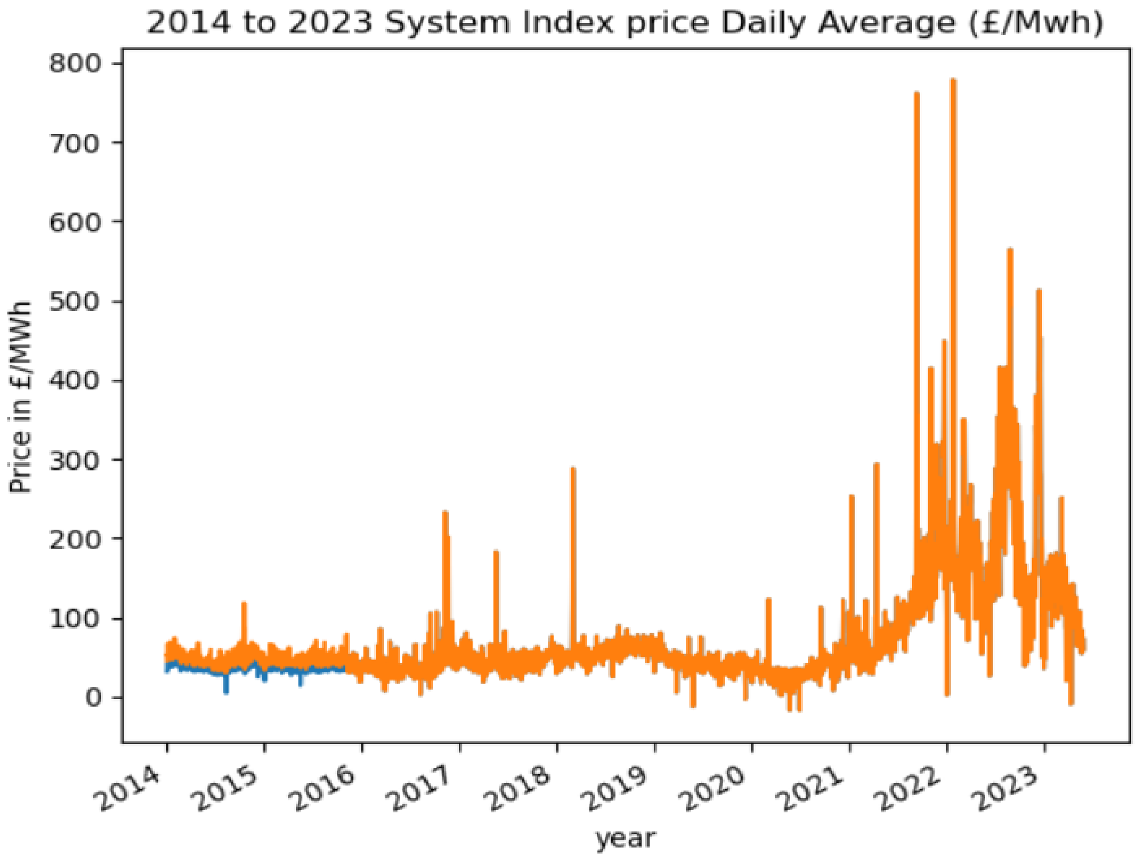



**Figure 5: System price daily average**

Figure 5 shows the 2014 to 2016 system sell and buy price data and the fluctuation between them. It is essential to analyses the system price data to determine how prices will change in the future.

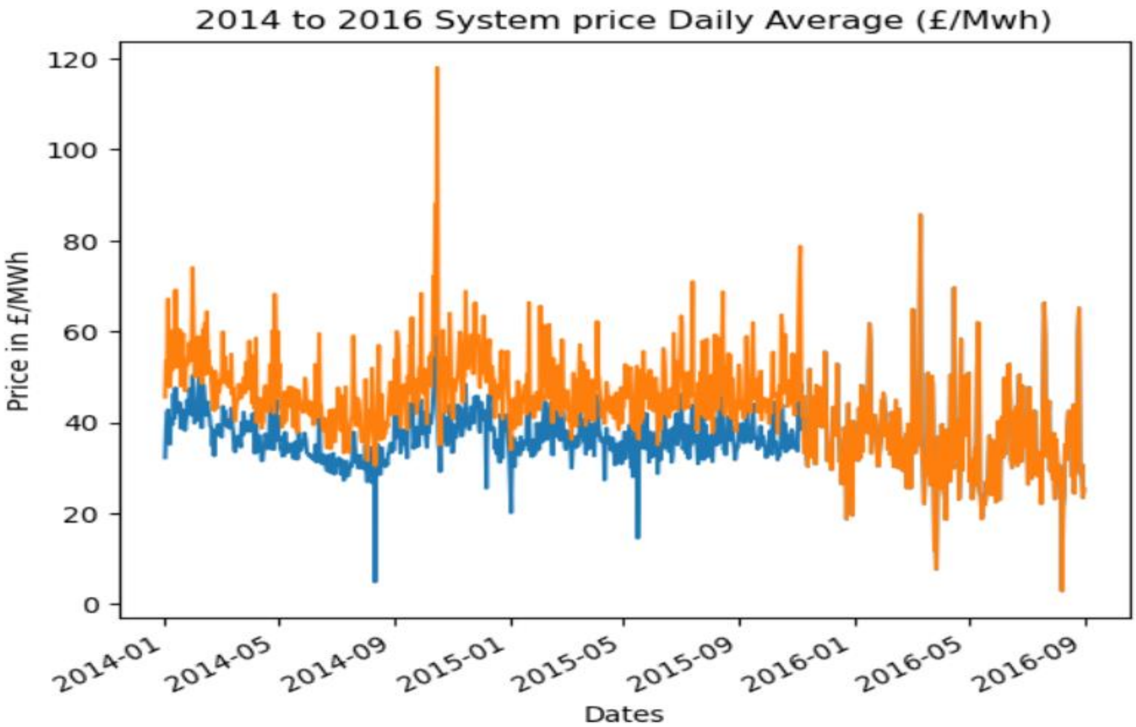



**Figure 6: System price for 2014 to 2016**

### 6.1.3 Time Series Analysis

It is important to understand patterns, trends, and relationships with data to make predictions of future values. Seasonality analysis is needed to find the sell price patterns that repeat over specific time intervals, weekdays/weekends/winter/summer.

Histogram and line plots were created using the data after it had been further divided into weekdays and weekends, as shown in Figures 7 and 8. Additionally, the data was divided into winter and summer periods, to check the seasonal changes that happened in system price. This seasonal division made it possible to investigate how various weather conditions affected energy demand and price.

Seasonal price analysis allows consumers and suppliers to identify periods of high electricity prices. It will help them to adjust their energy usage and save electricity costs. Also, It will help to understand when electricity is cheaper and more expensive.

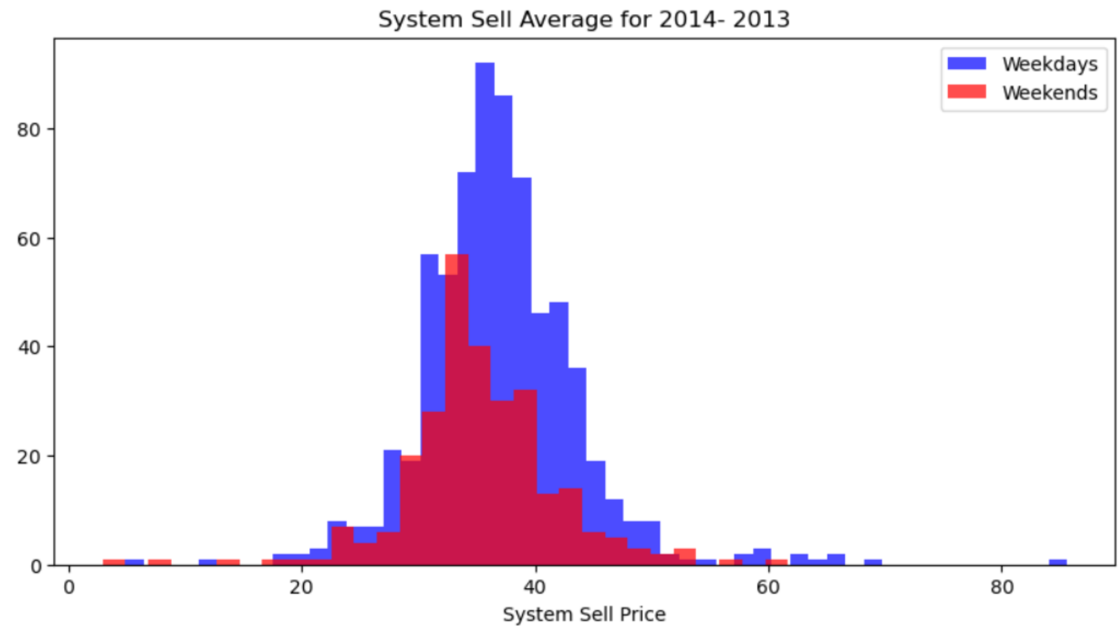



**Figure7: Histogram for selling daily average price weekdays and weekends**

This research may provide useful information for interpreting seasonal variations in consumer behaviour and market dynamics. The thorough examination of the data from the energy source using box plots, line graphs, histograms, and scatter plots revealed important trends, outliers, and seasonal fluctuations in the demand and price of energy.

It helps to understand when certain events tend to happen more often, such as increased sales during weekends or temperature changes with the seasons. Seasonal Analysis is highly important for predicting electricity prices. Electricity demand and prices often change with the weather and during holidays. It is useful to identify and understand peak and off-peak periods for electricity consumption. In Figure 7, we can see the system sell price is higher on weekdays compared with weekends.

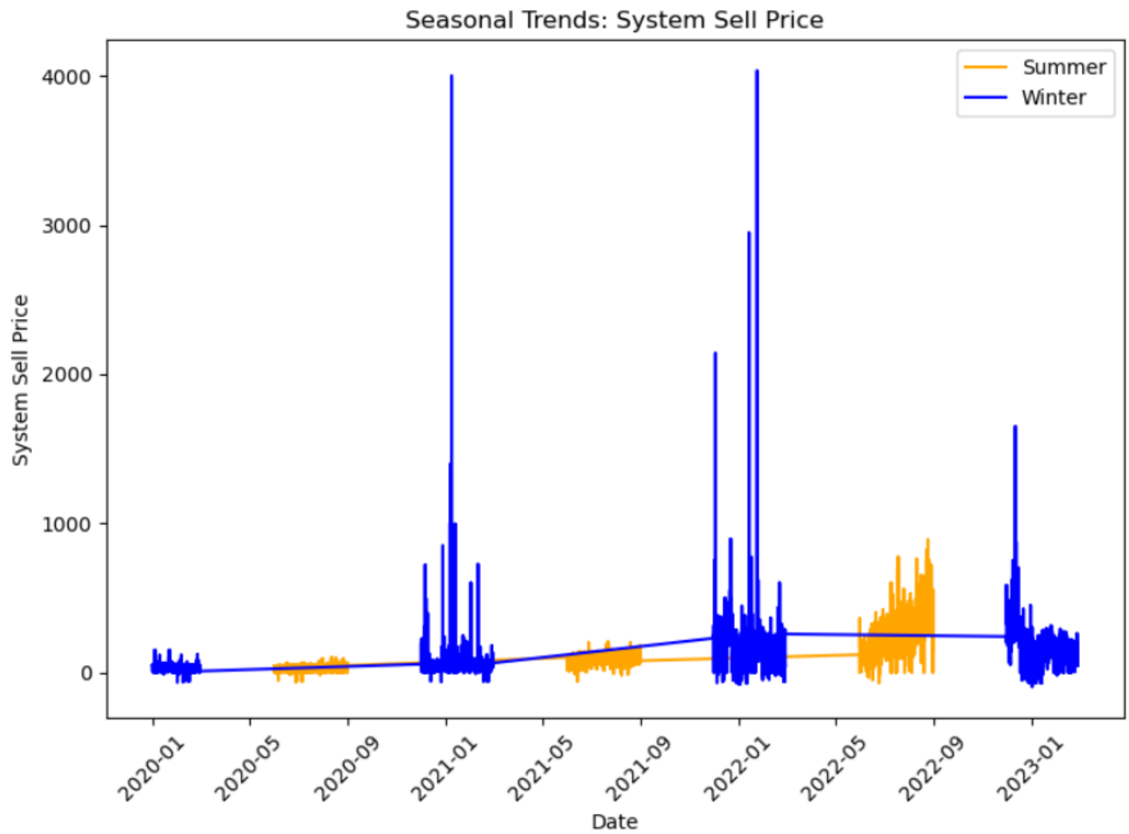



**Figure 8: System sell Seasonal comparison for 2020 to 2023**

The system sell price is highest during the winter season since consumers used more energy for heating. Winter is defined as December to February and Summer as June to August.

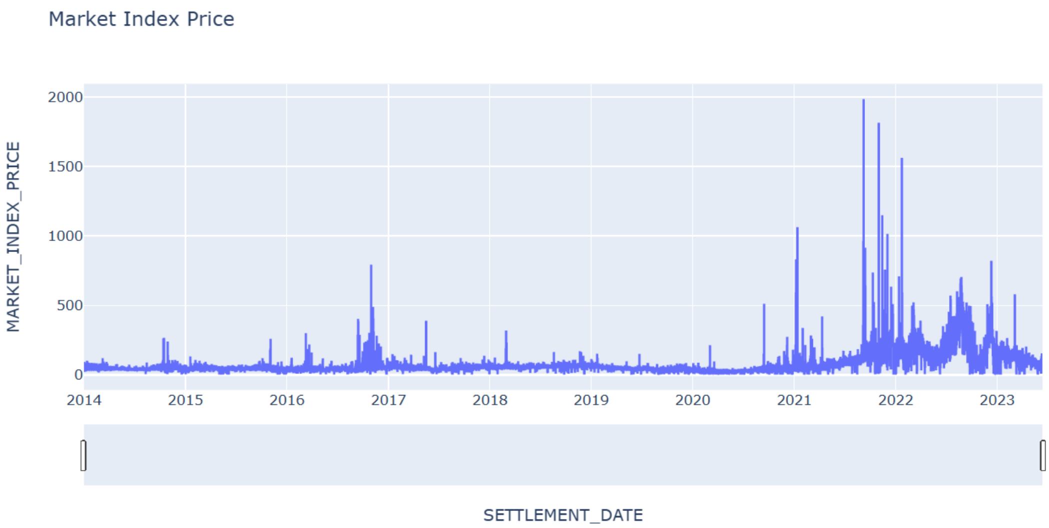



**Figure 9: Market Index price fluctuations for 2014 to 2023**

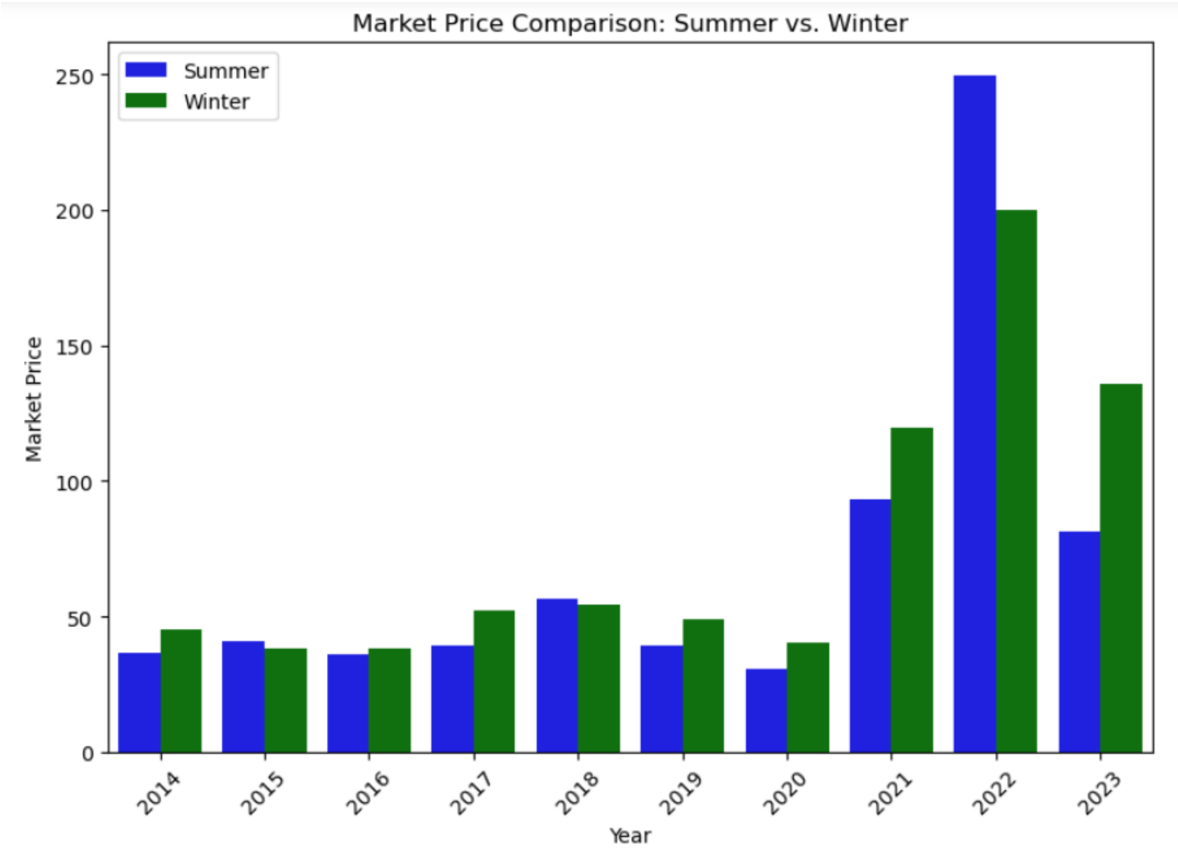



**Figure 10 Market price comparison between winter and summer**

I used the range slider visible attribute to visualize the market index price [22] as shown in Figure 11. To focus on a specific period of data, it can be moved left or right. When you adjust the slider to zoom in, the chart gives a closer picture of a smaller period. Using that, we can zoom the plot to see the COVID time Market price fluctuation. I noted that the September 9$^{th}$, 2021 Market Price was extremely high; it is the highest price in the entire data range from 2014 to 2023.

I searched why the high prices happened on the Elexon site. The reason is a continuation of rising demand, and low wind generation leading to future System price spikes. On Thursday 9$^{th}$ September, the System price reached £ 3,999/MWh in settlement periods 33 and 34. Price extended £ 1,000 /MWh for 13 Settlement Periods, the prolonged spike in System price led to a record-high daily average System Price of £ 957.77 /MWh. [4]

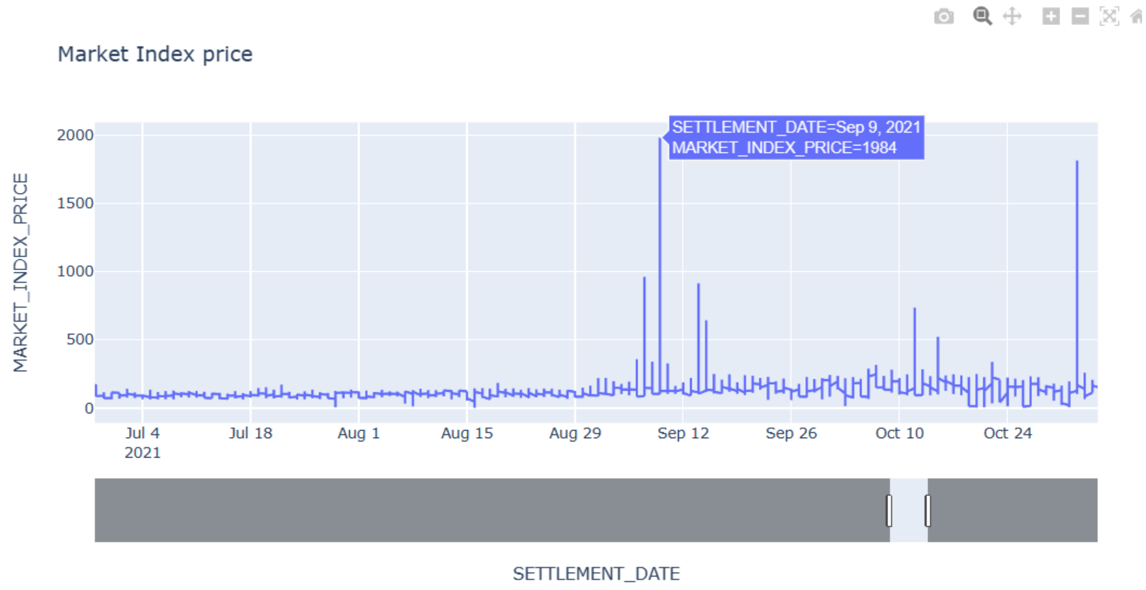



**Figure 11: Market Index price to view the Sep 9th 2021**

### 6.1.4 Correlation and Relationship

Demand data can be as understood as National Demand (ND) and Transmission System Demand (TSD). They are significantly correlated. We can utilize one of the variables to train the model for better prediction. The Sell Price and Buy price have the same values. so, we can then combine all data frames based on the settlement period to train the model.

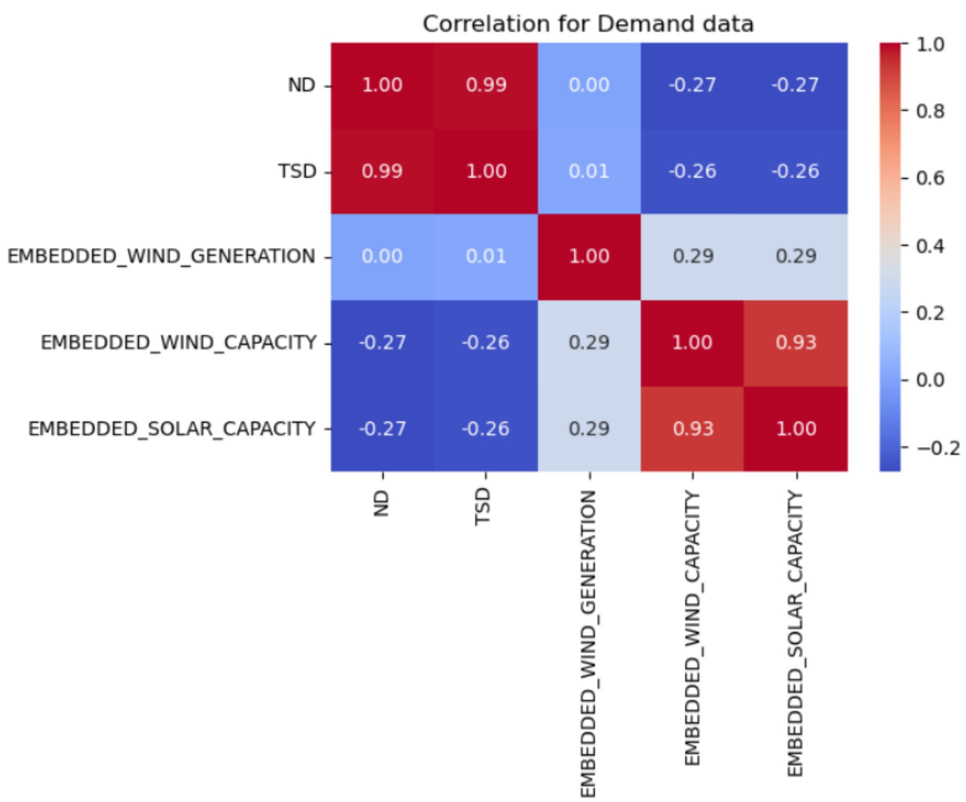



**Figure 12: Correlation of Demand Data**

### 6.2 Variable Selection

Based on the analysis of the demand, energy source, market index, and system price, to find the correlation between demand data, energy source, and market price I choose the independent variable (Market Index Volume, TSD, System Sell Price, Embedded Wind Generation, Embedded Wind Capacity, Hydro, Wind and Solar Percentage) and dependent variable (predicted variable is Market Index price) to train the model and also find the correlation. (Figure 12)

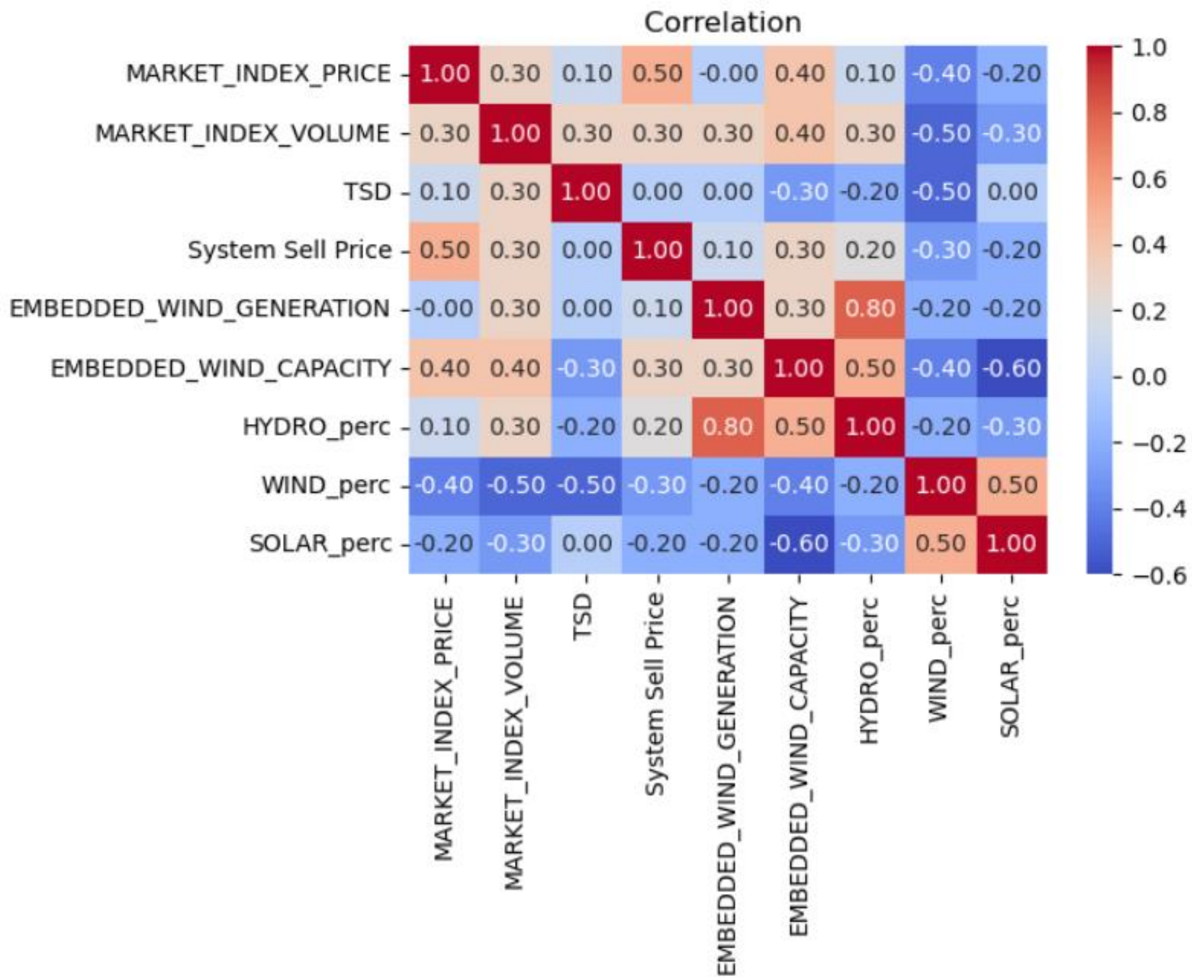



**Figure 13: Correlation independent variable and dependent Variable**

Train the model with data from 2014 to 2023. Randomly split the data into 80 % of training and 20 % of test data. I have installed the required libraries for a model.

Following are the models used to train to predict electricity prices.

- Linear Regression
- Random Forest Regression
- ANN (Artificial Neural Network)

# 7. Model and Algorithm

## 7.1 Multiple Linear Regression

It is used to forecast the value of the dependent variable (Market index price) based on the values of several independent variables. The goal is to train a predictor model with the best features for the given equation and the lowest Mean Square Error.

**y= b0+b1*x1+b2*x2+…+bn*xn**

**y** is the dependent variable (Market Index Price).

**x** is the independent variable (predictor variable).

**b 0** intercepts.

**b1, b2**, bn is the coefficient of each independent variable.

## 7.2 Random Forest Regression

Secondly, I trained my data with Random Forest Regression In comparison to other algorithms, it requires less training time. It performs well with a huge dataset and accurately predicts outcomes. Accuracy can still be preserved even though a sizable portion of the data is missing.

## 7.3 Artificial Neural Network (ANN)

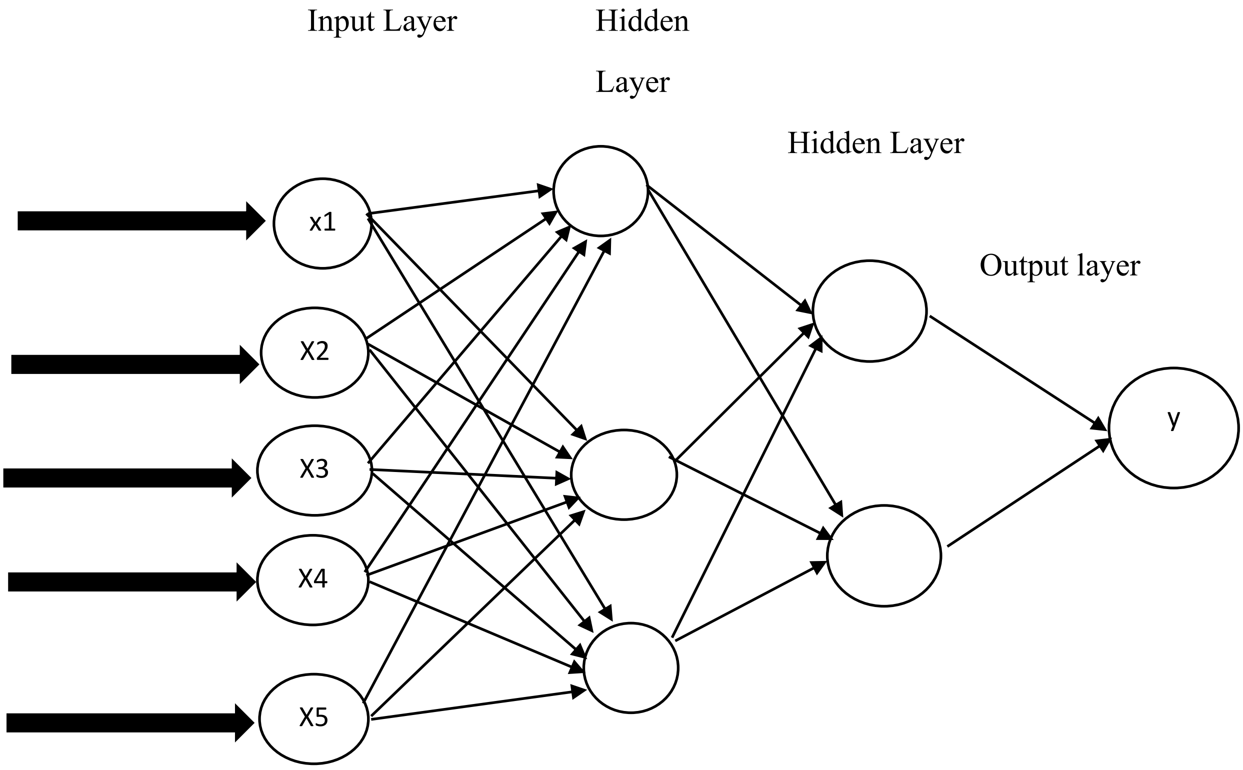
Powerful machine learning models are called Artificial Neural Networks. ANN models process data by sending information through layers of interconnected nodes, or neurons, arranged in layers of input, hidden, and output. To understand patterns and relationships in the data, the connections between neurons change their weights. The main advantage of ANN is that they can represent complex non-linear interactions, which makes linear models fail. The three-layer neural network is in the flow chart.

Keras and TensorFlow are two popular libraries for ANN models. In the ANN model, the Sequential function is used to build a succession of layers to perform one after the other in the stack. Then Each layer is defined using the Denes module, which is part of the Keras framework. It is Used to see the weight and activation function for each layer and neuron in the layer. Units are used to generate layers in neurons, for example. Suppose we provide unit 8, the values of the independent variables in the eight tiers. The same applies to all the independent variables.

The Activation Function is used to calculate within each neuron. The model uses 'ReLU' as the default activation function for regression. The size of the batch is used in the model to specify how many rows will travel through for each calculation. When it reaches that last stage of calculation for all the rows in the batch, it will call on one epoch. If the batch size is small, the ANN model will slow the processing of data. If the value is less than 10, the model will be overfitted. The model data will be processed faster if it is between 20 and 50. Epoch is used to train the data using the count specified in the epoch and to modify the weight.

An artificial neural network is made up of serval neurons connected by identical fundamental processing units. Each neuron then calculates its internal activity level by adding up its weighted inputs [6]

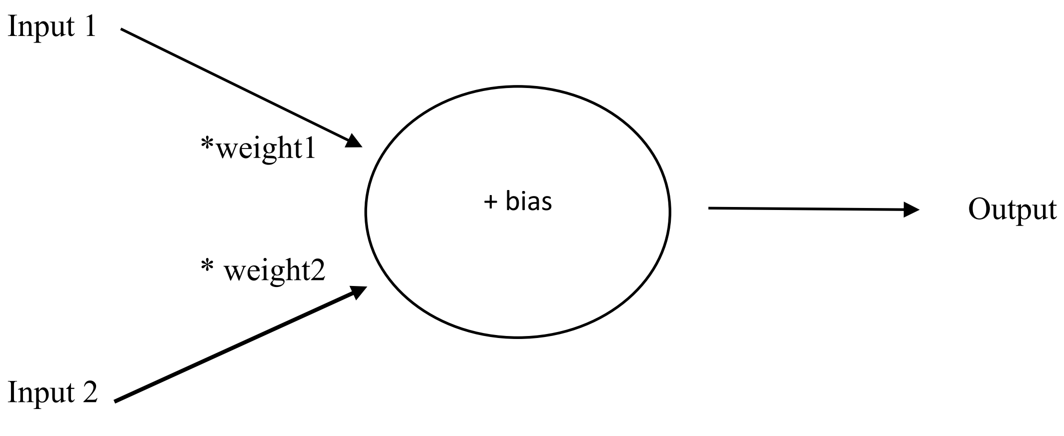
**The Generic architecture of the ANN model.**




In my model, there are eight input layers, two hidden layers and one output layer.

**Structure of single perceptron**

The input layer consists of every feature, each with an assigned weight added to the hidden layer.



Three layers make up an ANN: input layers, which accept unprocessed input data, hidden levels, which receive input from one layer and transmit output to the next, and output layers, which produce a prediction. Typically, the same activation function is used by all levels.

ANN contains three layers: input layers, which accept unprocessed input data, hidden levels, which receive input from one layer and transmit the results to the next, and output layers which produce a prediction. Typically, use the same activation function is used by all levels.

The output layer will utilize a different function than the hidden layer and depends on the prediction type required by the model. The project uses ReLU (Rectified Liner Unit Activation Function). It is the most frequent and default Activation Function for Multilayer perceptron (MLP).

**Activation Function**

Relu Function

$$\mathbf{F(x) = \begin{cases} x, & x >= 0 \\ 0, & x < 0 \end{cases}}$$

This means that if x is less than zero, it will take 0

# 8. Evaluation Metrics

## 8.1 R2 Score

R2 Score is a regression statistic that assesses how well the model fits the data. It is based on the correlation between actual and predicted values. It represents the proportion of the target variable's variance explained by the model. R2 values vary from 0 to 1. A value near 1 indicates a better model. It aids in evaluating model performance and comparing various model results. Also, the actual value was compared to the Predicted Value.

$$\boldsymbol{R2 = 1 - \frac{SS\ res}{SS\ tot}}$$

Where:

- SS res is the sum of the square difference between the observed values and the predicted values
- SS tot s tot sum of the squared difference between the observed values and the mean of the observed values.

## 8.2 Mean Squared Error

$$\boldsymbol{MSE = \frac{1}{n}\Sigma\left(y_{actual(i)} - y_{pred(i)}\right)^2}$$

Mean Squared Error (MSE) is based on the square of error. In regression problems, it is a widely used statistic for measuring the average squared difference between predicted and actual values. It measures the model's prediction accuracy with lower MSE representing higher performance. n stands for the number of data points.

## 8.3 Mean Absolute Error

MAE computes the average absolute difference between actual and predicted values. MAE, unlike MSE, does not square the errors, making it less susceptible huge errors. A small value indicates a better model. Values lie between 0 to infinity.

$$\boldsymbol{MAE = \frac{1}{n}\Sigma\left(\left|y_{actual(i)} - y_{pred(i)}\right|\right)}$$

MAE and MSE are assessment measures used to examine the accuracy of different prediction models in the context of predicting energy prices. Lower values of these indicators indicate higher performance, as they show that the model predictions are more accurate. We can select the most accurate and dependable model by comparing models using these metrics. These measures are essential for verifying model effectiveness and making data-driven decisions in the forecast of electricity prices. Figure 13 shows the model results in predicting the electricity price. In the comparison, I discovered that the expected value is inaccurate beyond 2020. The explanation for this is COVID and the war in Ukraine war. So, I split the data to retrain the model from 2014 to 2019 and 2020 to 2023. As a result, I retained the model using data from 2014 to 2019.

# 9. Result

## 9.1 First stage Analysis (2014- 2023)

In the first stage model's performance. I obtained a low R2 square value for linear regression, a high R2 Score for Random Forest regression, and an average R2 square value for the ANN model.

| Model | **Linear Regression** | **Random Forest** | **ANN** |
|---|---|---|---|
| **R2** | 0.40 | 0.86 | 0.52 |
| **MAE** | 53.00 | 11.42 | 27.45 |
| **RMSE** | 83.22 | 26.37 | 49.02 |

**Table 1: First stage predicted result for 2014- 2023**

### 9.1.1 Random Forest predicted Analysis 2014- 2023

I used a Scattered plot to analysed the predicted value, if the points are near the diagonal lines, the predictions are accurate. Figures 15 and 16 highlight the Random Forest and ANN Predicted value accuracy.

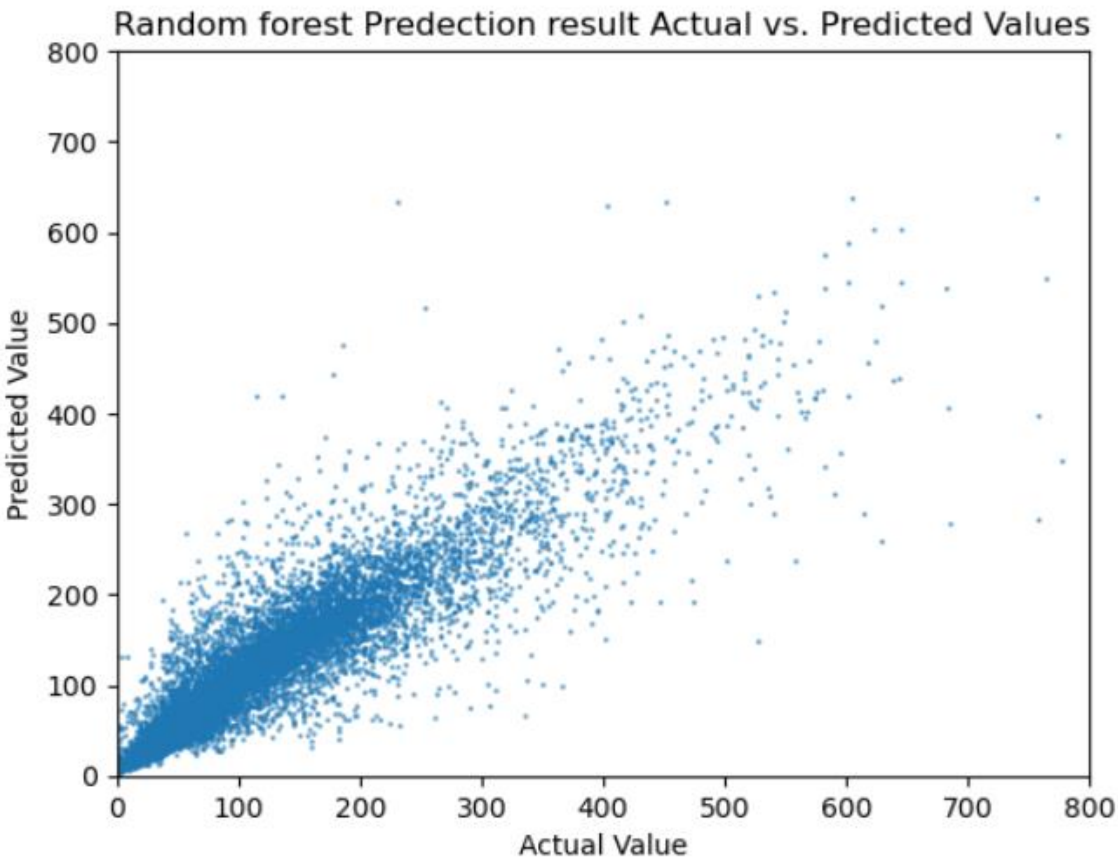



**Figure 15: Scattered plot for RFR Predicted Values of 2014 to 2023**

### 9.1.2 ANN model Predicted Analysis 2014-2023

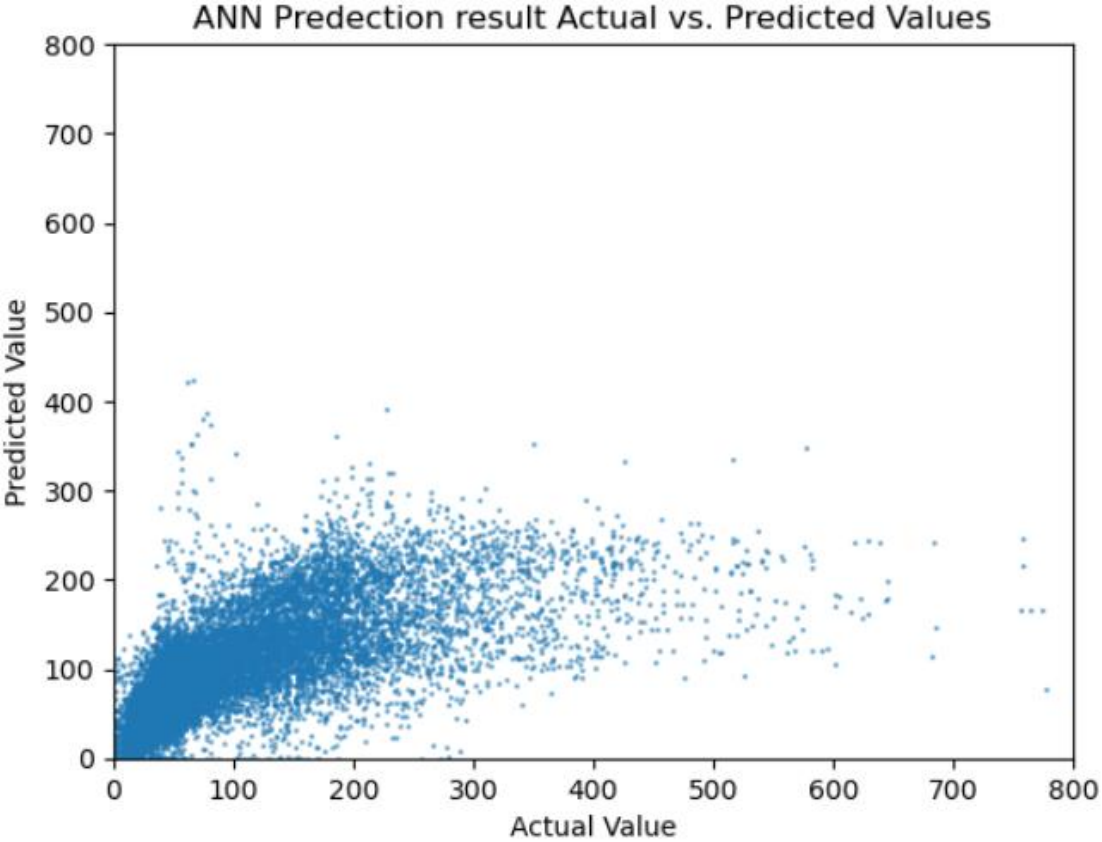



**Figure 16: Scattered plot for ANN predicted Values of 2014 to 2023**

```
mlr_diff_ann
```

| | Actual value | Predicted value |
|---|---|---|
| **130058** | 76 | 124.854073 |
| **13640** | 36 | 39.856609 |
| **49646** | 96 | 64.194199 |
| **25241** | 37 | 45.103634 |
| **76748** | 50 | 42.334888 |
| **...** | ... | ... |
| **87871** | 75 | 100.776840 |
| **60762** | 33 | 91.064682 |
| **4733** | 38 | 43.820194 |
| **43094** | 45 | 60.822735 |
| **113638** | 24 | 86.457611 |

32348 rows × 2 columns

**Figure 17: Actual vs Predicted values from ANN model 2014-2023**

The predicted value in the preceding screen is not accurate when compared to the actual value. The market price varies more after 2020; For Example, the real value is 76 GBP, whereas the predicted value is 124 GBP, as shown in Figure 17. Because of the highly fluctuating market price, I divided the data from 2014 to 2019 and 2020 to 2023. Below is an analysis of 2020-2023

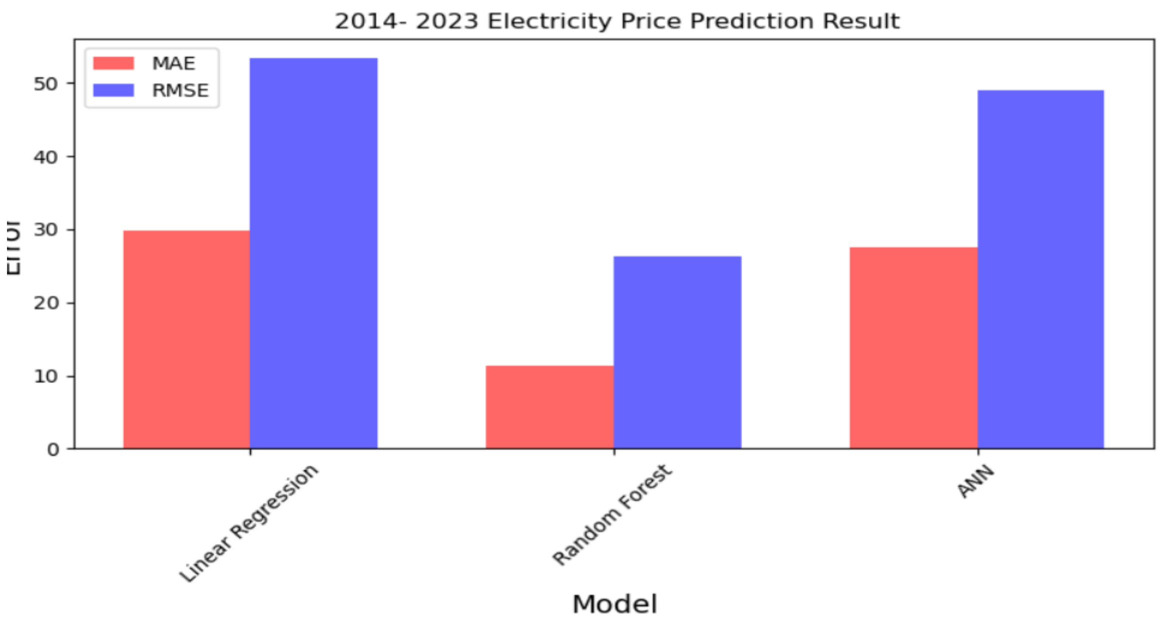



**Figure 18: Predicted Error Chat**

### 9.1.3 Analysis of the Error in Prediction Electricity Price 2014- 2023

The MAE and RMSE of the electricity price forecast are displayed in Figure 18. As previously stated, MAE will detect the absolute difference between the actual and predicted values, as well as the RMSE average square of the prediction and actual values.

In the R2 score comparison of three models in comparison to other models, the random forest model predicted an accurate value of 0.86. Random Forest outperforms other methods because it produces low MAE and RMSE values.  In the first stage, Prediction Random Forest Model Analysis yields the finest result.

## 9.2 Second Stage Analysis (2014-2019)

In the second stage of prediction, I retrained the models with 2014- 2019 period data. In the R2 score comparison of three models, the random forest model predicted an accurate value of 0.81.

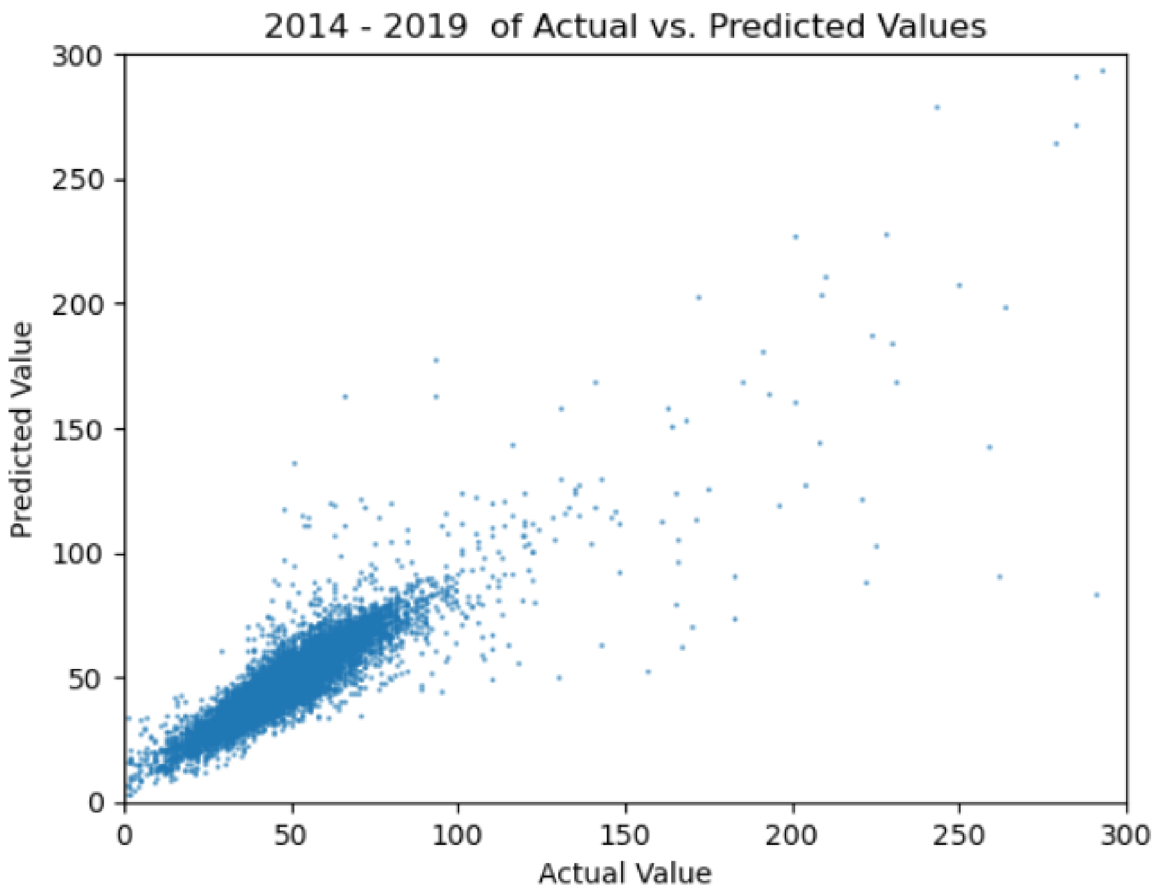



**Figure 19: Random Forest Actual vs Predicted Values 2014 – 2019**

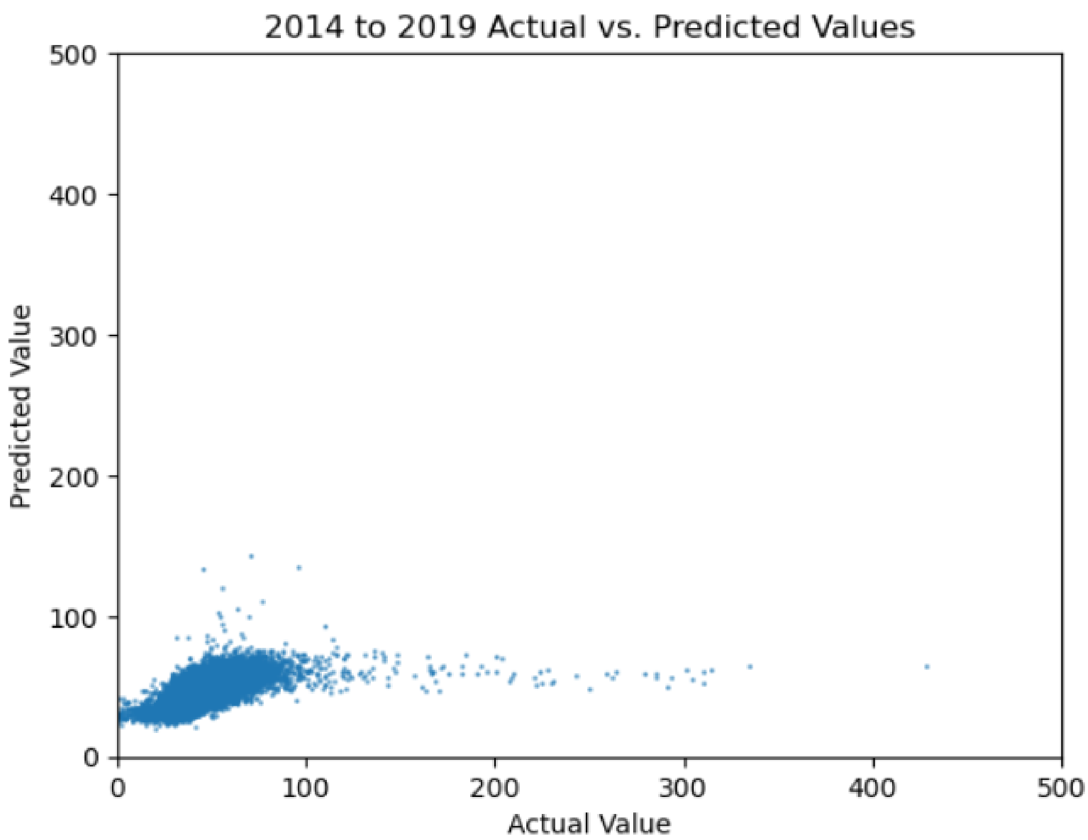



**Figure 20: ANN Predicted vs Actual value 2014- 2019**

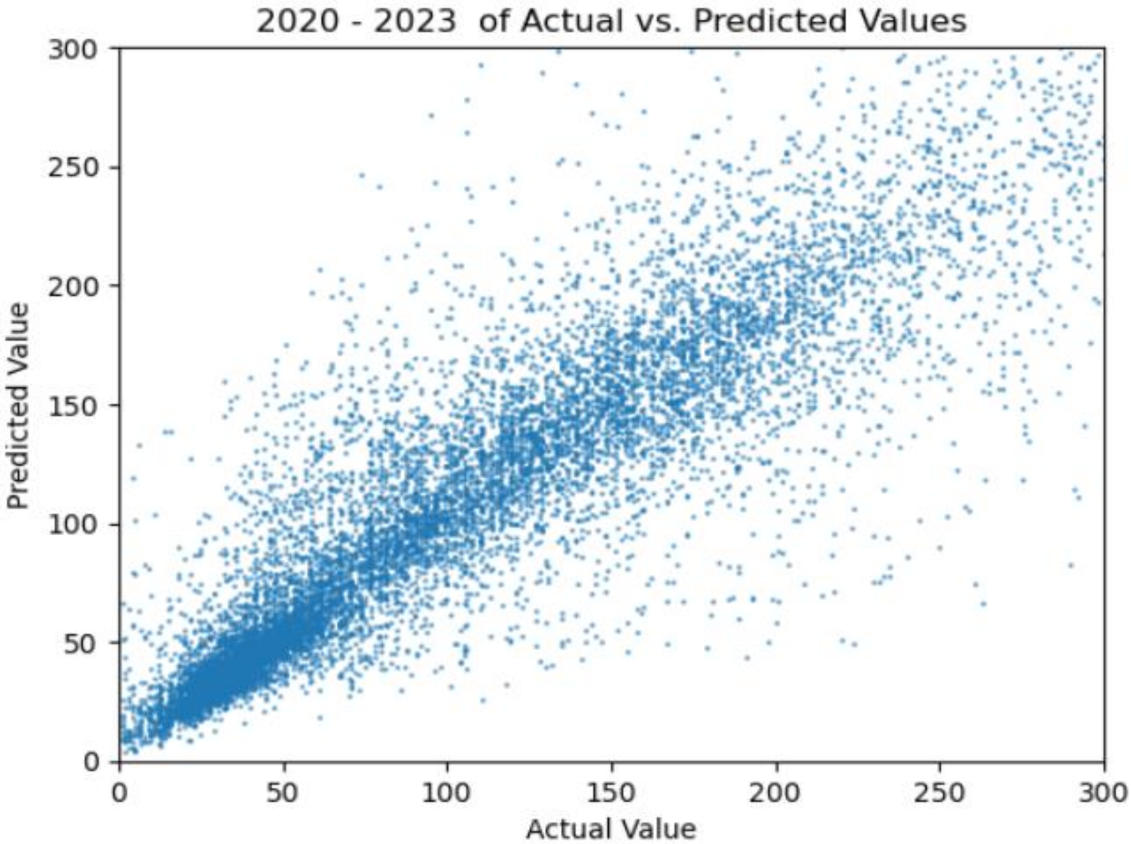
**Predicted result for 2014 – 2019**

| Model | **Linear Regression** | **Random Forest** | **ANN** |
|---|---|---|---|
| **R2** | 0.38 | 0.81 | 0.31 |
| **MAE** | 7.68 | 3.71 | 7.84 |
| **RMSE** | 13.44 | 7.37 | 14.19 |

**Table 2: 2014 – 2019 overall result**

### 9.3 The third stage of Prediction 2020-2023

I trained the model using data from 2020- 2023. The prediction result is an R2 score comparison of their model in comparison to other models; the random forest model predicted an accurate value of 0.79, the ANN model predicted an accuracy of 0.41, and the linear regression model projected an accuracy of 0.38. Random Forest predicts the best accuracy in the second stage of the forecast. I tuned the epochs to 10-50 for the ANN model and got the same result. The actual and projected values in this forecast are both reasonable.




**Random forest predicted value scattered plot Figures 21**

**9.3.1 Artificial Neural Network:** Difference between Actual value and Predicted value in 2020- 2023. Previously when trained with 2014- 2023, the predicted value was irrelevant. But the third stage of prediction got a reasonable predicted value. There is not much difference between the actual value and the predicted value.

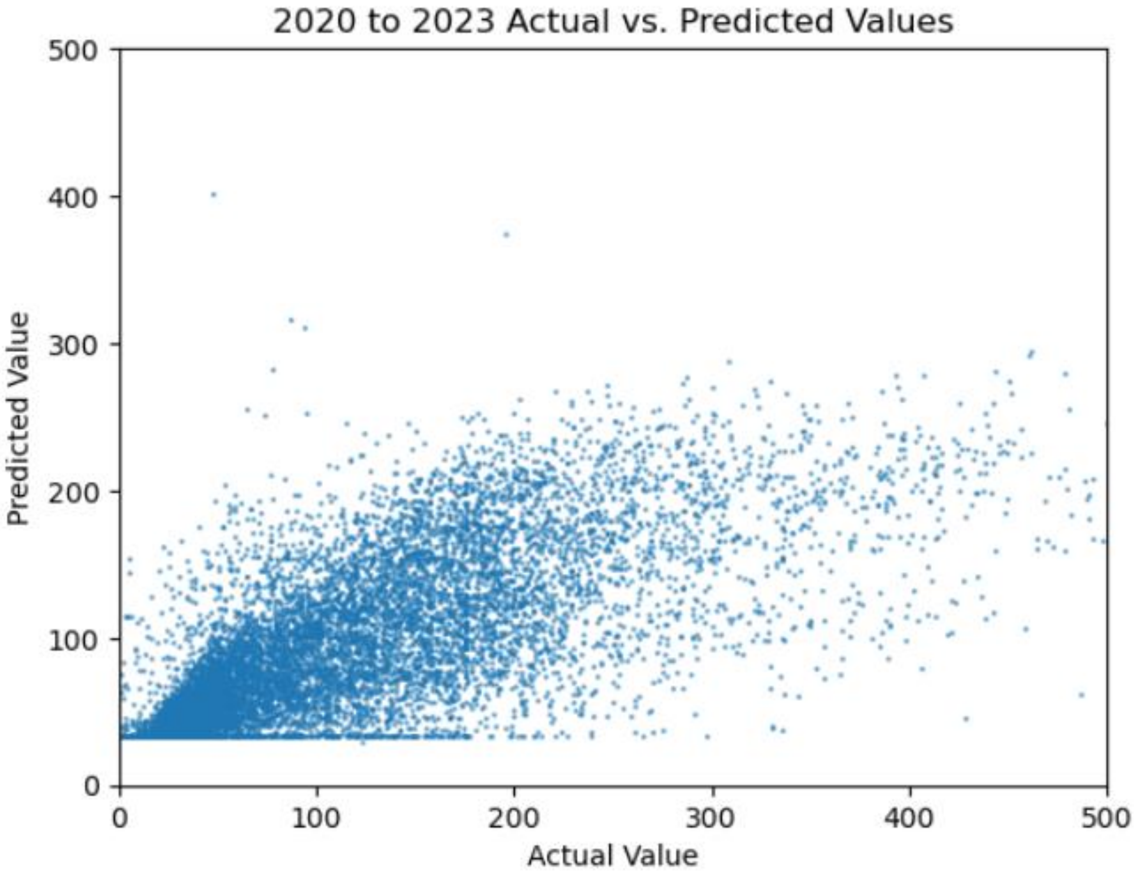



**Figure 22: Actual vs predicted electricity price scatterplot for 2020-2023**

**Predicted result for 2020 – 2023**

| Model | Linear Regression | Random Forest | ANN |
|---|---|---|---|
| **R2** | 0.42 | 0.79 | 0.41 |
| **MAE** | 51.36 | 23.88 | 44.25 |
| **RMSE** | 74.19 | 45.16 | 77.20 |

**Table 3 Outcomes of overall performance**

## 11. Main Result

Comparisons of all three models and all three stages of prediction reveal that the Random Forest has a highly accurate R2 score and MAE and RMSA have low values. I turned the parameters epoch and batch size for better results of the ANN model but still got low accuracy compared with Random Forest. I can conclude, therefore, that the Random Forest model has the best outcomes in performance because the predictions of the model are closer to the actual electricity prices.

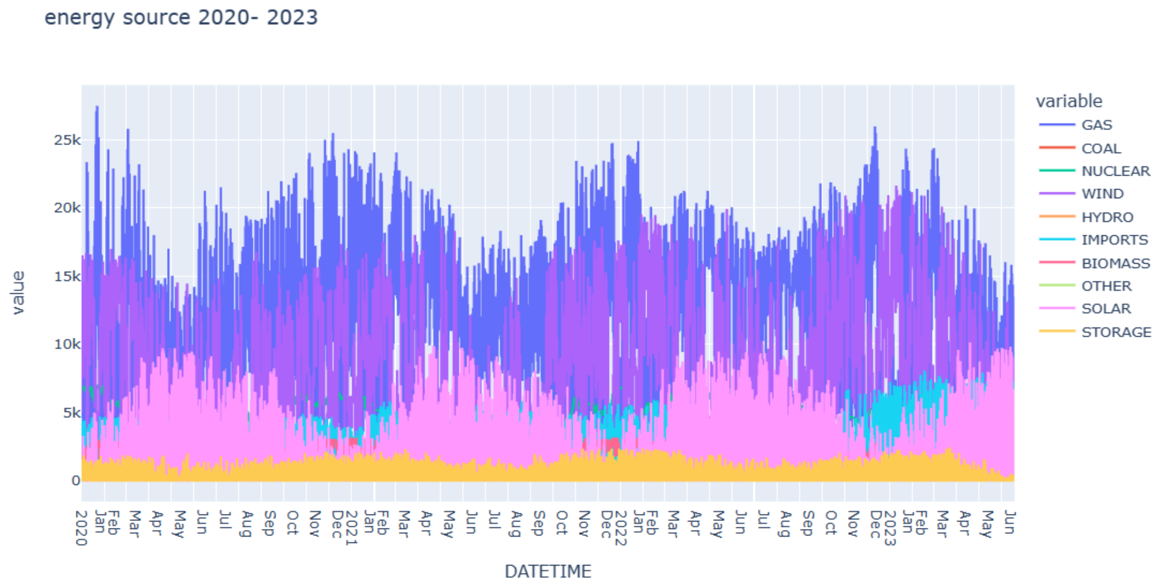
## 12. Conclusion

After Analysis of demand data, market index, Renewable energy source, and system price, a method to predict electricity prices for each settlement period using machine learning techniques is proposed. Among the Linear Regression, Random Forest Regression (RFR), and Artificial Neural Networks models, the RFR model has the most accurate price predictability power, with a 0.86 R2 Score. Renewable source energy generation is strongly connected to the market index price. Demand for renewable sources will increase the price. Market prices tend to be high in the winter season. It is suggested that the RFR algorithm can be used to predict power costs for each half-hour Settlement Period.

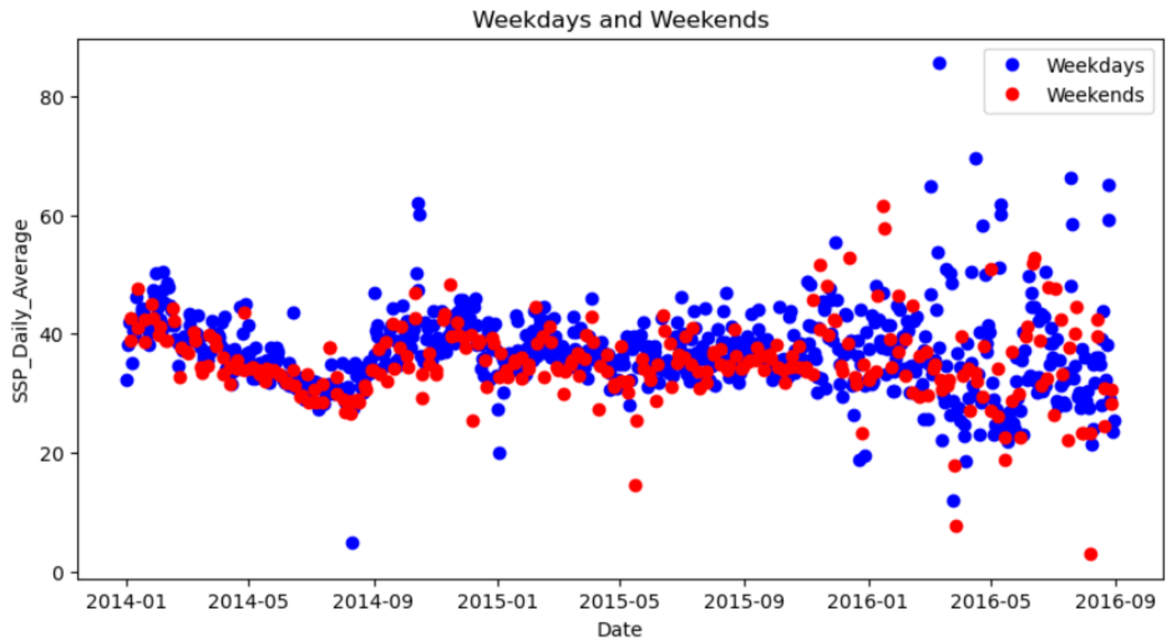
## 13.Acknowledgment

I would like to convey my heartfelt thanks to Prof. Ekkehard Ullner and Prof. Mamen Romano Blasco for their support and assistance to complete the project successfully. I thank for Elexon and National Grid who provided the data to predict the electricity price. My sincere thanks to my friends for the encouragement and guidance to do my project.

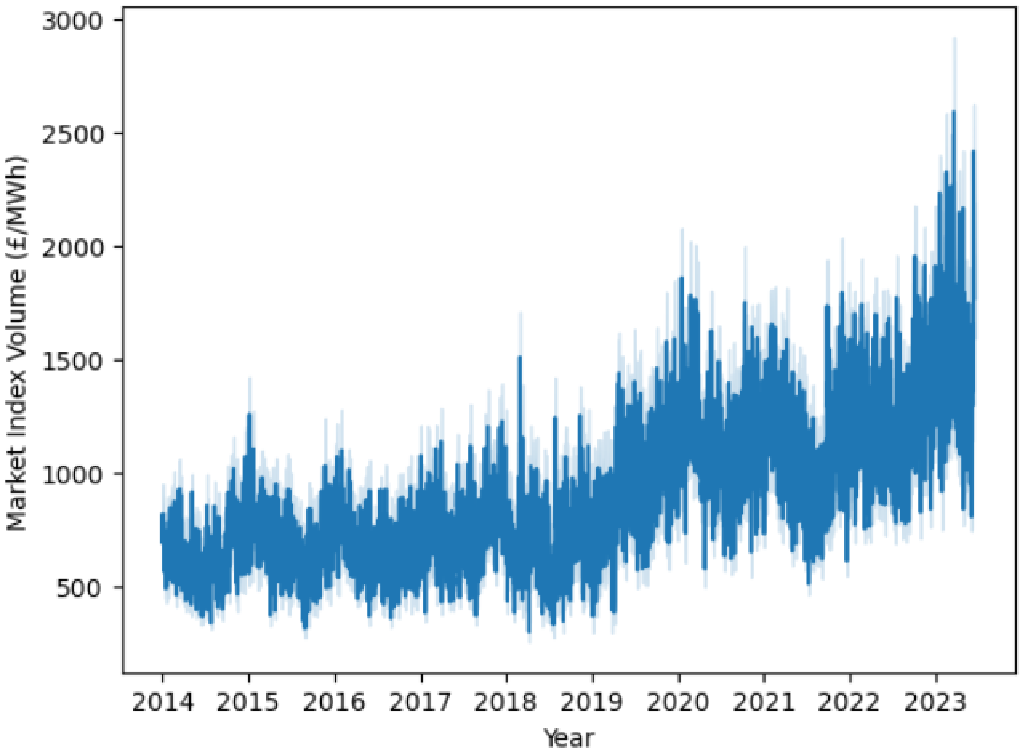
# 14.Appendix




Appendix 1-Energy source generation 2020-2023




Appendix 2 -System sells price weekdays vs weekends of 2014-2016




Appendix 3 -Market Index Volume for 2014-2023

References:

[1] Elexon Last accessed on 10/08/2023 https://www.elexon.com/about-elexon/

[2] https://data.nationalgrideso.com/demand/historic-demand-data

[3] https://bscdocs.elexon.co.uk/guidance-notes/the-electricity-trading-arrangements-a-beginners-guide

[4] https://www.elexon.co.uk/article/system-prices-spike-on-7-september-2021/

[5] https://thinkingneuron.com/using-artificial-neural-networks-for-regression-in-python/

[6] Hashmi. **Lead Data Scientist**. Last accessed on 10/08/2023 [ https://thinkingneuron.com/using-artificial-neural-networks-for-regression-in-python/ ]

[7] Gao, Gao, Kwoklun Lo, and Fulin Fan. "Comparison of ARIMA and ANN models used in electricity price forecasting for the power market." *Energy and Power Engineering* 9.4B (2017): 120-126.

[8] Gao, Gao, et al. "A short-term electricity price forecasting scheme for the power market." *World Journal of Engineering and Technology* 4.3D (2016): 58-65.

[9] Sridharan, Vasudharini, Mingjian Tuo, and Xingpeng Li. "Wholesale electricity price forecasting using integrated long-term recurrent convolutional network model." *Energies* 15.20 (2022): 7606.

[10] Yuan, Zhongbo, Raja Biswas, and Leong Hien Poh. "Accelerated offline setup of homogenized microscopic model for multi-scale analyses using a neural network with knowledge transfer." *International Journal for Numerical Methods in Engineering* 124.13 (2023): 3063-3086.

[11] Wenjie Zhang, Farwa Cheema, and Dipti Srinivasan Department of Electrical and Computer Engineering, National University of Singapore, Singapore, 117583.

[12] Lago, Jesus, et al. "Forecasting Day-ahead electricity prices in Europe: The importance of considering market integration." *Applied Energy* 211 (2018): 890-903.

[13] Castelli, Mauro, Aleš Groznik, and Aleš Popovič. "Forecasting electricity prices: A machine learning approach." *algorithms* 13.5 (2020): 119.

[14] Aggarwal, Sanjeev Kumar, Lalit Mohan Saini, and Ashwani Kumar. "Electricity price forecasting in deregulated markets: A review and evaluation." *International Journal of Electrical Power & Energy Systems* 31.1 (2009): 13-22.

[15] Hu, Linlin, and Gareth Taylor. "A novel hybrid technique for short-term electricity price forecasting in UK electricity markets." *Journal of International Council on Electrical Engineering* 4.2 (2014): 114-120.

[16] Ebrahimian, Homayoun, et al. "The price prediction for the energy market based on a new method." *Economic research-Ekonomska istraživanja* 31.1 (2018): 313-337.

[17] Nogales, Francisco Javier, et al. "Forecasting next-day electricity prices by time series models." *IEEE Transactions on power systems* 17.2 (2002): 342-348.

[18] Catalão, João Paulo da Silva, et al. "Short-term electricity prices forecasting in a competitive market: A neural network approach." *electric power systems research* 77.10 (2007): 1297-1304.

[19] Snell, Richard. "Thesis submitted to the University of Strathclyde in fulfillment of the for the degree of Master of Philosophy." (2020).

[20] Last accessed on 10/08/2023 https://stackoverflow.com/questions/55268578/split-dataframe-by-month

[21] Last accessed on 10/08/2023 https://www.w3resource.com/python-exercises/pandas/groupby/python-pandas-groupby-exercise-12.php

[22] Last accessed on 10/08/2023 https://plotly.com/python/time-series/