# Household Electricity Prices in Europe

**Students:** Sebastian Ionuț Cotinghiu, Merve Pakcan Tufenk, Tudor Ungureanu

**Abstract.** This project focuses on a quantitative analysis of electricity prices in each European country over a period starting from 2015 to 2024. The goals of the project are to search for the right dataset based on the project blueprint, analyse the data to identify insights, apply multiple algorithms to predict future values for the price metric and make data-driven conclusions and recommendations based on project findings. The dataset used includes monthly price values for 31 countries. The data analysis includes an introductory part where some insights regarding the price evolution over two major disruptive events in Europe (the COVID-19 pandemic and the Russian invasion of Ukraine) are observed and a main part where the prediction algorithms are applied to the dataset. The prediction algorithms used were ARIMA, LSTM for a Hybrid model with ARIMA, and Linear Regression. The algorithms are compared based on MAE, MSE and MAPE metrics. The project concludes that the hybrid algorithm (ARIMA with LSTM) gives the best results, however, the actual results are not solid enough which implies that further tuning of the parameters is needed or a different approach.

**Introduction.** This project aims to analyze the evolution of retail electricity prices in Europe over an extended period of time. The analysis involves identifying trends for the whole region or on a country-by-country basis, detecting anomalies in price evolution that may be explained by past disruptive events such as the COVID-19 pandemic or the Ukraine war, and using three prediction algorithms to derive data-driven insights and recommendations for the future evolution of electricity prices in Europe. The analysis is primarily quantitative.

The project is divided into multiple parts, centered around the dataset used. The initial part details how the data was obtained, followed by the data cleaning process. The subsequent part focuses on understanding the data through basic statistical analysis and interpretations. The next section covers the prediction algorithms, their results, and metric comparisons. Finally, conclusions and recommendations are presented.

The prediction algorithms used were ARIMA, a Hybrid model between ARIMA forecast and LSTM, and Linear Regression. Their key metrics were compared, and conclusions were drawn based on the comparisons. The experimental part of the project was conducted in Python, utilizing libraries such as Pandas, Matplotlib, NumPy, Seaborn, Statistics, Scikit-learn, and TensorFlow.

**Data gathering.** At the core of this project is the essential data that needs to be analyzed. A relatively large dataset with observations spanning a long period of time for each or the majority of European countries is required. Data crawling on the web was the chosen method for gathering the initial dataset. The following datasets were found and retrieved from the web:

- Electricity prices by type of users [1];

Description: This indicator presents electricity prices charged to final consumers. Electricity prices for non-household consumers are defined as follows: *Average national price in Euro per kWh without taxes applicable for the first semester of each year for medium-sized industrial consumers* (Consumption Band Ic with annual consumption between 500 and 2000 MWh). Electricity prices for household consumers are defined as follows: ***Average national price in Euro per kWh including taxes and levies applicable for the first semester of each year for medium-sized household consumers*** (Consumption Band Dc with annual consumption between 2500 and 5000 kWh) [1].

- European wholesale electricity prices - monthly [2];

The mentioned dataset contains the prices of electricity in Europe, which are centralised in a monthly configuration. However, these are not the end consumer prices as they do not include taxes, levies, network charges, subsidies, and supplier profits. These are prices on what is called the spot market [2]. Unlike the previous dataset, which may be more relevant for this analysis, end-customer prices make a more relatable analysis.

- European Union Energy Market Data [3];

This dataset was collected from a Kaggle data source and it contains a large amount of data related to hourly updates on power princess across various systems. The purpose of the dataset is to be used for research and correlation insights as the European energy markets are highly dynamic due to factors such as renewable energy integration, supply-demand balance, and geopolitical influences [3]. What is interesting about this dataset is that it contains a categorization between fossil and renewable energy sources. It is not stated whether the prices contained in the dataset are end-consumer prices or prices at a different stage on the supply chain but for simplicity, we will consider them as end-consumer prices.

- Renewable energy share of total production - Europe [4];

This dataset is the result of a tool/methodology used by Eurostat to collect the mentioned data. The tool involves a standard for the calculation of the indicators related to the share of energy from renewable sources [4]. The dataset contains exactly the share of renewable energy for every European country starting from 2004 to 2023.

In the end, only the second dataset was used, as the other four either contained an insufficient number of entries, as was the case with the first and third datasets or included data outside the scope of the project and unrelated to the price dimension, as with the fourth dataset. Although the second dataset does not include end-customer prices, the analysis focuses primarily on the evolution of this metric rather than its absolute value.

It could be argued that, before prices reach the end customer, the ratios for different countries may vary. However, this factor was considered in the project, and the results are not presented as fully conclusive.

**Data cleaning.** Data cleaning is required before any of the prediction algorithms are applied. Following this is a sample of the dataset. The dataset has 3504 observations, it contains monthly data for 31 countries starting with January 2015 and ending with December 2024. The sample includes the first 8 observations with the following features: *ISO3 Code*, *Date* and *Price* expressed in EUR/MWhe which is a unit that measures the power output of a power plant. For the scope of the project, it can be referred to as MWh or megawatt per hour as is a more common term.

| Country | ISO3 Code | Date | Price (EUR/MWhe) |
|---|---|---|---|
| Austria | AUT | 1/1/2015 | 29.94 |
| Belgium | BEL | 1/1/2015 | 42.33 |
| Czechia | CZE | 1/1/2015 | 29.47 |
| Denmark | DNK | 1/1/2015 | 27.12 |
| Estonia | EST | 1/1/2015 | 33.84 |
| Finland | FIN | 1/1/2015 | 33.81 |
| France | FRA | 1/1/2015 | 40.94 |
| Germany | DEU | 1/1/2015 | 29.94 |

Table 1. Dataset sample

The dataset was introduced into the coding environment, to better evaluate the spread of data, Figure 1 was ploted. As it can be seen from this figure, disregarding the versatility of the plot, there is no observation with value 0 or missing and there are some observations with a flat evolution, i.e. Muntenegro and North Macedonia (Figure 2 & 3).

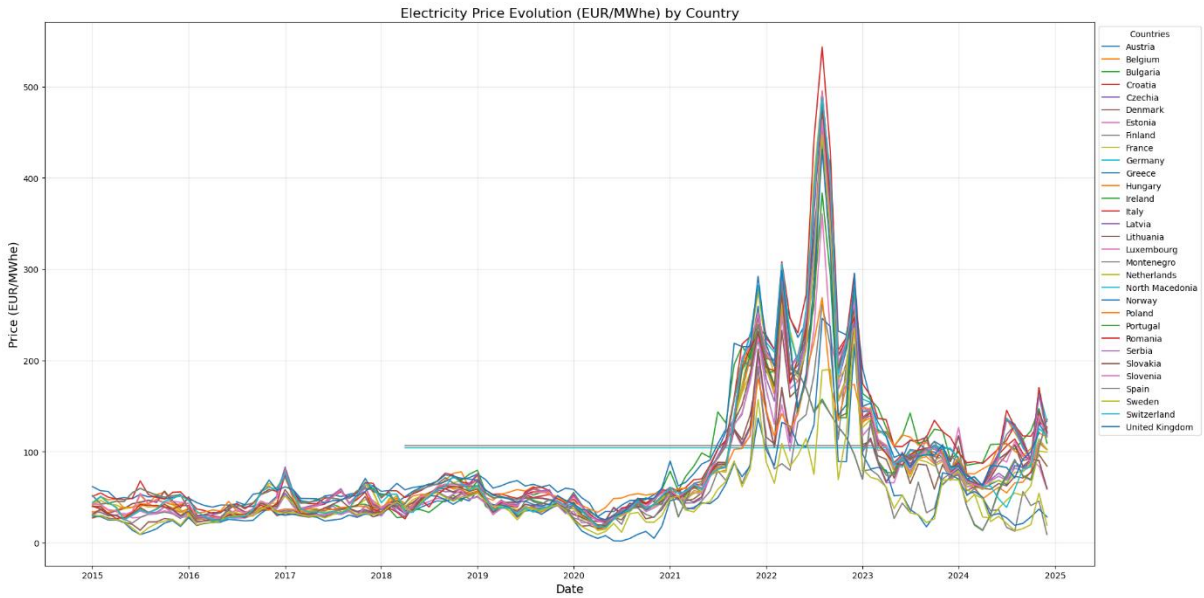


Figure 1. Dataset Visualization

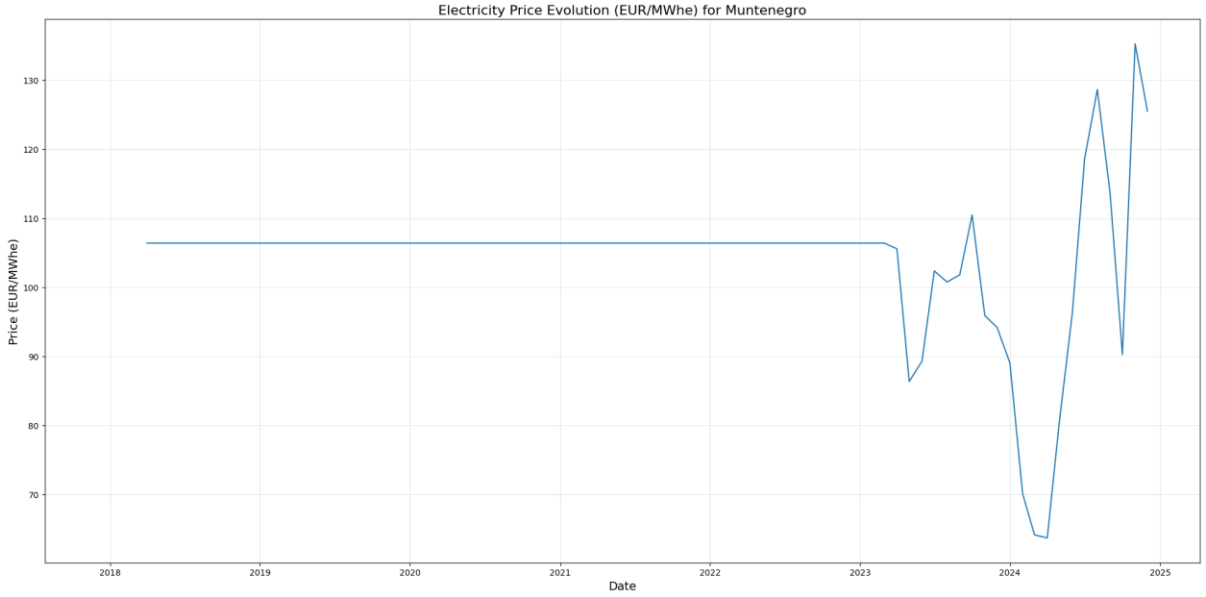


Figure 2. Price evolution of Muntenegro

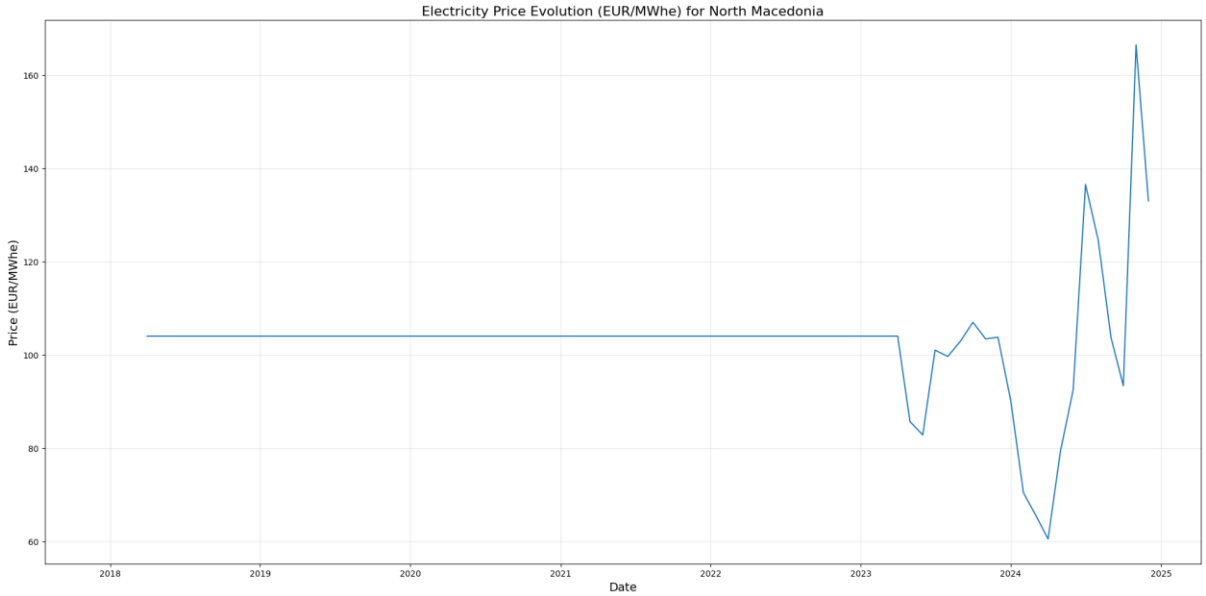


Figure 3. Price evolution of North Macedonia

The decision for these observations was to remove them, as they are not relevant as main drivers of the price, such as larger European economies. Once those were removed, the data was sorted by country name and then by date in order to organize it more categorically. Moreover, it was noticed that four countries (Bulgaria, Croatia, Serbia, and the United Kingdom) were missing price values for the first part of 2015. To use a uniform dataset, those values were filled using the average electricity price of that month for the entire group of countries, thus avoiding any bias or skewing of the data.

**Statistics.** As a prerequisite of the prediction algorithms, it would be necessary to briefly understand the price evolution of electricity in Europe over the analysed period. Figure 4 displays the evolution of the computed average price over the analysed period.

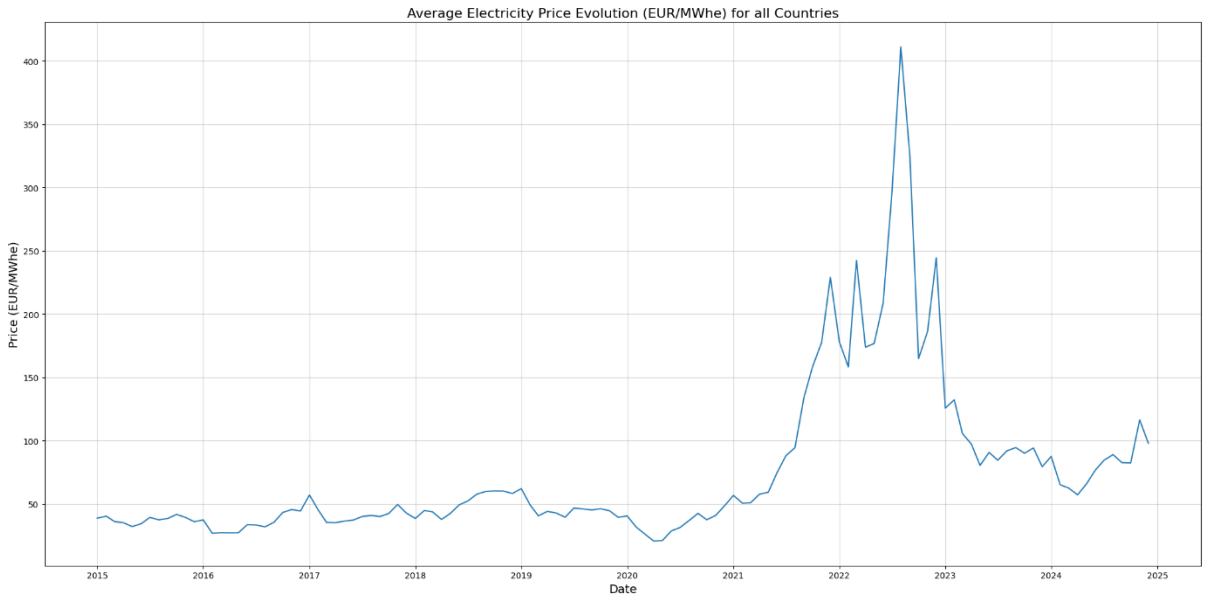


Figure 5. Average electricity price for all European countries.

It is evidently that the price mantained itself at a lower level belowe 60 euros while in the half of 2021 „in the wake of Covid-19 pandemic" the price started to increase considerably reaching values of around 175 euros at the start of 2022, when the Russia invasion of Ukrained lead to surging prices of 400 euros per MWh [5].

In [5] we learn that the sharp increase in energy import prices at the end of 2021, which more than doubled between December 2020 and December 2021, was driven by rising global demand and limited supply as economies recovered from the pandemic. This surge was unprecedented, as energy prices, despite their usual volatility, typically do not change by more than 30% in a year. The situation worsened in 2022 with Russia's war on Ukraine and its suspension of gas supplies to some EU member states, which pushed gas and electricity prices to record highs. Summer heatwaves further strained energy markets, increasing demand for cooling while droughts reduced hydropower supply.

On the other hand, it can also be observed that prices have started to decrease around 2023 and got to a level more relatable to the period before this two disruptive events described.

Looking back at figure 1, it could be observed that when the price disruptions appeared some countries performed better than others. In order to visualize which ones were those the following box plot was drafted (Figure 6).

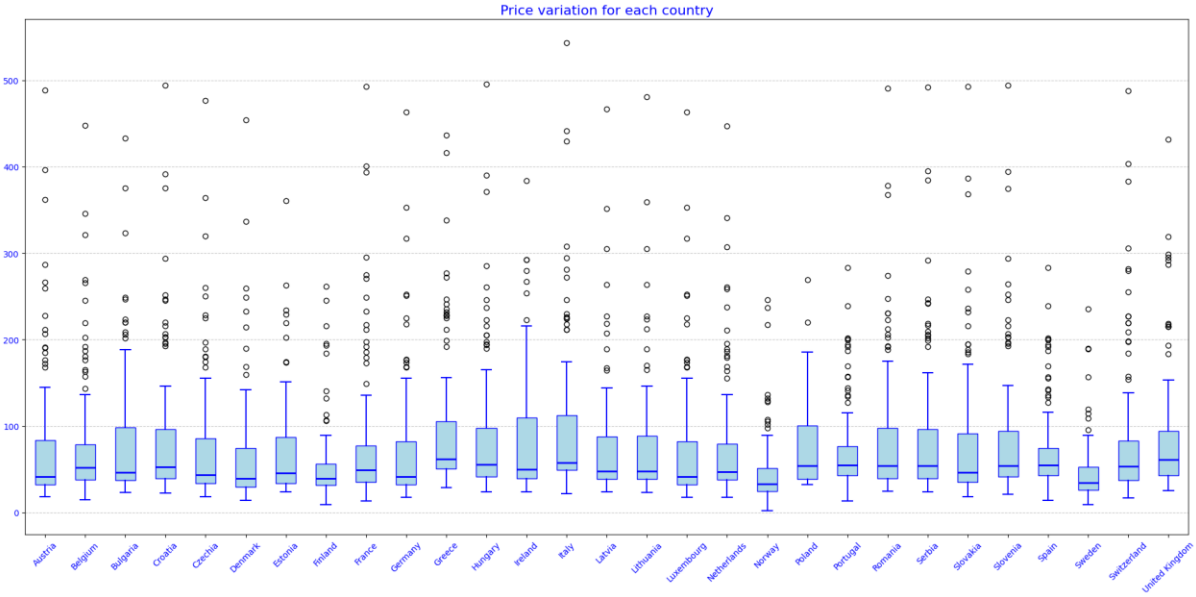


Figure 6. Price variation of each country

As it can be observed the mean values are rather similar for all countries. However, when looking at the outlier prices it can be spotted that some countries have much smaller values for upper outliers than other countries. This could be interpreted as having more stability during crises such as the two disruptive events. For this reason, the variance value of the price metric for each country was calculated and plotted in Figure 7 to identify which countries had the best price stability over the period analysed.

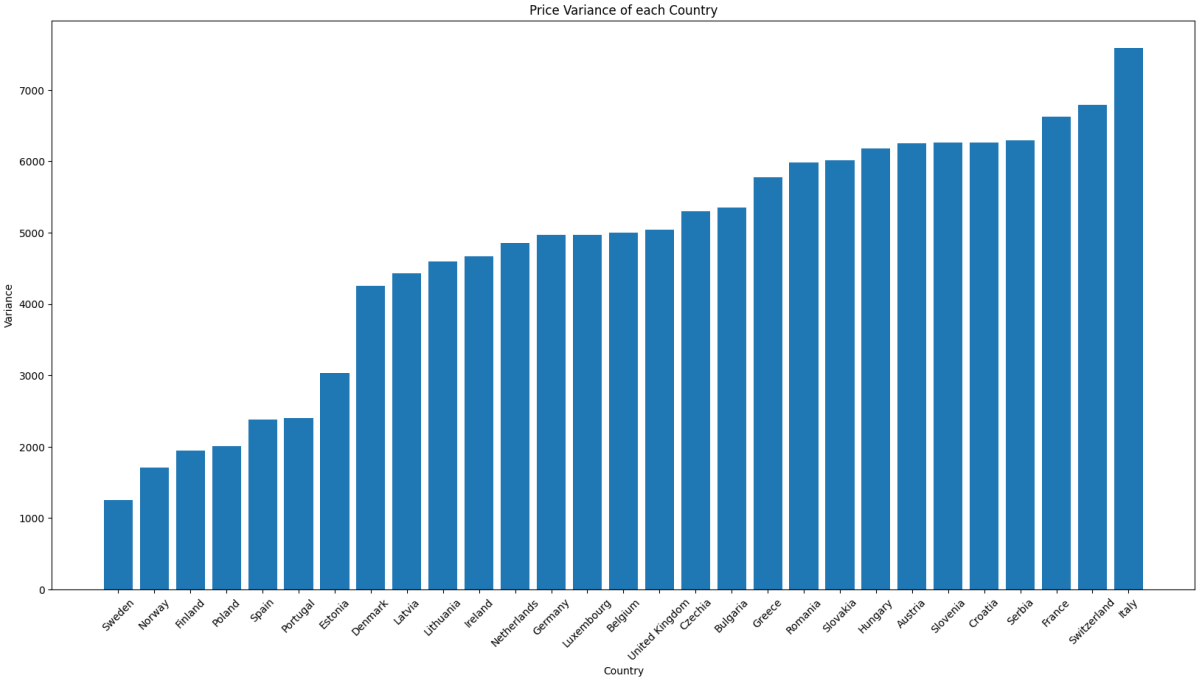


Figure 7. Price Variance of each country

The top 5 most stable countries were Sweden, Norway, Finland, Poland and Spain. One potential reason for this stability may be that according to [6] these Nordic countries as well as Spain, Estonia or Portugal have a high and even the highest share of renewable energy as an energy source. Such an asset may be used to cope with the dependency on external suppliers of energy and keep the price at a more stable level than the markets.

To support again the point already made, Figures 8 and 9 show the difference in price for the most unstable and stable countries in terms of price evolution. The difference can be best seen in the years 2021 and 2022 where the difference in prices is as much as 200 to 250 euros.

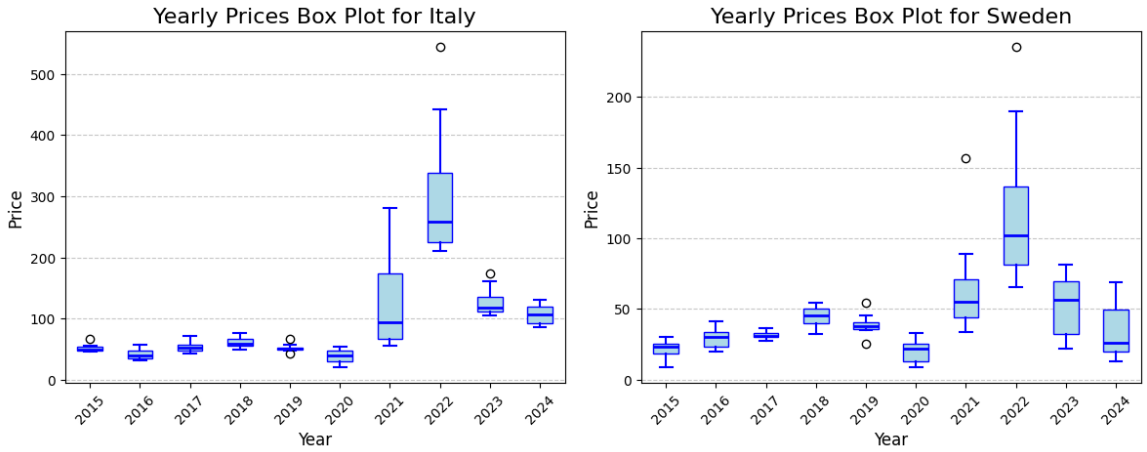


Figure 8. Price statistics yearly for Italy
Figure 9. Price statistics yearly for Sweden

Following is a final visualisation of the price variance for each of the analysed countries. The visualisation was created using *SAS Viya Data Analytics* platform and it shows again the countries with better stability in the electricity price over the others.

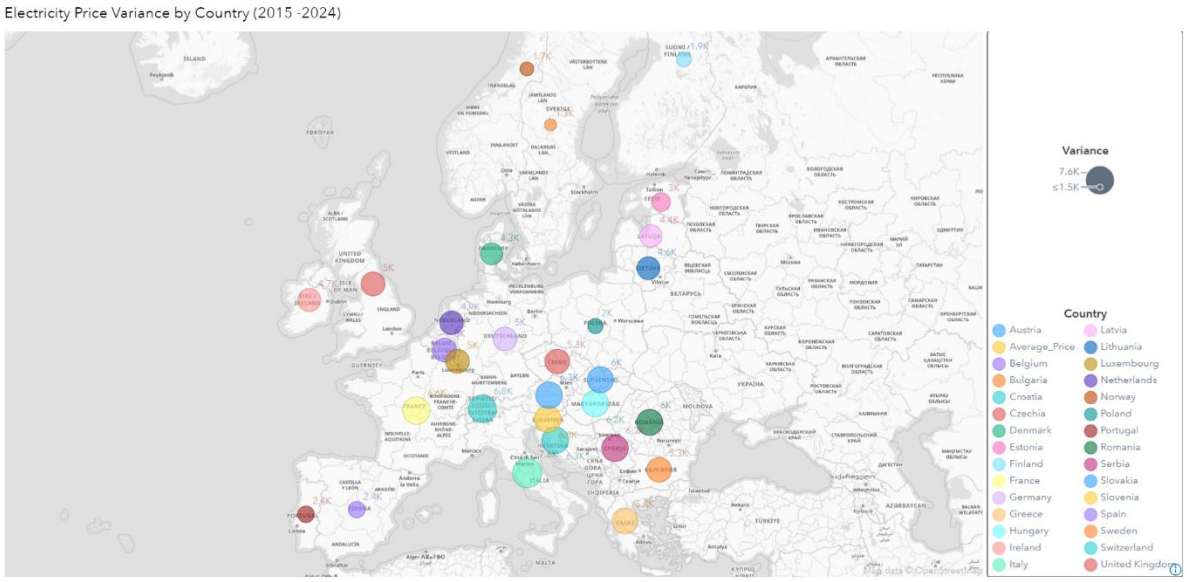


Figure 10. Price Variance of each country

**Data prediction**. Now that we have a better understanding of the data, the other part of the analysis, i.e. the price prediction, can be further detailed. A prediction can be defined as a statement or element of a future time that might or might not happen. The chance of being correct increases if the past is analyzed as well. In the case of this project, the aim is to predict the future price of electricity for each European country included in the dataset, based on the price evolution from 2015 until December 2024. Three prediction algorithms were used: *ARIMA (AutoRegressive Moving Average), LSTM (Long Short-Term Memory) for Hybrid Model with ARIMA, and Linear Regression.*

The ARIMA algorithm is a widely used statistical modeling technique for analyzing and forecasting time series data, such as the one used in this project. It is particularly effective for data that shows some degree of temporal correlation and trends [7]. For this experiment, the dataset was split into training data (up to 2023) and test data (from 2024 onward).

The *ARIMA* model was initiated using the SARIMAX class, then the model was fitted to the training data. The model was applied to each country using a loop function. Once the model was applied and the results were obtained, the residual values (loss) were calculated as the difference between the actual values and ARIMA predictions. Figure 11 shows the price evolution according to the ARIMA model.

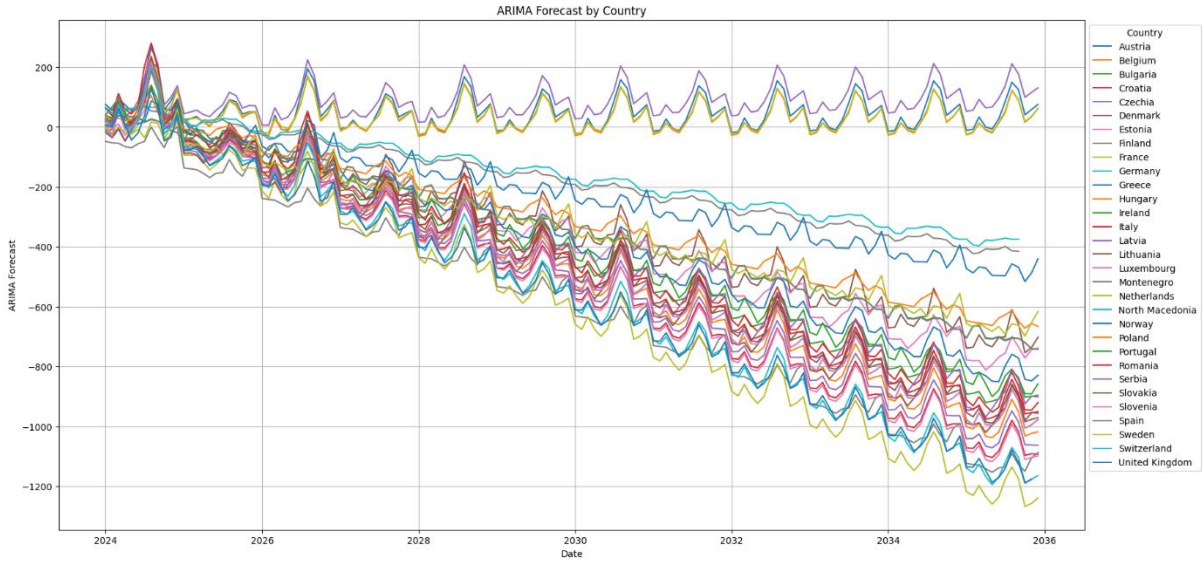


Figure 11. ARIMA forecast of each country

The results show a decreasing projection of the price and a repetitive pattern as well, which makes it clear that the prediction is not accurate. One reason for this may be the inappropriate choice of ARIMA parameters, the large variability in the data caused by two disruptive events, or non-linearity in the data [7].

Once ARIMA was applied, data for *LSTM* was then prepared using the residual values to create a more robust hybrid model. The LSTM model is used to support the ARIMA forecast. The residual values were amplified and scaled, then standardized. The LSTM training data was created using a sequence of 36 months. The LSTM model was designed with two LSTM layers for sequence learning, dropout layers to prevent overfitting, and dense layers for final regression output. The model was compiled and trained using the MSE loss function. Finally, the ARIMA and LSTM predictions were combined using a weight factor alpha of 0.7. Figure 12 shows the prediction results of the hybrid model.

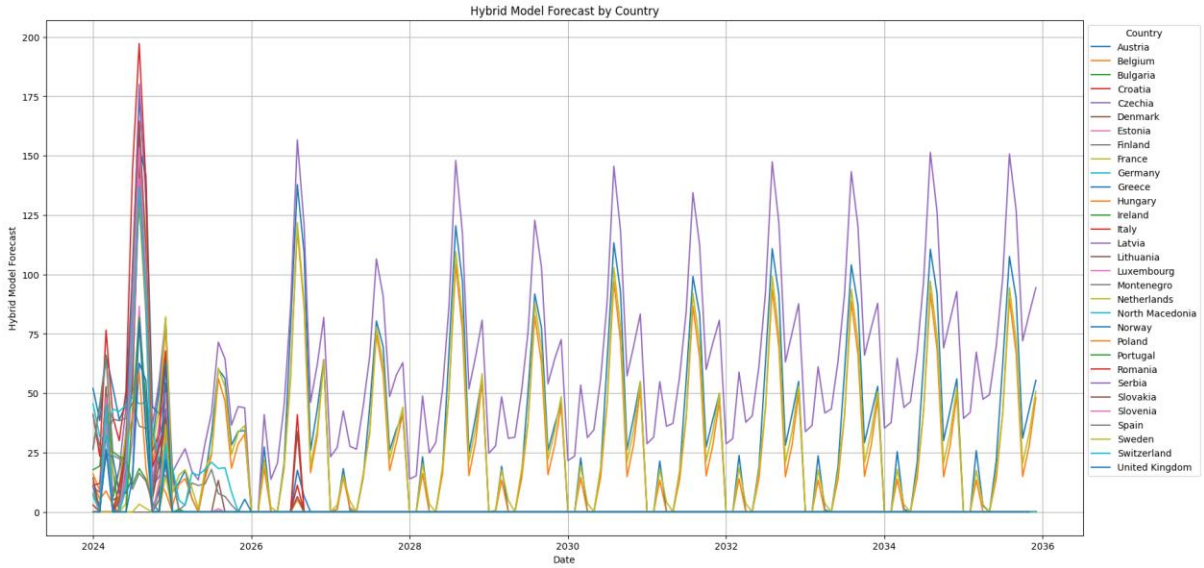


Figure 12. Hybrid Model forecast of each country

The results are looking more promising now, having no negative values, however, keeping the same pattern as earlier. The price ranges from 0 to 150 – 175 euros which in real-life cases might be hard to find.

Finally, the *Linear Regression* model was defined. The train data was defined and scaled. The model was applied using the SGDRegresor model. Figure 13 shows the forecasted private values for each country. The results for this model are more linear than the previous ones and show a constant increase in the price for each observation with different starting values.

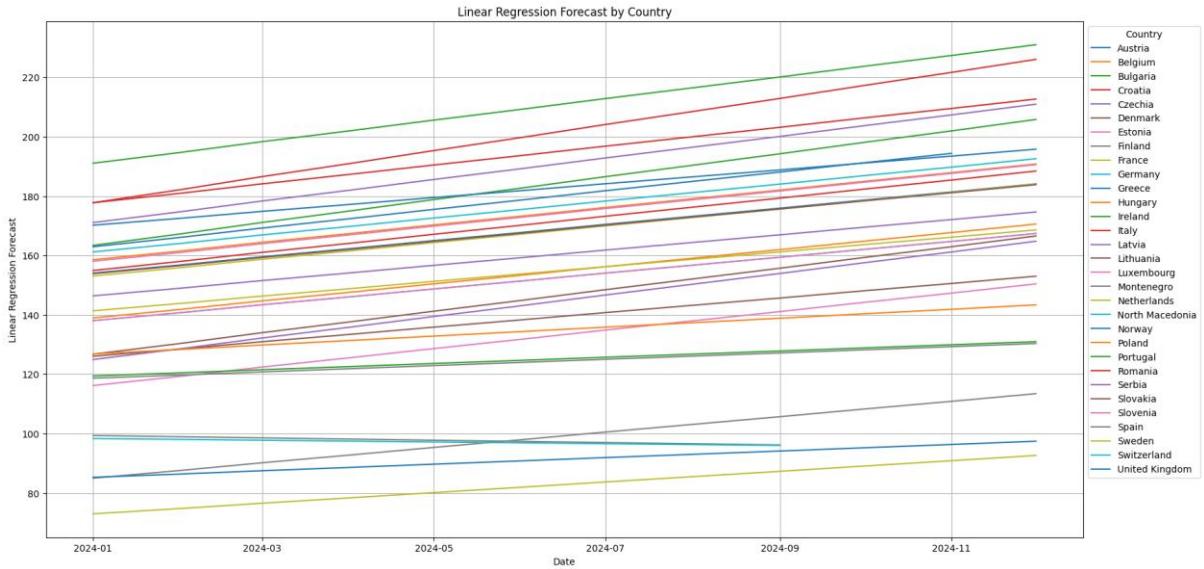


Figure 13. LR Model forecast of each country

Once all three models were applied and the price metric was forecasted, the MAE, MSE and MAPE parameters were calculated for each model and for each prediction observation, i.e. for each of the 29 countries. The following table contains the respective values along with a colour code for the smallest value of each comparison.

From the first look at the table, it seems that the most performant model was the Hybrid one, having the best metrics. However, it is also important to state that the values are overall high which leads to the conclusion that the models were either not the right options for the available dataset or the definition parameters used need more tuning in order to reach better results.

The hybrid model outperforms ARIMA and LSTM individually by combining their strengths and compensating for their weaknesses. ARIMA excels at capturing linear trends and seasonality but struggles with non-linear dynamics, while LSTM handles complex, non-linear patterns and long-term dependencies but can overfit or miss periodic seasonality. By using LSTM to model ARIMA's residuals, the hybrid model focuses on the patterns ARIMA cannot explain, leading to a more precise forecast. The weighted combination of ARIMA and LSTM predictions allows flexibility, adapting to the strengths of each model based on the data. This balanced approach reduces noise, improves generalization, and ensures better accuracy for complex time series data.

| Country | MAE_ARIMA | MAE_Hybrid | MAE_LR | MSE_ARIMA | MSE_Hybrid | MSE_LR | MAPE_ARIMA | MAPE_Hybrid | MAPE_LR |
|---|---|---|---|---|---|---|---|---|---|
| Austria | 61.983 | 56.576 | 88.840 | 6383.839 | 3596.395 | 8145.227 | 81.187 | 72.860 | 121.313 |
| Belgium | 55.648 | 50.534 | 84.725 | 5079.431 | 2985.570 | 7382.206 | 83.876 | 74.781 | 131.982 |
| Bulgaria | 54.862 | 56.508 | 83.590 | 3841.798 | 4377.110 | 7565.018 | 54.388 | 57.642 | 97.835 |
| Croatia | 67.253 | 63.623 | 108.773 | 5608.334 | 5389.934 | 12074.013 | 72.219 | 69.955 | 128.124 |
| Czechia | 56.411 | 54.060 | 76.224 | 6014.487 | 3399.889 | 6063.563 | 69.854 | 65.396 | 98.878 |
| Denmark | 54.741 | 50.384 | 69.459 | 4738.767 | 3184.536 | 4894.224 | 76.697 | 71.176 | 102.729 |
| Estonia | 57.062 | 62.848 | 49.921 | 4193.317 | 4823.942 | 2865.419 | 68.768 | 73.816 | 65.882 |
| Finland | 86.006 | 43.158 | 59.586 | 9075.502 | 2481.734 | 4122.784 | 204.439 | 100.000 | 279.801 |
| France | 71.071 | 52.542 | 110.518 | 6424.827 | 3430.918 | 12625.815 | 126.625 | 90.210 | 238.253 |
| Germany | 69.584 | 58.152 | 75.214 | 5960.109 | 4249.632 | 5749.111 | 89.437 | 74.248 | 102.076 |
| Greece | 38.333 | 38.678 | 82.988 | 2267.674 | 2039.015 | 7327.661 | 36.798 | 37.326 | 94.937 |
| Hungary | 63.318 | 61.342 | 75.800 | 5285.712 | 5004.089 | 6322.103 | 65.246 | 63.950 | 90.581 |
| Ireland | 62.942 | 82.640 | 104.639 | 5702.153 | 7992.894 | 11065.293 | 56.693 | 76.464 | 101.429 |
| Italy | 62.020 | 55.711 | 87.431 | 5163.666 | 3705.261 | 7709.097 | 55.451 | 51.467 | 83.772 |
| Latvia | 67.740 | 57.668 | 60.052 | 5896.306 | 4205.632 | 4092.251 | 81.527 | 68.886 | 78.305 |
| Lithuania | 72.691 | 59.336 | 61.858 | 6618.215 | 4374.976 | 4315.168 | 87.471 | 70.820 | 80.575 |
| Luxembourg | 69.584 | 57.431 | 75.214 | 5960.109 | 4179.254 | 5749.111 | 89.437 | 73.195 | 102.076 |
| Netherlands | 52.625 | 48.462 | 78.203 | 4426.005 | 2844.518 | 6243.636 | 69.832 | 64.135 | 108.179 |
| Norway | 35.042 | 32.946 | 53.775 | 1589.294 | 1236.648 | 3223.927 | 117.109 | 97.901 | 184.847 |
| Poland | 54.424 | 79.989 | 38.396 | 3423.799 | 6733.590 | 1587.952 | 56.923 | 83.313 | 42.276 |
| Portugal | 58.205 | 54.571 | 60.660 | 5020.441 | 4343.703 | 4591.265 | 107.903 | 86.308 | 194.034 |
| Romania | 61.006 | 61.059 | 69.919 | 4866.345 | 4919.611 | 5568.072 | 61.507 | 62.181 | 83.371 |
| Serbia | 64.962 | 62.645 | 90.798 | 5561.916 | 5344.897 | 8726.874 | 65.695 | 64.714 | 104.175 |
| Slovakia | 66.565 | 58.622 | 78.192 | 5247.291 | 4506.454 | 6386.459 | 72.865 | 65.273 | 95.384 |
| Slovenia | 69.396 | 59.801 | 84.402 | 5655.853 | 4837.550 | 7362.970 | 76.810 | 67.189 | 103.270 |
| Spain | 57.801 | 54.305 | 60.455 | 4990.889 | 4340.177 | 4541.672 | 103.979 | 83.786 | 188.985 |
| Sweden | 46.019 | 33.570 | 48.697 | 2587.094 | 1464.381 | 2858.523 | 144.760 | 96.377 | 227.359 |
| Switzerland | 72.558 | 63.686 | 101.658 | 6649.662 | 4912.970 | 10787.545 | 103.978 | 86.352 | 159.153 |
| United Kingdom | 79.688 | 71.534 | 94.795 | 8433.256 | 5895.464 | 9108.516 | 91.866 | 83.887 | 118.336 |

## Conclusions and Recomandations

This project focuses on a quantitative analysis of electricity prices in each European country over a period starting from 2015 to 2024. The goals of the project are to search for the right dataset based on the project blueprint, analyse the data to identify insights and apply multiple algorithms to predict future values for the price metric.

At the beginning of the first step, mutiple dataset were taken into account. In the end, only dataset number second was used as it contained a large enough number of observations. The dataset was analysed and cleaned and some insights could be drawn from it regarding the price evolution over two disruptive events which were the COVID-19 pandemic and the Russian invasion of Ukraine. This disruptive event led to surging price increases reaching values 4 to 5 times higher than the price one year before. One major conclusion from the analysis of the data would be that the countries with high renewable sources of energy showed increased price stability during the disruptive years, which may be or not a direct correlation. However, it is a proof which stands and can be further studied and proven.

For the price prediction, three models were used. The first one was ARIMA which returned strange values such as negative values. The second one was a Hybrid model between the results of the ARIMA model and the LSTM model applied to the residual values. The LSTM used to model ARIMA's residuals, focuses more on the patterns ARIMA cannot explain, leading to a more precise forecast, which was the final result therefore. The Hybrid model returned more appropriate values and better performance. The final model applied was a Linear Regression which returned good enough price forecasts however the performance metrics were not as good as the other two. Finally, all model performances were compared using the MAE, MSE and MAPE parameters and the comparison showed that the Hybrid model was the best performant for the majority of the predictions.

Further analysis can be done in order to improve the used models. Primarily focusing on tuning the model parameters to improve the performance as much as possible and secondarily either using another model or using a similar dataset but with hourly observations which means a much larger number of observations than the one used for this analysis.

**References:**


[1]. Eurostat Database. Electricity prices by type of user. Source: https://ec.europa.eu/eurostat/databrowser/view/ten00117/default/table?lang=en&category=t_nrg.t_nrg_indic

[2]. European Wholesale Electricity Price Data. Source: https://ember-energy.org/data/european-wholesale-electricity-price-data/

[3]. Kaggle. European Union Energy Market Data. Source: https://www.kaggle.com/datasets/pythonafroz/european-union-energy-market-data?resource=download

[4]. Eurostat Database. Short Assessment of Renewable Energy Sources (SHARES). Source: https://ec.europa.eu/eurostat/web/energy/database/additional-data#Short%20assessment%20of%20renewable%20energy%20sources%20(SHARES)

[5] European Council of the European Union. Energy price rise since 2021. Srouce: https://www.consilium.europa.eu/en/infographics/energy-prices-2021/

[6] European Environment Agency. 31 Oct 2024. Share of energy consumption from renewable sources in Europe. Source: https://www.eea.europa.eu/en/analysis/indicators/share-of-energy-consumption-from

[7] Investopedia. ARIMA. Adam Hayes. 31 July 2024. Source: https://www.investopedia.com/terms/a/autoregressive-integrated-moving-average-arima.asp