ANALYSIS OF THE DYNAMICS OF QUEENSLAND'S ELECTRICITY MARKET

DATA7001 Introduction to Data Science 25/10/2024

GROUP 16

Xiaoqing Yue
William Cruickshank
Keerthi Parthipan
Wenjun Xie
Howard (Zhe) Chen

We give consent for this report and the video of our final presentation to be used as a teaching resource.

Table of Contents

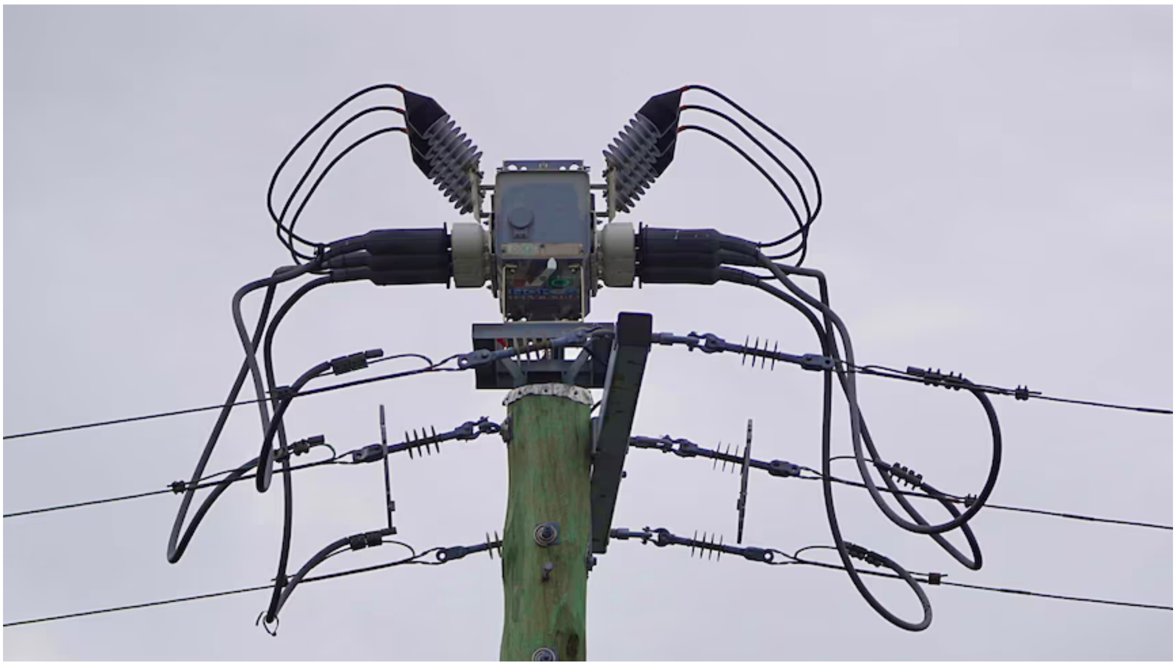
1. Abstract

This project investigated Queensland's electricity market to examine the connections between energy prices, energy demand, and weather temperature. Data on energy and weather from five Australian states were analysed over a period of one year. By applying regression analysis, the project identified significant factors affecting energy prices, with noticeable changes in demand throughout the day and week. However, it found no correlation between temperature and energy prices. The results could help energy producers enhance their generation strategies, assist consumers in making better decisions, and support policymakers in promoting sustainability and efficiency.

2. Problem-solving with data

2.1. Why choose this topic?

Electricity is essential to our lives. It drives economic growth and supports important services in our everyday life. In recent years, however, the rising electricity costs in Queensland have become a major concern for both householders and businesses. Many households are struggling with higher utility bills, leading to less disposable income and increasing worries about energy affordability. In response to this, the Queensland government has implemented a measure to help households by providing a $1,300 reduction on their electricity bills for the year 2024 to 2025. This will include a $1,000 upfront Cost of Living Rebate from the Queensland Government and an additional $300 rebate distributed in quarterly instalments by the Australian Government.



Despite these measures, it is important to understand the factors driving the increase in electricity prices. The increase in electricity prices can be caused by several factors. These include supply-demand imbalances due to outages in the wholesale market, fluctuations in fuel costs, changes in regulations, and periods of high demand. Our research aims to identify the main factors that affect energy demand. Additionally, we will provide practical suggestions that can help households and businesses manage the challenges of rising electricity costs. By understanding these factors, we hope to empower consumers to make informed decisions about their energy use and find ways to reduce their expenses.

2.2. Stakeholders

In this project, the key stakeholders are energy producers, consumers and policymakers.

- Energy Producers: this research might help them optimize operations and make informed decisions about investments in infrastructure and renewable energy sources by discovering demand patterns. By understanding when and how energy demand fluctuates, producers can better align their supply strategies to meet consumer needs.
- Consumers: our project's findings can help consumers understand their energy usage patterns. With this information, households can adopt strategies to use energy more efficiently and lower their costs, which allows them to make better decisions about their energy consumption.
- Policymakers: our analysis may provide insights on making regulations and policies that promote more efficient energy practices, enhance market stability, and address affordability issues for consumers.

2.3. Research Questions

In this project, the following 3 research questions will be investigated:

- What seasonal trends are observable in energy demand?
- How do economic and social factors influence fluctuations in energy demand?
- In what ways does the integration of renewable energy impact overall demand patterns?

3. Where to get the data

In order to analyse the dynamics of the Queensland electricity market in depth, we need to collect relevant and reliable data. The source of the data directly affects the accuracy of the analysis and the credibility of the results. Therefore, the selection of appropriate data sources is the first and crucial step in our entire study. Relevant details on data source, size, format and description are as follows.

3.1. Data Sources

Our data is mainly derived from two authoritative sources: Open-Meteo (2022) and AEMO, the Australian Energy Market Operator (2022). Open-Meteo provides us with weather data in detail. The Open-Meteo data is based on the information from Australian Bureau of Meteorology. AEMO provides emissions and price data from the electricity market. As an important operator in the Australian energy market, AEMO ensures quality and accuracy of the data. Thus, these data are able to provide a solid basis for our analysis. Links to collect database resources are shown below.

3.2. Data Size and Format

We processed two main datasets: the price dataset and the weather dataset. The price dataset contains 60 files, each of which is about 400KB in size. Each file contains five columns of data, and each file has over 8,000 rows. This data records several key parameters ensuring that we can extract rich information related to the energy market. When it comes to the weather dataset, the file size of the weather dataset is about 2.2MB, covering 32 columns and about 8,900 rows of data. The data in both datasets are numerical data, which means the data can be easily imported into data analysis tools, such as Excel or Python.

	A	B	C	D	E	F
1	REGION	SETTLEMI	TOTALDE	RRP	PERIODTYPE	
2	QLD1	#######	5737.46	87.42	TRADE	
3	QLD1	#######	5781.31	87.99	TRADE	
4	QLD1	#######	5726.56	87.91	TRADE	
5	QLD1	#######	5693.06	87.86	TRADE	
6	QLD1	#######	5706.07	85.6	TRADE	
7	QLD1	#######	5655.75	85.6	TRADE	
8	QLD1	#######	5700.05	85.6	TRADE	
9	QLD1	#######	5650.63	85.55	TRADE	
10	QLD1	#######	5698.36	89.39	TRADE	
11	QLD1	#######	5662.52	85.6	TRADE	
12	QLD1	#######	5625.36	61.82	TRADE	
13	QLD1	#######	5623.51	67.71	TRADE	
14	QLD1	#######	5580.43	59.6	TRADE	
15	QLD1	#######	5618.3	61.37	TRADE	
16	QLD1	#######	5534.07	60.89	TRADE	
17	QLD1	#######	5511.93	59.55	TRADE	
18	QLD1	#######	5566.99	60.91	TRADE	
19	QLD1	#######	5552.43	59.6	TRADE	
20	QLD1	#######	5527.61	51.29	TRADE	

Figure 1 Energy Price sample data

temperatur	relative_hu	dew_point_	apparent_t	precipitatio	rain	snowfall	snow_depth	weather_co	pressure_m	surface_pre
17.4195	32.64337	0.8195	14.128002	0	0	0	0	0	1019.4	987.9889
17.1695	33.523308	0.9695	14.042986	0	0	0	0	0	1019.1	987.67163
15.3695	38.970646	1.4695001	12.191429	0	0	0	0	0	1018.8	987.18835
15.0195	40.721653	1.7695	12.037466	0	0	0	0	0	1018.3	986.66626
15.3195	40.230682	1.8695	12.222904	0	0	0	0	0	1017.6	986.0202
14.8695	42.006744	2.0695	11.838459	0	0	0	0	0	1017.6	985.9718
14.3195	43.99224	2.2195	11.364604	0	0	0	0	0	1017.9	986.2032
13.9695	45.483746	2.3695	11.009619	0	0	0	0	0	1018.6	986.8434
14.669499	43.936516	2.5195	11.625097	0	0	0	0	0	1019.5	987.79126
17.9195	36.61895	2.8695	14.824326	0	0	0	0	0	1020.2	988.81726
20.9195	34.971596	4.8694997	17.928114	0	0	0	0	0	1020.8	989.71313
24.219501	32.86599	6.8694997	20.999386	0	0	0	0	0	1020.7	989.9549
26.119501	26.55629	5.4195	23.387861	0	0	0	0	0	1020	989.46765
27.3195	22.91887	4.3195	25.486225	0	0	0	0	0	1018.8	988.42316
28.119501	21.044216	3.7695	26.961178	0	0	0	0	0	1017.2	986.9499
28.5195	19.848438	3.2695	27.403196	0	0	0	0	0	1016.3	986.1161
28.8195	18.431278	2.4695	26.742146	0	0	0	0	0	1015.3	985.1753
28.4195	18.864216	2.4695	25.159822	0	0	0	0	0	1015	984.8448
27.4195	19.785404	2.3195	23.803516	0	0	0	0	0	1015.1	984.8434
25.2695	22.145084	2.1195	22.341316	0	0	0	0	0	1015.4	984.9203
22.4195	25.913195	1.9195	19.195961	0	0	0	0	0	1015.8	985.01965
22.3195	25.701368	1.7195001	19.095951	0	0	0	0	0	1016.1	985.3003
20.5695	28.91693	1.8695	17.174513	0	0	0	0	1	1016.3	985.3141
19.969501	29.902014	1.8195	16.933224	0	0	0	0	0	1015.8	984.7669
18.0195	36.00459	2.7195	15.074108	0	0	0	0	0	1015.8	984.5632
17.469501	38.344517	3.1195	14.607079	0	0	0	0	0	1015.4	984.1175
17.1695	38.257343	2.8195	14.155098	0	0	0	0	0	1015.3	983.98883
17.0195	38.213715	2.6695	13.817493	0	0	0	0	0	1014.7	983.3913
17.5695	37.7012	2.9695	14.111317	0	0	0	0	2	1014.1	982.8679
17.9195	36.87916	2.9695	14.356249	0	0	0	0	2	1013.7	982.5172
18.1695	36.048157	2.8695	14.429489	0	0	0	0	1	1013.8	982.64044
17.369501	37.91075	2.8695	13.917721	0	0	0	0	1	1014.7	983.42847
16.719501	40.066963	3.0695	13.6846695	0	0	0	0	1	1015.7	984.32874

temperatur	cloud_cove	cloud_cove	cloud_cove	cloud_cove	et0_fao_ev	vapour_pre	wind_speed	wind_speed	wind_direct
17.4195	3.6000001	0	6	0	0.07543394	1.3416431	8.891344	14.618837	68.62931
17.1695	3	0	5	0	0.06545269	1.3033767	7.928177	17.610588	50.52757
15.3695	0	0	0	0	0.06198398	1.0670437	8.854829	17.068707	26.564985
15.0195	0	0	0	0	0.05195812	1.0134027	7.8625183	19.855558	15.945477
15.3195	0	0	0	0	0.05959025	1.0416641	8.766573	22.79684	19.179106
14.8695	0.6	0	1	0	0.05497361	0.9819121	8.557102	22.35481	14.620926
14.3195	5.4	0	9	0	0.04925237	0.9152494	8.217153	21.53089	28.810698
13.9695	7.8	0	13	0	0.04823947	0.87092316	8.435069	21.758419	50.194473
14.669499	0	0	0	0	0.07624938	0.937086	9.19939	23.188583	30.579157
17.9195	1.2	0	2	0	0.13630266	1.3027887	9.983106	22.127556	25.640957
20.9195	0	0	0	0	0.2798682	1.6103381	11.966954	16.516901	15.708701
24.219501	0	0	0	0	0.45170867	2.030541	17.771753	23.177402	6.980963
26.119501	0	0	0	0	0.5753016	2.487072	17.643673	22.33161	358.83087
27.3195	0	0	0	0	0.6376366	2.8009858	13.849476	17.518356	351.02744
28.119501	2.4	0	4	0	0.64544106	3.0061505	10.137692	12.0694	343.49554
28.5195	1.2	0	2	0	0.5943424	3.1234004	7.491114	9	324.78232
28.8195	4.8	0	8	0	0.52491605	3.2343779	8.217153	9.826088	298.8107
28.4195	2.4	0	4	0	0.43330896	3.1435394	10.587917	12.979984	287.81897
27.4195	4.8	0	8	0	0.31104147	2.9320924	12.849528	15.745627	281.3099
25.2695	3	0	5	0	0.12485585	2.5072882	7.8954163	16.263872	294.22772
22.4195	4.2000003	0	7	0	0.10637226	2.0107074	9.686609	15.124284	318.01288
22.3195	19.2	0	32	0	0.10392158	2.0042577	9.449572	11.808878	319.63553
20.5695	21	0	35	0	0.1068295	1.7229178	10.805998	14.264361	299.98157
19.969501	13.8	0	23	0	0.07963304	1.6372924	8.287822	9.659814	2.4894993
18.0195	2.4	0	4	0	0.07097417	1.3237083	8.766573	18.331741	19.179106
17.469501	0	0	0	0	0.06605212	1.2319229	8.707238	24.288662	7.12493
17.1695	0	0	0	0	0.07063571	1.2105281	9.3669195	23.0653	2.2025495
17.0195	13.200001	0	22	0	0.07849409	1.1999502	10.464798	25.982424	3.9451082
17.5695	51.000004	0	85	0	0.09599704	1.2526503	12.599999	29.16889	360
17.9195	57.600002	0	96	0	0.10363226	1.2974379	13.32	29.528778	360
18.1695	40.800003	0	68	0	0.11356928	1.3353128	14.404499	29.88217	1.4320632
17.369501	32.4	0	54	0	0.09382261	1.2327945	12.429127	27.595823	349.99207
16.719501	31.2	0	52	0	0.09165313	1.1420397	9.826088	23.770906	8.426887

Figure 2 Weather data sample data

3.3. Data Description and Coverage

In the price and emissions dataset, we focus on four important attributes, including energy consumption, total emissions, emission intensity index as well as regional reference prices. The data covers five Australian states: New South Wales, Queensland, South Australia, Tasmania and Victoria. In the weather dataset, important weather variables such as temperature, wind speed, humidity and soil moisture are collected. Each row in the price and emission dataset corresponds to a 5-minute interval while each row in the weather dataset corresponds to a one-hour interval.

Our data covers the full annual record from September 2023 to August 2024. The data for this time period is sufficient to reflect seasonal variations, daily fluctuations, and relevant factors that may affect electricity prices and electricity demand.

4. Making the data fit for use

4.1. Data Quality

The two data sets accessed from the AEMO - the price dataset and the emissions dataset - are both kept to an incredibly high standard, by participators in the market who are interested in market transparency, planning and investment decision making.

After exploring all the data collected from the AEMO, no missing values were found in the data.

The data freshness was key to the analysis as attempting to identify future trends requires up-to-date analysis of recent data. The AEMO provides real-time, current data which is easily and consistently accessible through their website and API, which allows models developed to be easily kept up-to-date.

As mentioned, BOM was the recording source for the weather data, which being a trusted Australian government organisation, gives the data a high level of accuracy.

When exploring the data only one feature was found to have missing values. This was the snow_depth column. Exploring the values within this column revealed that they all had the exact same value of 0, which would be expected from a Brisbane snow depth value. Because all of the values were the same, no information useful for modelling was lost by dropping the column to deal with the missing values.

Like the AEMO data, Open-Meteo provides an up-to-date API making the data fresh and consistently accessible.

4.2. Data Transformation

We obtained five one-year datasets for state prices and five one-year datasets for emissions from the AEMO. The column names in these datasets include the respective state's name, which is appended based on the state information contained in the metadata. Then all 10 datasets were joined using the 5-minute interval datetime column, as each file used the same 5-minute intervals.

This resulted in a dataset with 21 columns…

SETTLEMENTDATE, QLD_Energy, QLD_Total_Emissions, QLD_Intensity_Index, QLD_RRP, ..., TAS_Energy, TAS_Total_Emissions, TAS_Intensity_Index, TAS_RRP

With the 4 columns for each state.

Because the weather data contains data every hour rather than every 5 minutes, the time intervals had to be aligned before joining it to the other dataset. The assumption has been made that the weather data over the course of an hour will change approximately linearly between the two stored values. Using this assumption, the weather data was interpolated between each hour to add 11 values corresponding to the 5-minute intervals between each hour. The interpolation was carried out linearly, utilising the two hourly data points that surround the values to be interpolated.

Once this was done the two datasets could be joined on their datetime columns with consistent 5-minute intervals. The resulting dataset comprises 52 columns: a timestamp for each record, 10 columns for price data, 10 columns for emissions data, and 31 columns for weather data. With approximately 105,408 rows, it created a CSV file of 57.5MB for analysis.

5. Making the data confess

5.1. Correlations Analysis of Energy Demand, Price and Influencing Variables

Following the exploratory data analysis, a correlation matrix was generated to quantitatively evaluate the relationships between Queensland's price and other variables. The first category of variables considered was the demand and price of different states of the NEM. This was to identify the influence of inter-regional interconnectors on price. There are six regional operational interconnectors: two interconnectors operating between NSW-QLD and SA-VIC and one interconnector between VIC-TAS and NSW-VIC. There is no direct physical interconnection between QLD-SA and NSW-SA.

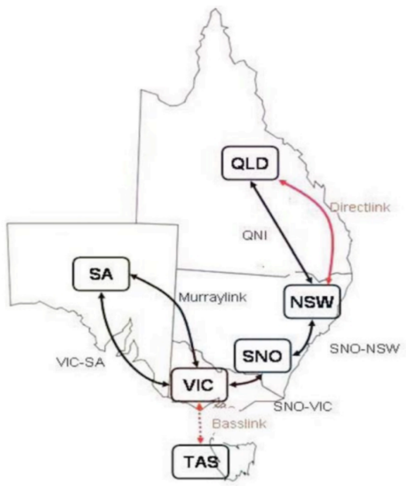


Figure 3 Interstate Energy Relationships

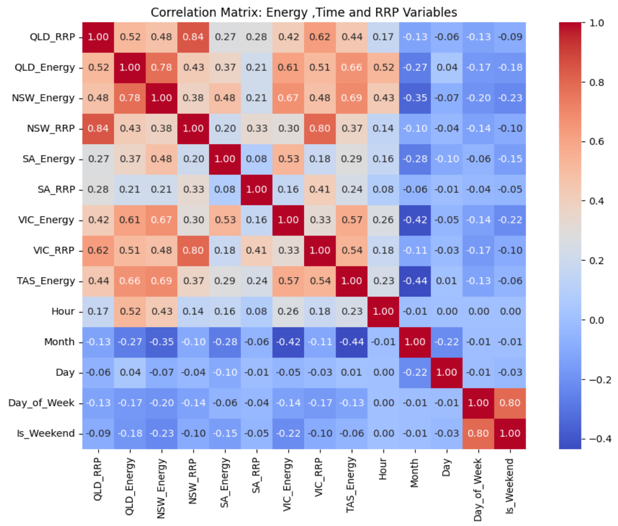


Figure 4 Correlation Matrix: Energy, Time and Price Variables

There is a strong correlation between the prices and demand of Queensland and New South Wales. However, there are only weak or moderate correlations between other states that have interconnection. These relationships need further investigation through more refined analysis.

The second category of variables chosen for correlation analysis are weather variables however they also showed no significant relationship. This may be because of the data quality.

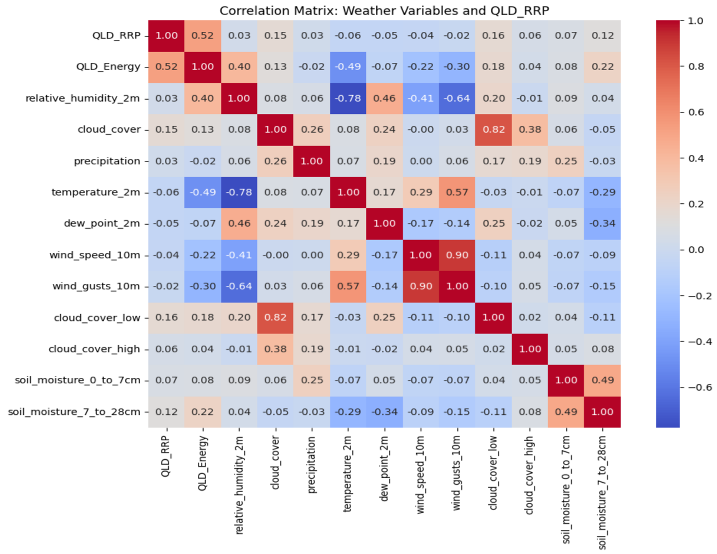


Figure 5 Correlation Matrix: Weather Variables and Energy Price in QLD

5.2. Evaluating Predictive Accuracy of Regression Models in Energy Pricing

To understand the relationships better, the variables were fitted in linear regression and higher-degree polynomial models. However, the model's performance at predicting the relationships was poor. The presence of outliers was also speculated. Taking this information and the suggestions from peer reviews into consideration, regression analysis with confidence intervals was implemented. A 95% confidence interval was used. As visualised in the plot, the prediction line closely follows the observed data indicating that the regression model was able to predict the overall pattern. In terms of relationships between Queensland's energy price and other variables, it can be interpreted that for the model, the independent variables are likely effective at predicting typical price patterns, which implies that there can be a **reasonable relationship** between the independent variables and Queensland's energy price when the price behaves normally. However, the presence of outliers can be detected, and it can also be observed that the independent variables fail to capture outliers.

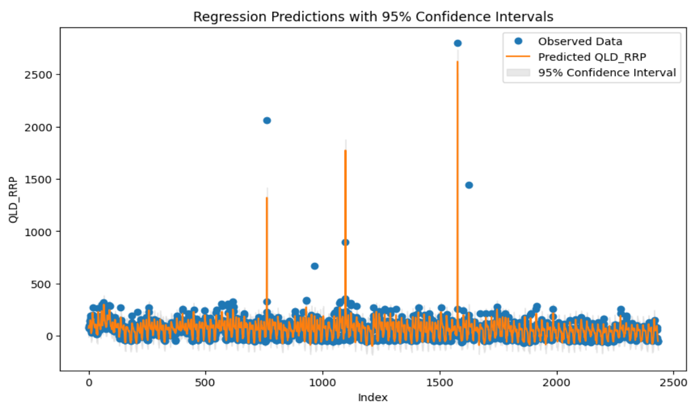


Figure 6 Regression Predictions with 95% Confidence Intervals

5.3. Assessing Outlier Effects on Model Predictions and Price Variations

In order to overcome the challenge of outliers, a robust regression model was used and compared to the previous regression model, it shows improvement in managing outliers. While it captures normal price variations well, it can be concluded that some additional variables may be needed to fully explain the extreme price spikes seen in the data.

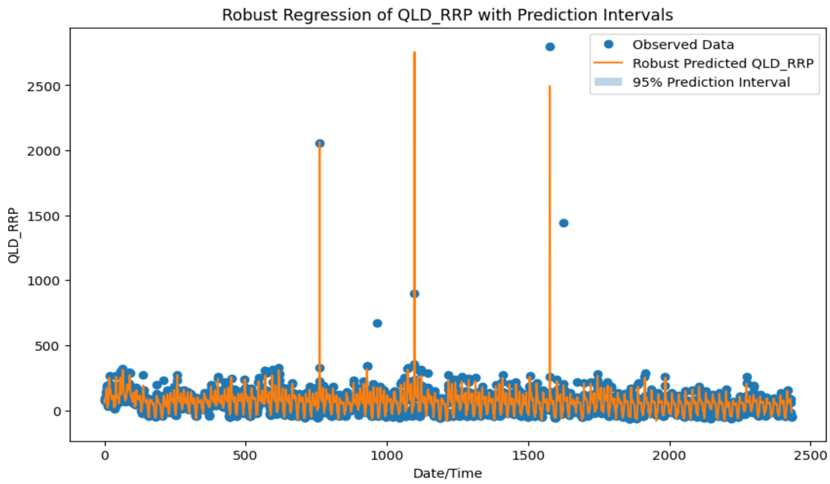


Figure 7 Robust Regression of QLD Energy price with prediction intervals

5.4. Determining Key Drivers of Energy Prices Through Feature Importance Analysis

In order to rank the feature importance of the existing variables, Support Vector Regression was used and the results suggest that Queensland's prices are primarily influenced by New South Wales's price, Queensland's energy demand, New South Wales's energy demand, Victoria's price and Victoria's energy demand. This relationship might be due to the transmission interconnectors and geographical closeness.

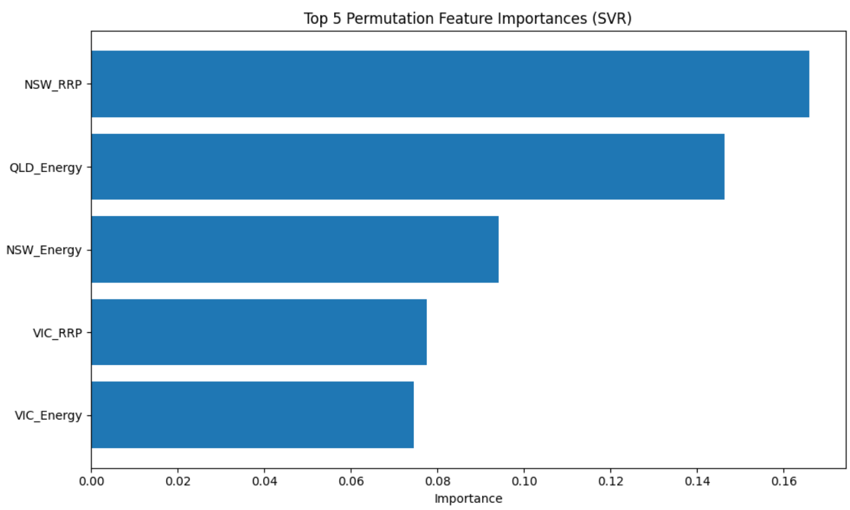


Figure 8 Top 5 Permutation Features Importances

6. Storytelling with Data

Analysis of the data revealed interesting patterns in energy demand and price across times of day and days of the week. This section of the report specifically considers Queensland data as it reflects the fluctuations observed in the national data.

6.1. QLD Energy Demand by Hour of Day

As can be seen from the chart below, it can be observed that the energy data demand changes throughout the day. There is a minor dip in the early morning around 3 am when most people are asleep and the low energy consumption is present for most homes (climate control most likely); then a peak in the early mornings around 6 am can be seen, likely due to people getting up and ready to go to work, using appliances like hot water systems, cooktops, and other personal appliances. Continuing into the day, another dip happens around 11 am when energy demand drops to the lowest of the day. This is likely to be from the fact that most people are out of their home and at work, and the residential address tends to be unoccupied; then it can be seen that the spike of energy demand starts from 3 pm all the way to peak at 6 pm when children are anticipated to come home in the afternoon, commuters return home around 5 pm and dinner and cooking happens around 6 pm; last but not least, the demand slowly drops from 7 pm onwards as people go to bed after 9 pm generally.

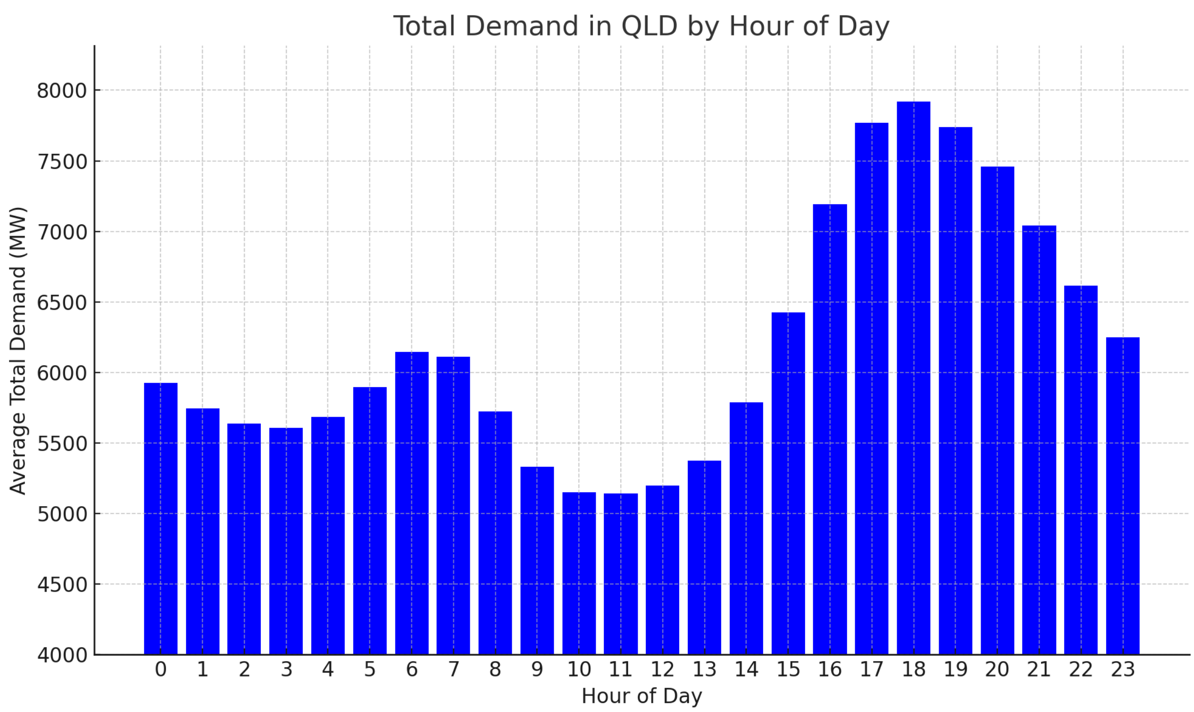


Figure 9 Total Demand in QLD by Hour of Day

6.2. QLD Energy Price by Hour of Day

The below graph demonstrates that electricity prices fluctuate greatly throughout the hours of the day. As we can see the hours between 8 pm and 7 am are quite stable and low cost, likely due to the fact that people are asleep and not many appliances are used. There is a slight dip at 3 am but not particularly strong. From the hour between 8 am to 3 pm, it can be observed that energy prices are the lowest, which may be due to the solar generation of houses with solar feed-in back to the grid, leaving an abundance of energy in those hours (10 am-12 pm being the cheapest where the sun is usually the strongest). Lastly, energy prices spike from 5 pm to 7 pm where the demand for electricity surges to the peak and solar is no longer providing energy by that point.

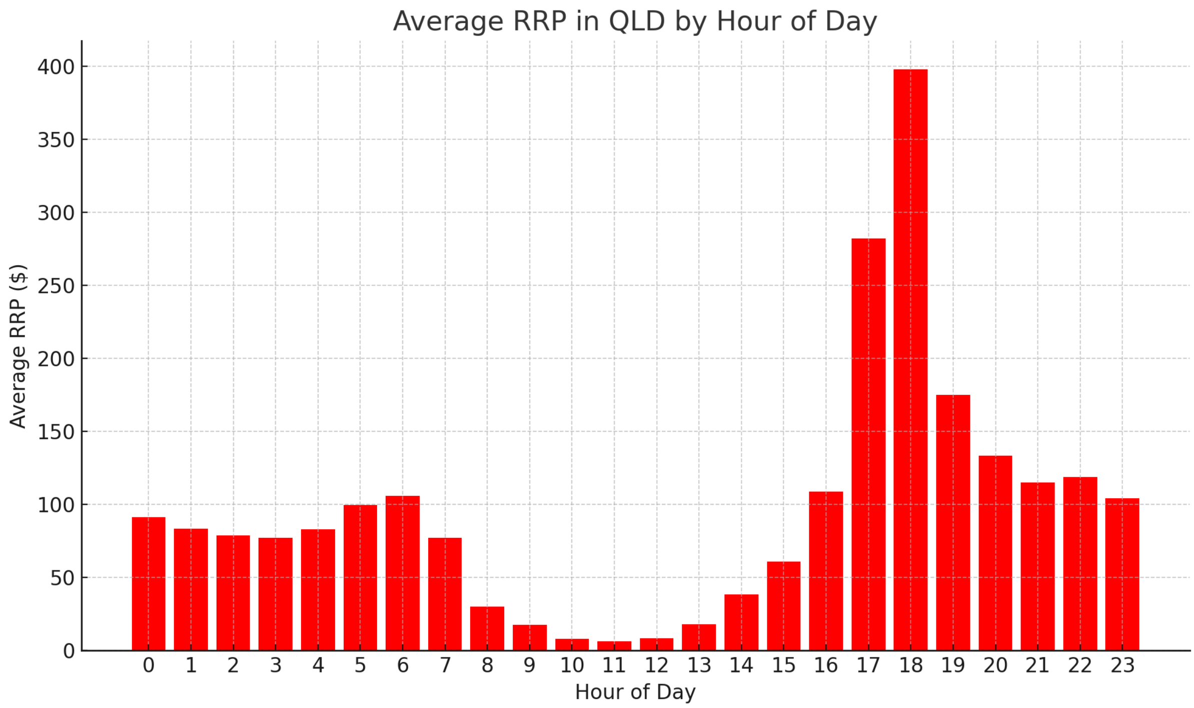


Figure 10 Average QLD Energy Price by Hour of Day

6.3. QLD Energy Demand by Energy Price by Hour of Day

This is a chart that combines the 2 layers of data to enable a better visualisation. It demonstrates that cost of energy almost perfectly matches the demand for energy.

Both data sets showed a minor dip around 3 am and a minor peak around 6 am. The lowest point is around 11 am and the absolute peak is around 6 pm with both data on a downward trajectory from 7 pm till midnight.

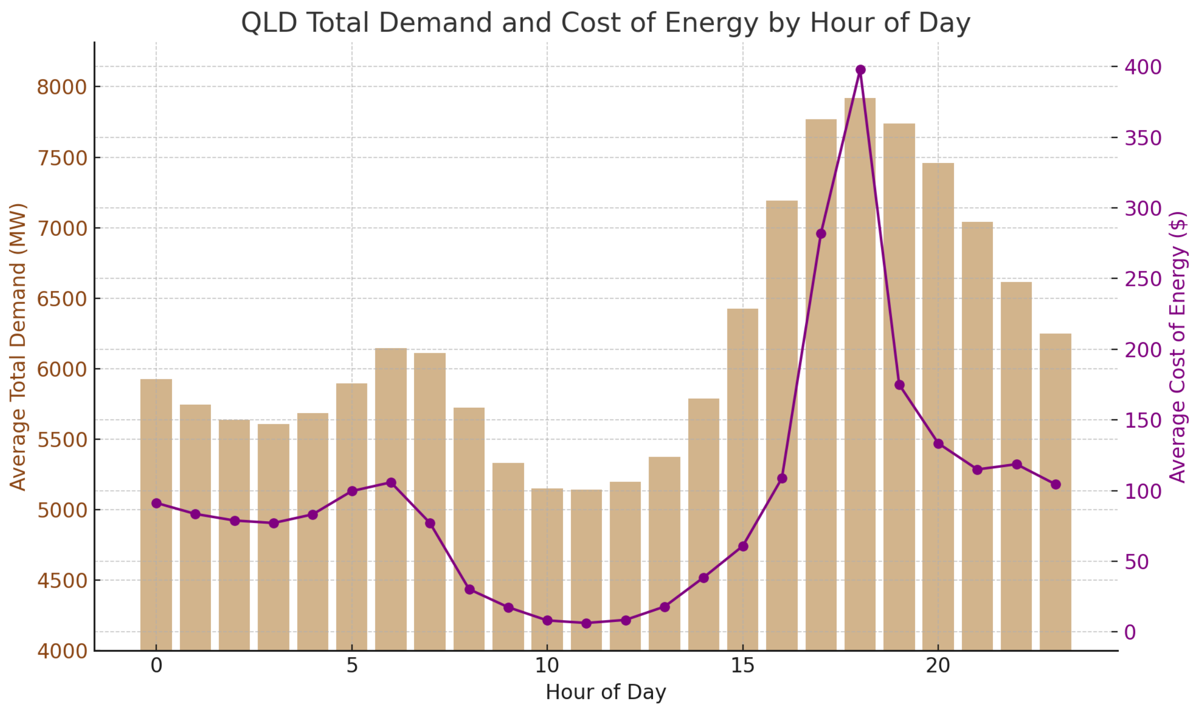


Figure 11 QLD Total Demand and Cost of Energy by Hour of Day

6.4. QLD Energy Demand by Energy Price by Day of Week

Another aspect we are seeking to better understand is whether the demand and cost of energy vary throughout the days of the week. As observed in Figure 12, demand for energy drops during the weekend and it perfectly explains why the energy cost drops for the weekend as well. This is likely due to people going out of their house during the weekend and leaving the house to minimum use.

However, it is difficult to tell the reason for why fluctuation of energy prices on the weekdays when the demand is very much level across the 5 weekdays but the energy cost is the highest on Monday and Friday but lowest on Tuesday and Wednesday.

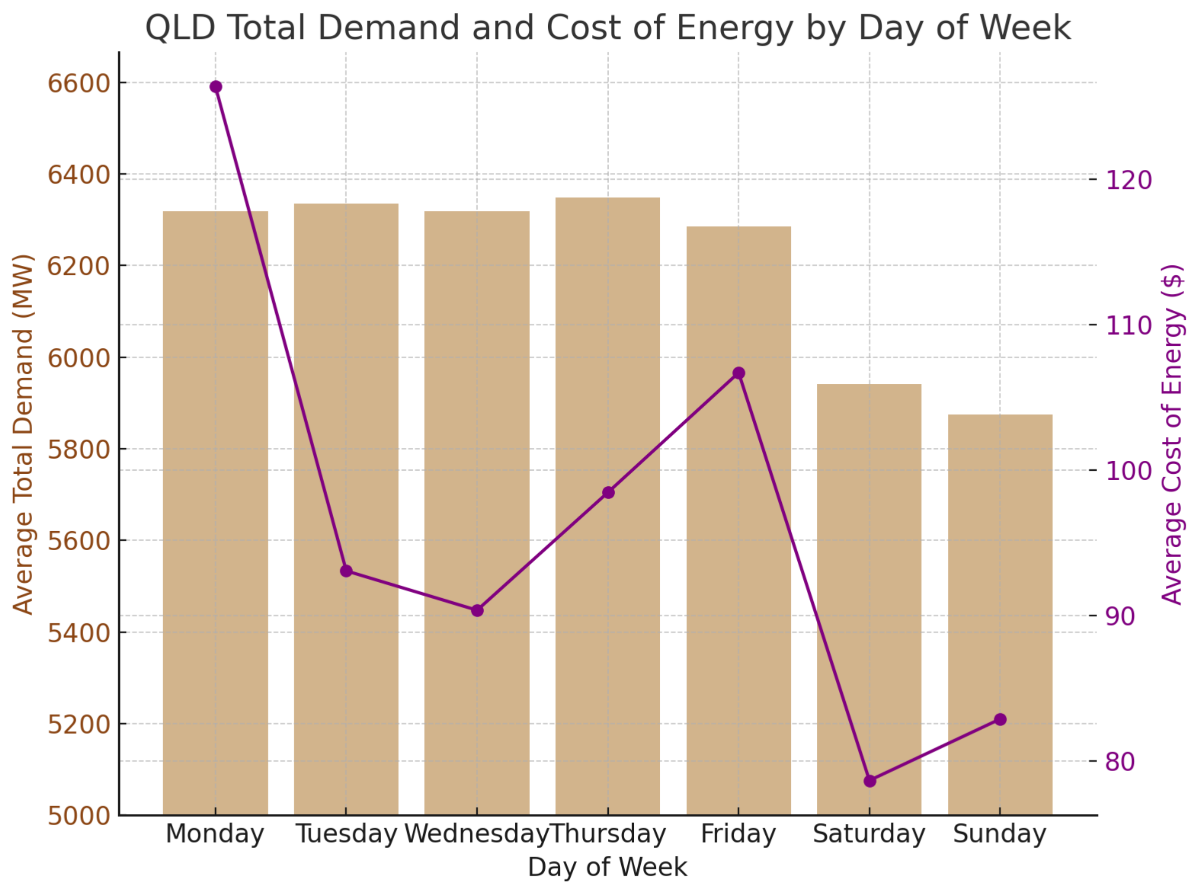


Figure 12 QLD Total Demand and Cost of Energy by Day of Week

6.5. QLD Energy price in correlation to temperature data

Alongside the observations noted above, we also examined weather data to explore potential correlations between weather patterns and energy costs.

Exploratory analysis revealed very weak relationships with most of the weather attributes. By way of example, Figure 13 is a visualisation focused on temperature at 2 meters. It shows that the temperature and cost of energy do not bear any positive correlation which leads to the conclusion that there is no correlation between energy price and temperature on the ground.

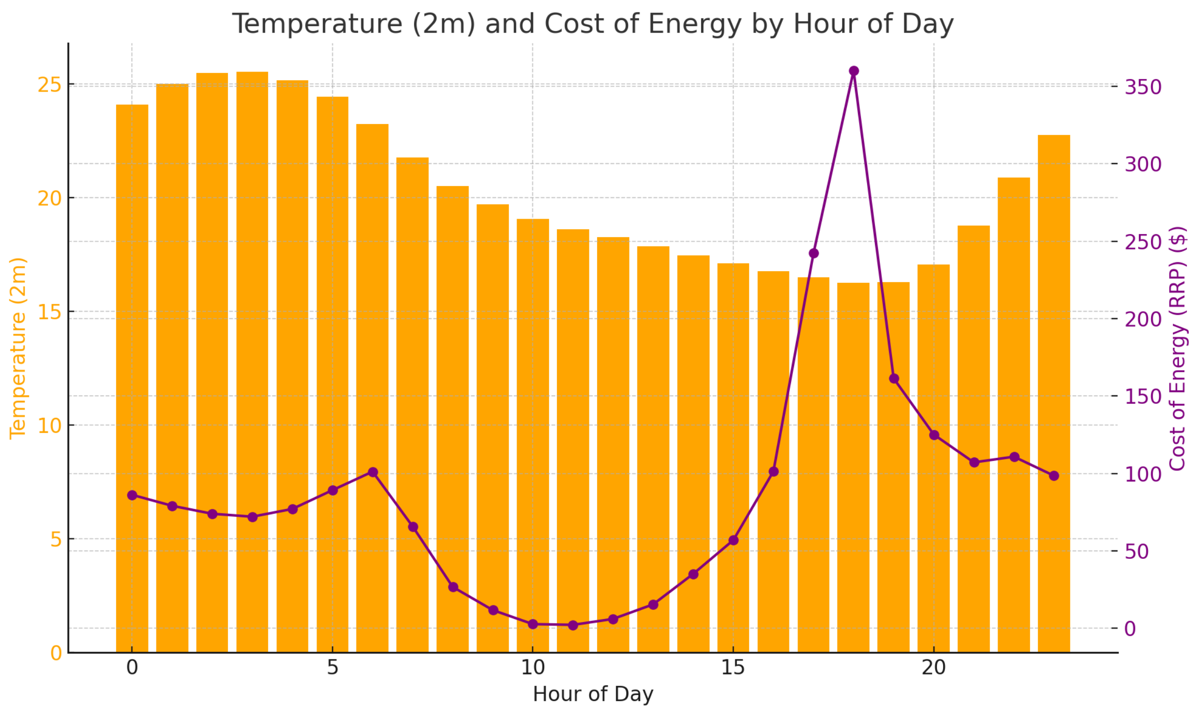


Figure 13 QLD Energy Price and Temperature by Hour of Day

7. Responses to feedback comments

7.1. Problem-solving with Data

A peer review recommended that we specify the distinct value our project delivers to each stakeholder. While time constraints prevented us from including this in our presentation, we have detailed the benefits for various stakeholders in the essay.

7.2. Getting Data I Need

After receiving the feedback comments, numerical data as the data type was the focus point of the analysis instead of csv data which is a format. Besides, we added a new dataset, the weather dataset for analysis.

7.3. Is my Data Fit for Use

It was mentioned in the peer-reviews that rather than just using the one internal data set from the AEMO, other weather or economic indicators should be added. Therefore 2 other data sets were collected and joined to the initial dataset - a weather dataset and an emissions dataset, as these may influence both the demand and supply of energy.

7.4. Making the Data Confess

The outputs presented during the trial presentations used linear regression and polynomial models of higher degree. However, the performance of the models was poor and they were not good at capturing the relationships. The peer reviews suggested implementing confidence intervals to help quantify the uncertainty in model predictions, given that in complex systems like energy pricing, external factors might influence trends. This would also provide stakeholders a clearer understanding of the range of potential outcomes and help identify the reliability of the model's predictions. Adhering to this, data was fitted in a regression model with confidence intervals.

The other suggestion was to implement complex models to better capture the trends and handle outliers. Therefore, a robust regression model was implemented. Additionally, to explore the feature importance among variables, Support Vector Regression was also used.

Both of the peer review suggestions resulted in better outputs and refined results

7.5. Storytelling with Data

The lecturer's feedback suggested that less use of technical graphs and instead use of plain language and non-technical charts to explain to stakeholders who do not have a data science background. This is incorporated into the final presentation as well as this report.

8. Conclusion

It has been discovered that there is a strong correlation between energy price and energy demand and throughout the hours of the day, we also discovered a pattern between energy cost and energy demand across the days of the week, with weekends showing the highest correlation and weekdays showing minor correlation. However, no evident relationship was found between energy data and weather data.

8.1. Answer the research questions

Seasonal trends in energy demand

The analysis identified a clear pattern in the cost of energy and demand throughout day and week. It is cheapest to use energy throughout the day after 10am till 3pm and most expensive to use energy around 6pm. Weekends are cheaper than weekdays in general and Tuesday and Wednesday are cheaper than the other weekdays.

Economic and social factors influencing energy demand

The study highlights that weekdays and weekends are quite different in terms of energy demand with weekends being much lower and weekdays being much higher. This suggests variations in energy demand in residential versus commercial energy use, thus explaining the workweek schedule.

Impact of renewable energy on demand patterns

As can be seen in the daily fluctuation patterns, the energy cost remains low throughout the daytime from sunrise to sunset. This suggests that the solar energy input could be contributing to the grid in those hours, thus reducing the cost of energy for when the sun is up.

8.2. Suggestions to the Stakeholders

For the energy producers, it is recommended to adapt the energy production schedule to align with the increase of demand during early morning and particularly early evenings and to consider a scale-back solution for production when energy demand drops in the middle of the day.

For consumers, it would be wise to adjust usage habits to shift energy usage to off-peak times to lower the overall cost. It is also good practice to invest in energy-efficient appliances and systems to reduce overall consumption.

For the policymakers, our suggestions would be to incentivise off-peak usage and implement policies that encourage time-of-use pricing schemes to help flatten the demand curves; promote energy efficient practices; and to support renewable energy generations.

References

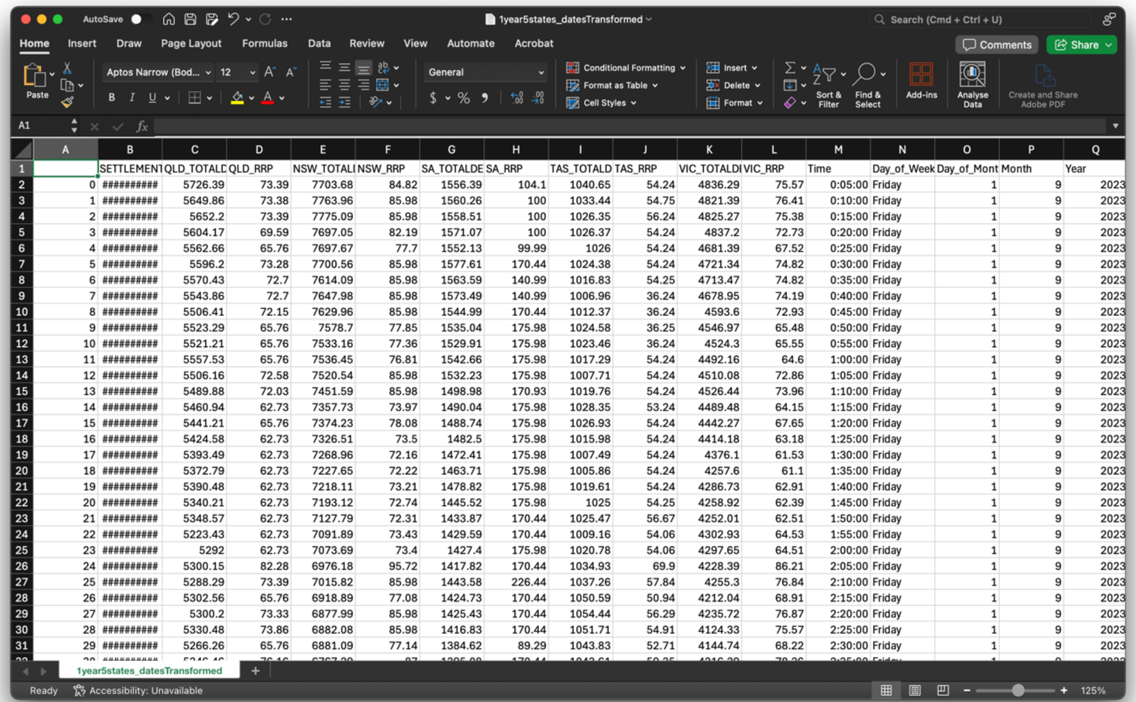
Open-Meteo. (2022). Historical weather API. Retrieved October 24, 2024, from https://open-meteo.com/en/docs/historical-weather-api

Australian Energy Market Operator. (2022). Data NEM: Aggregated data. Retrieved October 24, 2024, from https://aemo.com.au/energy-systems/electricity/national-electricity-market-nem/data-nem/aggregated-data

Appendix

Appendix A - Datasets

1year5states_datesTransformed.csv



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1		SETTLEMENT	QLD_TOTALD	QLD_RRP	NSW_TOTALI	NSW_RRP	SA_TOTALDE	SA_RRP	TAS_TOTALD	TAS_RRP	VIC_TOTALDI	VIC_RRP	Time	Day_of_Week	Day_of_Mont	Month	Year
2	0	##########	5726.39	73.39	7703.68	84.82	1556.39	104.1	1040.65	54.24	4836.29	75.57	0:05:00	Friday	1	9	2023
3	1	##########	5649.86	73.38	7763.96	85.98	1560.26	100	1033.44	54.75	4821.39	76.41	0:10:00	Friday	1	9	2023
4	2	##########	5652.2	73.39	7775.09	85.98	1558.51	100	1026.35	56.24	4825.27	75.38	0:15:00	Friday	1	9	2023
5	3	##########	5604.17	69.59	7697.05	82.19	1571.07	100	1026.37	54.24	4837.2	72.73	0:20:00	Friday	1	9	2023
6	4	##########	5562.66	65.76	7697.67	77.7	1552.13	99.99	1026	54.24	4681.39	67.52	0:25:00	Friday	1	9	2023
7	5	##########	5596.2	73.28	7700.56	85.98	1577.61	170.44	1024.38	54.24	4721.34	74.82	0:30:00	Friday	1	9	2023
8	6	##########	5570.43	72.7	7614.09	85.98	1563.59	140.99	1016.83	54.25	4713.47	74.82	0:35:00	Friday	1	9	2023
9	7	##########	5543.86	72.7	7647.98	85.98	1573.49	140.99	1006.96	36.24	4678.95	74.19	0:40:00	Friday	1	9	2023
10	8	##########	5506.41	72.15	7629.96	85.98	1544.99	170.44	1012.37	36.24	4593.6	72.93	0:45:00	Friday	1	9	2023
11	9	##########	5523.29	65.76	7578.7	77.85	1535.04	175.98	1024.58	36.25	4546.97	65.48	0:50:00	Friday	1	9	2023
12	10	##########	5521.21	65.76	7533.16	77.36	1529.91	175.98	1023.46	36.24	4524.3	65.55	0:55:00	Friday	1	9	2023
13	11	##########	5557.53	65.76	7536.45	76.81	1542.66	175.98	1017.29	54.24	4492.16	64.6	1:00:00	Friday	1	9	2023
14	12	##########	5506.16	72.58	7520.54	85.98	1532.23	175.98	1007.71	54.24	4510.08	72.86	1:05:00	Friday	1	9	2023
15	13	##########	5489.88	72.03	7451.59	85.98	1498.98	170.93	1019.76	54.24	4526.44	73.96	1:10:00	Friday	1	9	2023
16	14	##########	5460.94	62.73	7357.73	73.97	1490.04	175.98	1028.35	53.24	4489.48	64.15	1:15:00	Friday	1	9	2023
17	15	##########	5441.21	65.76	7374.23	78.08	1488.74	175.98	1026.93	54.24	4442.27	67.65	1:20:00	Friday	1	9	2023
18	16	##########	5424.58	62.73	7326.51	73.5	1482.5	175.98	1015.98	54.24	4414.18	63.18	1:25:00	Friday	1	9	2023
19	17	##########	5393.49	62.73	7268.96	72.16	1472.41	175.98	1007.49	54.24	4376.1	61.53	1:30:00	Friday	1	9	2023
20	18	##########	5372.79	62.73	7227.65	72.22	1463.71	175.98	1005.86	54.24	4257.6	61.1	1:35:00	Friday	1	9	2023
21	19	##########	5390.48	62.73	7218.11	73.21	1478.82	175.98	1019.61	54.24	4286.73	62.91	1:40:00	Friday	1	9	2023
22	20	##########	5340.21	62.73	7193.12	72.74	1445.52	175.98	1025	54.25	4258.92	62.39	1:45:00	Friday	1	9	2023
23	21	##########	5348.57	62.73	7127.79	72.31	1433.87	170.44	1025.47	56.67	4252.01	62.51	1:50:00	Friday	1	9	2023
24	22	##########	5223.43	62.73	7091.89	73.43	1429.59	170.44	1009.16	54.06	4302.93	64.53	1:55:00	Friday	1	9	2023
25	23	##########	5292	62.73	7073.69	73.4	1427.4	175.98	1020.78	54.06	4297.65	64.51	2:00:00	Friday	1	9	2023
26	24	##########	5300.15	82.28	6976.18	95.72	1417.82	170.44	1034.93	69.9	4228.39	86.21	2:05:00	Friday	1	9	2023
27	25	##########	5288.29	73.39	7015.82	85.98	1443.58	226.44	1037.26	57.84	4255.3	76.84	2:10:00	Friday	1	9	2023
28	26	##########	5302.56	65.76	6918.89	77.08	1424.73	170.44	1050.59	50.94	4212.04	68.91	2:15:00	Friday	1	9	2023
29	27	##########	5300.2	73.33	6877.99	85.98	1425.43	170.44	1054.44	56.29	4235.72	76.87	2:20:00	Friday	1	9	2023
30	28	##########	5330.48	73.86	6882.08	85.98	1416.83	170.44	1051.71	54.91	4124.33	75.57	2:25:00	Friday	1	9	2023
31	29	##########	5266.26	65.76	6881.09	77.14	1384.62	89.29	1043.83	52.71	4144.74	68.22	2:30:00	Friday	1	9	2023

rrp_weather_emiss.csv

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1		SETTLEMENT	NSW_Energy	NSW_Total_E	NSW_Intensi	NSW_RRP	QLD_Energy	QLD_Total_E	QLD_Intensit	QLD_RRP	SA_Energy	SA_Total_Em	SA_Intensity_	SA_RRP	TAS_Energy	TAS_Total_E
2	0	##########	0	0	0	114.26865	0	0	0	115.101	0	0	0	113.61003	0	0
3	1	##########	652.303436	453.679930	0.69550443	114.29432	529.11339	370.113890	0.6994982	114.2334	84.5747808	14.66857	0.17343905	114.17166	142.104884	0
4	2	##########	651.496523	455.850746	0.69969789	115.05796	522.36119	365.681743	0.70005535	114.94	84.7558995	14.2760931	0.16843775	114.17302	141.35985	0
5	3	##########	652.575715	455.97181	0.69872629	97.33	518.01806	362.657123	0.70008587	97.73575	85.5170254	14.3644588	0.16797191	99.37507	140.220523	0
6	4	##########	654.478152	456.547808	0.69757532	91.17679	516.448118	361.862308	0.700675	91.51198	86.5755316	14.5991477	0.168629	91.70988	140.581935	0
7	5	##########	654.546964	456.648426	0.69765570	93.00305	514.349221	361.219116	0.7022838	92.70147	84.7631933	14.4878289	0.17092122	94.64056	140.942354	0
8	6	##########	650.360444	455.041539	0.69967591	93.70933	511.821710	359.861215	0.70309877	93.93909	83.0814604	14.2976896	0.17209242	95.1032	141.61759	0
9	7	##########	643.66548	450.565731	0.6999998	89.73143	510.297266	358.437785	0.70240977	89.29919	82.4234441	14.2301870	0.17264732	91.09939	142.162533	0
10	8	##########	637.195504	443.769465	0.69644161	75.05012	509.309767	357.890392	0.70269689	74.11395	84.056074	14.1299781	0.16810181	76.23824	140.053458	0
11	9	##########	631.808368	437.996188	0.69324214	67.07047	510.433028	358.621923	0.70258369	65.74	86.111639	14.2860545	0.16590155	67.82206	138.346641	0
12	10	##########	627.146189	433.843496	0.69177410	86.22931	511.921774	359.57217	0.70239670	84.55961	85.4934866	14.5657091	0.17037215	86.51273	138.338060	0
13	11	##########	621.128346	428.734668	0.6902513	88.95	510.391365	359.057344	0.70349415	86.03164	84.5345883	14.4115146	0.17048068	89.15992	139.777138	0
14	12	##########	615.516565	423.70463	0.68837242	89.2098	509.916598	358.663184	0.70337617	87.30663	83.8507949	14.1546049	0.16880704	88.9531	140.086737	0
15	13	##########	614.36982	421.850205	0.68663887	71.0099	511.272617	359.03137	0.70223077	68.14475	83.060521	14.4090084	0.17347601	69.86282	137.660505	0
16	14	##########	611.94406	420.186011	0.68664120	75.84089	511.5416	358.642858	0.70110203	73.26169	82.4566770	14.6515691	0.17768808	75.14674	136.481563	0
17	15	##########	605.54091	416.756033	0.68823761	88.59608	511.125960	357.922177	0.7002622	86.12602	79.2157316	14.5629953	0.18383968	87.46353	137.954122	0
18	16	##########	598.33573	412.796091	0.68990714	88.9	510.47188	357.213785	0.69977171	85.80228	75.5996908	14.4319615	0.19089974	88.13278	138.89475	0
19	17	##########	590.21315	408.175813	0.69157357	69.57586	508.888209	356.010842	0.6995856	65.74	73.5193824	14.2638338	0.1940146	69.49854	136.977852	0
20	18	##########	586.099877	405.817240	0.69240287	88.9	502.062945	355.990539	0.70905559	87.28558	74.1646229	13.9703938	0.18837005	48.4209	137.061344	0
21	19	##########	586.85993	406.199948	0.69215825	89.2102	496.715867	357.046050	0.71881345	86.96802	78.0608891	13.6785186	0.17522883	58.74489	138.662286	0
22	20	##########	585.780267	405.361416	0.69200251	76.68656	497.367353	357.166204	0.7181135	74.23	79.9659920	13.6014709	0.17009069	56.7587	138.730112	0
23	21	##########	584.727227	404.355957	0.69152920	88.9	498.694055	357.50845	0.71688933	85.42393	78.5652945	13.6353262	0.17355406	88.922	138.416489	0
24	22	##########	582.38788	401.866629	0.69003261	67.96765	498.800245	357.234067	0.7161866	65.74	73.5841991	13.4801902	0.18319408	43.86094	138.27484	0
25	23	##########	577.175112	398.091785	0.68972444	84.48459	498.58954	357.127869	0.7162763	81.11789	68.8931866	13.3468586	0.19373263	52.38084	139.76274	0
26	24	##########	572.365117	394.730561	0.68964818	84.43453	499.087693	357.636458	0.7165804	81.00406	67.3243958	13.3686631	0.19857085	54.11372	141.479658	0
27	25	##########	569.460462	392.060059	0.6884763	68.60212	498.863012	357.377945	0.71638493	65.74	66.9237933	13.5347009	0.20224049	45.01202	142.212555	0
28	26	##########	564.250656	388.812780	0.68907811	86.35891	500.725518	358.767857	0.71649605	82.82	65.913624	13.4801900	0.20451295	54.19794	141.766856	0
29	27	##########	557.91664	384.128207	0.68850466	69.56521	503.335242	360.405661	0.71603502	65.74	63.9047866	13.380922	0.20938841	139.43	139.987355	0
30	28	##########	555.28026	380.552095	0.68533337	68.39042	502.4237	359.205538	0.71494545	65.49	62.9147945	13.4463	0.21372294	39.14479	138.168985	0

Appendix B - Code

data_transformation.py

```python
import pandas as pd


def load_months_states(state_list=['QLD', 'NSW', 'SA', 'TAS', 'VIC'], month_list=['202309', '202310', '202311', '202312',
    final_df = pd.DataFrame()
    for state in state_list:
        month_dfs = []
        for month in month_list:
            df = pd.read_csv(f"{path}PRICE_AND_DEMAND_{month}_{state}1.csv")

            df['SETTLEMENTDATE'] = pd.to_datetime(df['SETTLEMENTDATE'], format='%Y/%m/%d %H:%M:%S')

            df[f'{state}_TOTALDEMAND'] = df['TOTALDEMAND']
            df[f'{state}_RRP'] = df['RRP']
            df.drop(columns=['PERIODTYPE', 'REGION', 'TOTALDEMAND', 'RRP'], inplace=True)

            month_dfs.append(df)

        state_df = pd.concat(month_dfs, ignore_index=True)

        if final_df.empty:
            final_df = state_df
        else:
            final_df = pd.merge(final_df, state_df, on='SETTLEMENTDATE', how='inner')

    final_df['Time'] = final_df['SETTLEMENTDATE'].dt.time
    final_df['Day_of_Week'] = final_df['SETTLEMENTDATE'].dt.day_name()
    final_df['Day_of_Month'] = final_df['SETTLEMENTDATE'].dt.day
    final_df['Month'] = final_df['SETTLEMENTDATE'].dt.month
    final_df['Year'] = final_df['SETTLEMENTDATE'].dt.year

    return final_df

df = load_months_states()

df.to_csv('1year5states_datesTransformed.csv')

```

weather_and_prices.py

```
import nemed

import nemosis

import pandas as pd

import numpy as np

import logging

import openmeteo_requests

import requests_cache

from retry_requests import retry


def load_price_data(start_date="2023-09-01", end_date="2024-09-01", cache="E:\\TEMPCACHE_cemed_demo") -> pd.DataFrame:

    logging.getLogger("nemosis").setLevel(logging.WARNING)

    prices = nemosis.dynamic_data_compiler(start_time=start_date.replace("-", "/") + " 00:00:00", end_time=end_date.replace("-", "/") + " 00:00:00", table_name='DISPATCHPRICE', raw_data_location=cache, fformat='csv', select_columns='all', keep_csv=False)

    prices = prices[['SETTLEMENTDATE', 'REGIONID', 'RRP']]

    prices['SETTLEMENTDATE'] = pd.to_datetime(prices['SETTLEMENTDATE'], format='%Y/%m/%d %H:%M:%S')

    return prices


def load_emmisions_data(start_date="2024-08-01", end_date="2024-09-01", cache="E:\\TEMPCACHE_cemed_demo"):

    logging.getLogger("nemed").setLevel(logging.WARNING)

    total = nemed.get_total_emissions(start_time=start_date.replace("-", "/") + " 00:00", end_time=end_date.replace("-", "/") + " 00:00", cache=cache, by=None)

    total['SETTLEMENTDATE'] = pd.to_datetime(total['TimeEnding'], format='%Y/%m/%d %H:%M:%S')

    total.drop(columns=['TimeEnding'], inplace=True)

    return total


def get_weather_data(start_date="2023-09-01", end_date="2024-09-01") -> pd.DataFrame:

    cache_session = requests_cache.CachedSession('.cache', expire_after=-1)

    retry_session = retry(cache_session, retries=5, backoff_factor=0.2)

    openmeteo = openmeteo_requests.Client(session=retry_session)

    url = "https://archive-api.open-meteo.com/v1/archive"

    params = {

        "latitude": -27.4679,

        "longitude": 153.0281,

        "start_date": start_date,

        "end_date": end_date,

        "hourly": ["temperature_2m", "relative_humidity_2m", "dew_point_2m", "apparent_temperature", "precipitation", "rain", "snowfall", "snow_depth", "weather_code", "pressure_msl", "surface_pressure", "cloud_cover", "cloud_cover_low", "cloud_cover_mid", "cloud_cover_high", "et0_fao_evapotranspiration", "vapour_pressure_deficit", "wind_speed_10m", "wind_speed_100m", "wind_direction_10m", "wind_direction_100m", "wind_gusts_10m", "soil_temperature_0_to_7cm", "soil_temperature_7_to_28cm", "soil_temperature_28_to_100cm", "soil_temperature_100_to_255cm", "soil_moisture_0_to_7cm", "soil_moisture_7_to_28cm", "soil_moisture_28_to_100cm", "soil_moisture_100_to_255cm"],
```

```
        "timezone": "Australia/Sydney"
    }
    responses = openmeteo.weather_api(url, params=params)
    response = responses[0]
    hourly = response.Hourly()
    hourly_data = {"date": pd.date_range(start=pd.to_datetime(hourly.Time(), unit="s", utc=True), end=pd.to_datetime(hourly.TimeEnd(), unit="s", utc=True), freq=pd.Timedelta(seconds=hourly.Interval()), inclusive="left")}
    for i, variable in enumerate(hourly.Variables()):
        hourly_data[variable.Description()] = variable.ValuesAsNumpy()
    hourly_dataframe = pd.DataFrame(data=hourly_data)
    hourly_dataframe['SETTLEMENTDATE'] = pd.to_datetime(hourly_dataframe['date'].astype(str).str.replace('+00:00', '', regex=True), format='%Y/%m/%d %H:%M:%S')
    hourly_dataframe.drop(columns='date', inplace=True)
    return hourly_dataframe


def five_min_weather_interpolation(df: pd.DataFrame) -> pd.DataFrame:
    df.set_index('SETTLEMENTDATE', inplace=True)
    df_resampled = df.resample('5T').interpolate(method='linear')
    df_resampled.reset_index(inplace=True)
    return df_resampled


def join_all_data(start_date="2023-07-01", end_date="2024-07-01", cache="E:\\TEMPCACHE_cemed_demo"):
    print('Starting weather download')
    weather_df = get_weather_data(start_date=start_date, end_date=end_date)
    weather_df = five_min_weather_interpolation(weather_df)
    print('Weather download complete... starting price download')
    price_df = load_price_data(start_date=start_date, end_date=end_date, cache=cache)
    print('Price download complete... starting emissions download')
    emiss_df = load_emmisions_data(start_date=start_date, end_date=end_date, cache=cache)
    print('Emissions download complete')
    final_df = pd.DataFrame()
    for region in price_df['REGIONID'].unique():
        region_cleaned = region.replace('1', '')
        region_df_price = price_df[price_df['REGIONID'] == region].copy()
        region_df_price.drop(columns='REGIONID', inplace=True)
        region_df_price.rename(columns={'RRP': f'{region_cleaned}_RRP'}, inplace=True)
        region_df_emiss = emiss_df[emiss_df['Region'] == region].copy()
        region_df_emiss.drop(columns='Region', inplace=True)
        region_df_emiss.rename(columns={'Energy': f'{region_cleaned}_Energy', 'Total_Emissions': f'{region_cleaned}_Total_Emissions', 'Intensity_Index': f'{region_cleaned}_Intensity_Index'}, inplace=True)
        region_df = pd.merge(region_df_emiss, region_df_price, on='SETTLEMENTDATE', how='inner')
```

```
        if final_df.empty:
            final_df = region_df
        else:
            final_df = pd.merge(final_df, region_df, on='SETTLEMENTDATE', how='inner')
    final_df = pd.merge(final_df, weather_df, on='SETTLEMENTDATE', how='inner')
    cols = ['SETTLEMENTDATE'] + [col for col in final_df.columns if col != 'SETTLEMENTDATE']
    final_df = final_df[cols]
    print('Transformation complete')
    return final_df

df = join_all_data(start_date="2023-07-01", end_date="2024-07-01", cache="E:\\TEMPCACHE_cemed_demo")
df.to_csv('rrp_weather_emiss.csv')
```

weather_api.py

```
import openmeteo_requests
import requests_cache
import pandas as pd
from retry_requests import retry
import pytz

def get_weather_data(start_date="2023-08-29", end_date="2024-09-02") -> pd.DataFrame:
    cache_session = requests_cache.CachedSession('.cache', expire_after=-1)
    retry_session = retry(cache_session, retries=5, backoff_factor=0.2)
    openmeteo = openmeteo_requests.Client(session=retry_session)
    url = "https://archive-api.open-meteo.com/v1/archive"
    params = {
        "latitude": -23.5535,
        "longitude": 145.2854,
        "start_date": start_date,
        "end_date": end_date,
        "hourly": ["temperature_2m", "relative_humidity_2m", "dew_point_2m", "apparent_temperature", "precipitation", "rain", "snowfall", "snow_depth", "weather_code", "pressure_msl", "surface_pressure", "cloud_cover", "cloud_cover_low", "cloud_cover_mid", "cloud_cover_high", "et0_fao_evapotranspiration", "vapour_pressure_deficit", "wind_speed_10m", "wind_speed_100m", "wind_direction_10m", "wind_direction_100m", "wind_gusts_10m", "soil_temperature_0_to_7cm", "soil_temperature_7_to_28cm", "soil_temperature_28_to_100cm", "soil_temperature_100_to_255cm", "soil_moisture_0_to_7cm", "soil_moisture_7_to_28cm", "soil_moisture_28_to_100cm", "soil_moisture_100_to_255cm"],
        "timezone": "Australia/Sydney"
    }
    responses = openmeteo.weather_api(url, params=params)
    response = responses[0]
```

```
    hourly = response.Hourly()

    hourly_data = {

        "date": pd.date_range(start=pd.to_datetime(hourly.Time(), unit="s", utc=True), end=pd.to_datetime(hourly.TimeEnd(), unit="s",
utc=True), freq=pd.Timedelta(seconds=hourly.Interval()), inclusive="left"),

        "temperature_2m": hourly.Variables(0).ValuesAsNumpy(),

        "relative_humidity_2m": hourly.Variables(1).ValuesAsNumpy(),

        "dew_point_2m": hourly.Variables(2).ValuesAsNumpy(),

        "apparent_temperature": hourly.Variables(3).ValuesAsNumpy(),

        "precipitation": hourly.Variables(4).ValuesAsNumpy(),

        "rain": hourly.Variables(5).ValuesAsNumpy(),

        "snowfall": hourly.Variables(6).ValuesAsNumpy(),

        "snow_depth": hourly.Variables(7).ValuesAsNumpy(),

        "weather_code": hourly.Variables(8).ValuesAsNumpy(),

        "pressure_msl": hourly.Variables(9).ValuesAsNumpy(),

        "surface_pressure": hourly.Variables(10).ValuesAsNumpy(),

        "cloud_cover": hourly.Variables(11).ValuesAsNumpy(),

        "cloud_cover_low": hourly.Variables(12).ValuesAsNumpy(),

        "cloud_cover_mid": hourly.Variables(13).ValuesAsNumpy(),

        "cloud_cover_high": hourly.Variables(14).ValuesAsNumpy(),

        "et0_fao_evapotranspiration": hourly.Variables(15).ValuesAsNumpy(),

        "vapour_pressure_deficit": hourly.Variables(16).ValuesAsNumpy(),

        "wind_speed_10m": hourly.Variables(17).ValuesAsNumpy(),

        "wind_speed_100m": hourly.Variables(18).ValuesAsNumpy(),

        "wind_direction_10m": hourly.Variables(19).ValuesAsNumpy(),

        "wind_direction_100m": hourly.Variables(20).ValuesAsNumpy(),

        "wind_gusts_10m": hourly.Variables(21).ValuesAsNumpy(),

        "soil_temperature_0_to_7cm": hourly.Variables(22).ValuesAsNumpy(),

        "soil_temperature_7_to_28cm": hourly.Variables(23).ValuesAsNumpy(),

        "soil_temperature_28_to_100cm": hourly.Variables(24).ValuesAsNumpy(),

        "soil_temperature_100_to_255cm": hourly.Variables(25).ValuesAsNumpy(),

        "soil_moisture_0_to_7cm": hourly.Variables(26).ValuesAsNumpy(),

        "soil_moisture_7_to_28cm": hourly.Variables(27).ValuesAsNumpy(),

        "soil_moisture_28_to_100cm": hourly.Variables(28).ValuesAsNumpy(),

        "soil_moisture_100_to_255cm": hourly.Variables(29).ValuesAsNumpy()

    }

    hourly_dataframe = pd.DataFrame(data=hourly_data)

    return hourly_dataframe


df = get_weather_data()

df.to_csv('weatherData.csv')
```