Assignment 2 - Report
DATA0006
August 2023

Jigar Shah, Hillar Parbo, Anastasiia Sulakova, Chloe Gerak

Overview: This report investigates the relationships between weather data, price data and maximum daily energy demand across the state of Victoria for the time period between November 1st 2022 and April 24th 2023, identifies factors that have the most predictive use and creates a machine learning model to accurately predict the energy demand to aid energy companies in planning for future usage.

Word Count: 2499

1 Data Visualisation

1.1 Datasets

1.1.1 BOM Weather Dataset

The weather data used in this report is originally from the Bureau of Meteorology (BOM) and was provided by the University of Melbourne (UoM)[1]. The data contains 175 daily observations with 22 features from a central Melbourne location, with features representing various temperature, wind, pressure, rainfall, and humidity measurements/determinations. The original dataset is available as monthly CSV files that each contain approximately 30-31 rows and 21 features. Given the weather data is automatically generated and handled, it's possible that there are erroneous, missing or imputed values[2].

1.1.2 AEMO Historical Aggregated Price and Demand Dataset

A subset of the historical aggregated price and demand data from the Australian Energy Market Operator (AEMO) and was provided by the UoM[3]. The subset contains 8,352 rows and 5 features, with data displaying ~50,000 rows and 5 features for the time period, and observations are shown in 5-minute time increments. UoM selected observations at 30-minute increments, therefore it's likely the subset does not fully represent the original dataset as no aggregative methods were performed to reduce the data from 5-minute increments to 30-minute time increments.

All data is for the Victorian region, with the features representing Date/Time, Total Demand (in Megawatt Hours) over the period of time for a region, Regional Reference Price and Period Type[4]. The original dataset is available as monthly CSV files.

1.2 Data Cleaning

Visual inspection of both datasets showed an overall high quality of data with very few missing, inconsistent or possibly erroneous values. Therefore, Pandas was used over alternatives like OpenRefine to perform preprocessing operations, allowing for efficient data cleaning, while keeping the source data files in their original form.

1.2.1 BOM Weather Dataset

The following data cleaning processes were applied to the dataset using Pandas:

- Whitespace in column headings was removed for consistency and ease of cleaning operations.
- Columns 'Evaporation (mm)', 'Sunshine (hours)', '9am cloud amount (oktas)' and '3pm cloud amount (oktas)' were empty thus removed.
- Row 174 representing 24th April 2023 had missing 9 values, and given that the price and demand dataset only had one observation for that particular date (opposed to 48 for other dates), it was decided to entirely remove row 174. This helped maintain a high quality of data and avoided imputation, which had the potential to dilute the statistical accuracy of any subsequent analysis.
- 'Location' column entirely comprised of identical values, therefore removed.
- For the '9am wind speed (km/h)' column, each instance of the value 'Calm' was imputed with an integer, 0, as per the Beaufort Wind scale[5].

- Converted the '9am wind speed (km/h)' and '9am relative humidity (%)' columns into decimal-based floats for consistency with other columns, and to enable certain calculations otherwise not possible with integer-based 0 values.
- Set the index to "Date" and converted values to datetime format to perform datetime operations.

Table 1. Final Weather Dataset - Snapshot

Date	Minimum temperature (°C)	Maximum temperature (°C)	Rainfall (mm)	Direction of maximum wind gust	Speed of maximum wind gust (km/h)	Time of maximum wind gust	9am Temperature (°C)	9am relative humidity (%)	9am wind direction	9am wind speed (km/h)	9am MSL pressure (hPa)	3pm Temperature (°C)	3pm relative humidity (%)
01/11/2022	8.5	13.3	3.0	SW	44.0	11:36	12.0	64.0	NW	13.0	991.3	13.2	59.0
02/11/2022	6.9	15.7	2.8	SSW	43.0	11:33	11.0	65.0	W	9.0	1006.7	15.1	54.0
03/11/2022	9.1	15.4	0.4	SSW	31.0	8:07	11.5	70.0	SSW	9.0	1019.2	13.8	67.0
04/11/2022	10.2	17.8	0.2	S	24.0	14:50	12.3	84.0	WSW	6.0	1028.1	16.9	56.0
05/11/2022	11.8	22.7	0.0	N	31.0	11:58	14.0	78.0	N	9.0	1026.3	18.8	65.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
19/04/2023	14.4	17.5	1.2	SW	28.0	11:32	14.8	66.0	WSW	9.0	1022.0	15.8	53.0
20/04/2023	7.8	18.5	0.0	SSW	17.0	12:47	12.4	72.0	NNE	7.0	1026.8	17.5	58.0
21/04/2023	11.3	19.0	0.0	SSW	13.0	10:48	14.8	78.0	NE	6.0	1030.5	18.6	64.0
22/04/2023	14.6	19.2	0.0	SSW	22.0	16:09	15.2	84.0	SW	6.0	1031.9	18.8	64.0
23/04/2023	14.8	19.0	0.0	SSW	26.0	15:45	16.7	79.0	SSE	7.0	1034.4	17.8	71.0

Table 2. Weather Data Cleaning Examples

```
Date                                  object
Minimum temperature (°C)             float64
Maximum temperature (°C)             float64
Rainfall (mm)                        float64
Direction of maximum wind gust        object
Speed of maximum wind gust (km/h)    float64
Time of maximum wind gust             object
9am Temperature (°C)                 float64
9am relative humidity (%)              int64
9am wind direction                    object
9am wind speed (km/h)                 object
9am MSL pressure (hPa)               float64
3pm Temperature (°C)                 float64
3pm relative humidity (%)            float64
3pm wind direction                    object
3pm wind speed (km/h)                float64
3pm MSL pressure (hPa)               float64
dtype: object
```

```
Date                                  object
Minimum temperature (°C)             float64
Maximum temperature (°C)             float64
Rainfall (mm)                        float64
Direction of maximum wind gust        object
Speed of maximum wind gust (km/h)    float64
Time of maximum wind gust             object
9am Temperature (°C)                 float64
9am relative humidity (%)            float64
9am wind direction                    object
9am wind speed (km/h)                float64
9am MSL pressure (hPa)               float64
3pm Temperature (°C)                 float64
3pm relative humidity (%)            float64
3pm wind direction                    object
3pm wind speed (km/h)                float64
3pm MSL pressure (hPa)               float64
dtype: object
```

Original Format

Date
1/11/2022
2/11/2022
3/11/2022
4/11/2022

Datetime format

Date
01/11/2022
02/11/2022
03/11/2022
04/11/2022

1.2.2 AEMO Price and Demand Dataset

The following data cleaning processes were applied using Pandas:

- Created headers based on the original AEMO dataset for readability[6].
- 'Region' and 'Period_Type' columns were constants, therefore removed.
- Split 'Date & Time' column into separate 'Date' and 'Time' columns to allow for a daily aggregate to be calculated, subsequently removed the original 'Date & Time' column to avoid duplicate data.
- Set 'Date' column as the index and set the index heading to 'Date' for consistency with weather dataset.
- There was only 1 row representing the 24th of April 2023 (opposed to ~48 for other dates), so it was removed to allow for consistency with the weather dataset, and maintain data integrity.
- Inserted the missing values for the date and time 1/11/2022 0:00 based on the original AEMO dataset to avoid imputation, which could have diluted the statistical accuracy of subsequent analysis[6].
- Converted 'Date' values to datetime format for consistency with weather data and to allow datetime operations to take place.
- Aggregated the values of 'Total Demand' and 'Price' columns using summing to get daily totals of each column, assigned the outputs to two new columns, 'Max Demand' and 'Max Price', and removed the original 'Total Demand' and 'Price' columns to avoid duplicate data.
- Aggregated the data into daily observations by dropping duplicate rows, allowing for efficient merging of datasets and ease of subsequent analysis.

Table 3. Energy Data Cleaning Examples

	Total Demand	Price	Time
01/11/2022	4455.59	44.58	0:00
01/11/2022	4178.18	8.94	0:30
01/11/2022	4086.02	0.14	1:00
01/11/2022	4033.37	0.02	1:30
01/11/2022	3985.64	0.00	2:00

	Total Demand	Price	Time	Max Demand	Max Price
01/11/2022	4455.59	44.58	0:00	211885.05	-132.26
01/11/2022	4178.18	8.94	0:30	211885.05	-132.26
01/11/2022	4086.02	0.14	1:00	211885.05	-132.26
01/11/2022	4033.37	0.02	1:30	211885.05	-132.26
01/11/2022	3985.64	0.00	2:00	211885.05	-132.26

Table 4. Final AEMO Price

	Max Demand	Max Price
Date		
01/11/2022	211885.05	-132.26
02/11/2022	233630.93	-120.73
03/11/2022	235672.34	3034.04
04/11/2022	224895.28	5325.08
05/11/2022	189190.35	3502.20
...	...	...
19/04/2023	220897.61	4912.03
20/04/2023	227925.10	6273.96
21/04/2023	232043.13	6310.13
22/04/2023	207431.82	3858.79
23/04/2023	195560.35	3034.26

1.2.3 Merged Final Dataset

After data cleaning, both datasets were merged based on their common index using Pandas, as this required no additional pre or post processing. A high quality of data was maintained, and the merge facilitated a thorough data analysis process that led to insights regarding variable relationships. All columns relating to wind' were excluded due to inadequate domain knowledge for proper classification. The resulting DataFrame contains 174 daily observations, 14 features, encompassing essential weather data, Max Price, and the predictive target, Max Demand.

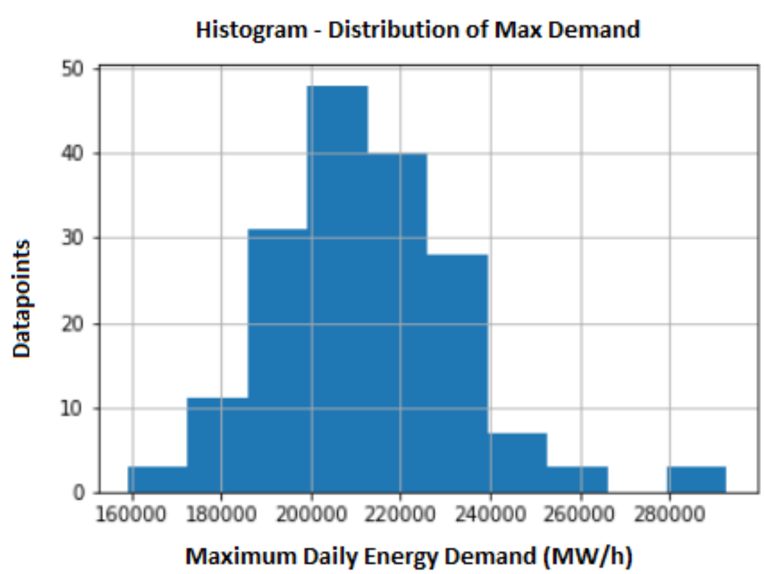
Table 5. Merged Final Dataset - Snapshot

Date	Minimum temperature (°C)	Maximum temperature (°C)	Rainfall (mm)	Speed of maximum wind gust (km/h)	9am Temperature (°C)	9am relative humidity (%)	9am wind speed (km/h)	9am MSL pressure (hPa)	3pm Temperature (°C)	3pm relative humidity (%)	3pm wind speed (km/h)	3pm MSL pressure (hPa)	Max Demand	Max Price
01/11/2022	8.5	13.3	3.0	44.0	12.0	64.0	13.0	991.3	13.2	59.0	11.0	991.5	211885.05	-132.26
02/11/2022	6.9	15.7	2.8	43.0	11.0	65.0	9.0	1006.7	15.1	54.0	13.0	1008.4	233630.93	-120.73
03/11/2022	9.1	15.4	0.4	31.0	11.5	70.0	9.0	1019.2	13.8	67.0	9.0	1021.3	235672.34	3034.04
04/11/2022	10.2	17.8	0.2	24.0	12.3	84.0	6.0	1028.1	16.9	56.0	9.0	1026.6	224895.28	5325.08
05/11/2022	11.8	22.7	0.0	31.0	14.0	78.0	9.0	1026.3	18.8	65.0	11.0	1023.2	189190.35	3502.20
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
19/04/2023	14.4	17.5	1.2	28.0	14.8	66.0	9.0	1022.0	15.8	53.0	9.0	1022.1	220897.61	4912.03
20/04/2023	7.8	18.5	0.0	17.0	12.4	72.0	7.0	1026.8	17.5	58.0	4.0	1025.5	227925.10	6273.96
21/04/2023	11.3	19.0	0.0	13.0	14.8	78.0	6.0	1030.5	18.6	64.0	2.0	1029.2	232043.13	6310.13
22/04/2023	14.6	19.2	0.0	22.0	15.2	84.0	6.0	1031.9	18.8	64.0	11.0	1029.7	207431.82	3858.79
23/04/2023	14.8	19.0	0.0	26.0	16.7	79.0	7.0	1034.4	17.8	71.0	13.0	1032.4	195560.35	3034.26

1.3 Data Exploration

1.3.1 General Descriptive Statistics

Figure 1 - Histogram - Distribution of Max Demand



The frequency distribution of Max Demand (Figure 1) exhibits characteristics of normality, albeit slightly skewed towards the higher end. Although the high values clustered around 280,000 could signify statistical outliers, no outlier processing has been applied as energy market operators will still need to fulfill those demands.

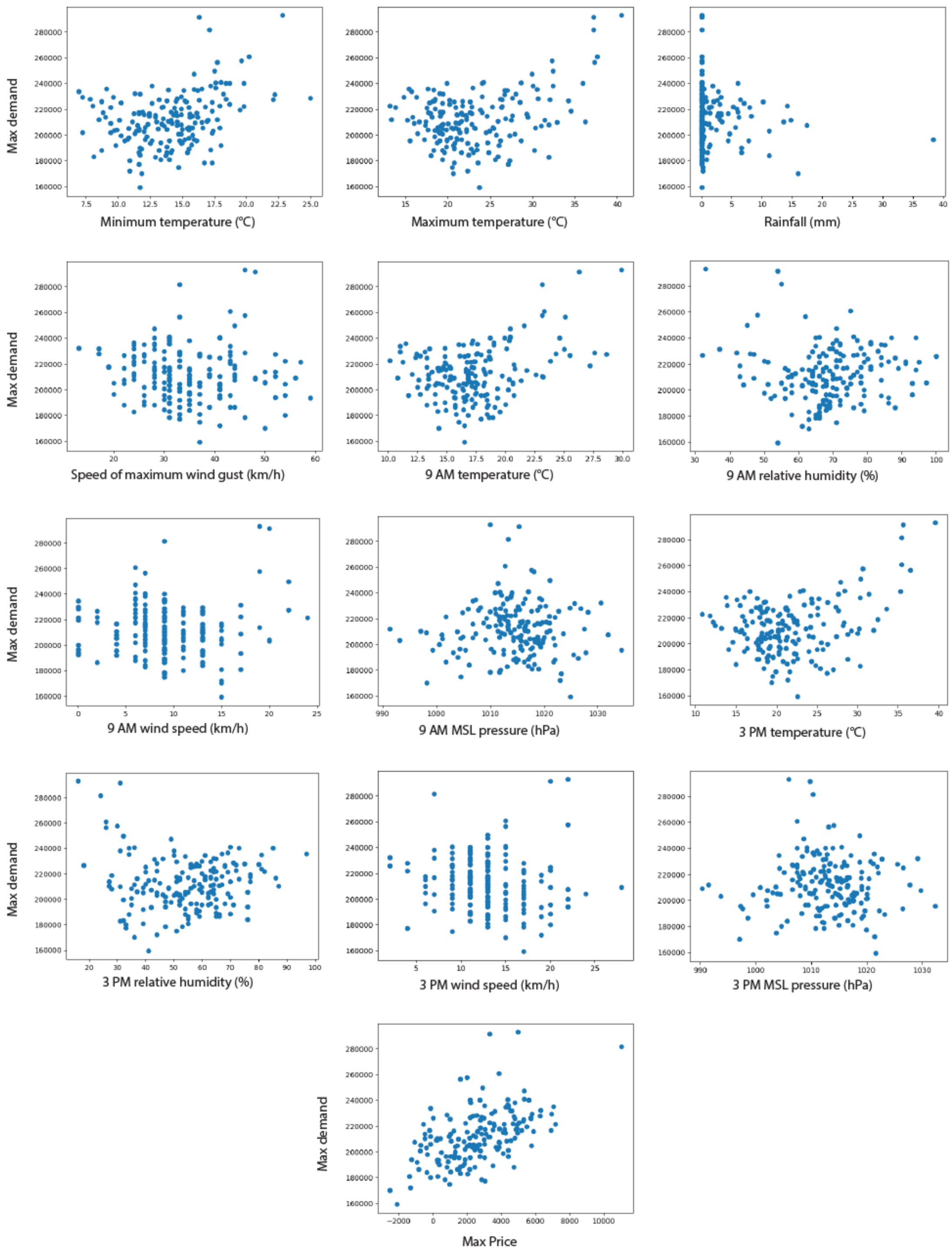
Table 7. Summary Statistics - Snapshot

Summary Statistics

	Minimum temperature (°C)	Maximum temperature (°C)	Rainfall (mm)	Speed of maximum wind gust (km/h)	9am Temperature (°C)	3pm Temperature (°C)	Max Price	Max Demand
count	174.000000	174.000000	174.000000	174.000000	174.00000	174.000000	174.000000	174.000000
mean	14.072414	23.121839	1.786207	34.563218	17.30000	21.546552	2622.096207	212200.532816
std	3.228188	5.419904	4.318081	9.083048	3.55235	5.255205	2120.692383	20891.472357
min	6.900000	13.100000	0.000000	13.000000	10.10000	10.800000	-2518.380000	159249.920000
25%	11.700000	19.300000	0.000000	28.000000	14.90000	18.125000	1102.460000	198263.087500
50%	14.150000	21.700000	0.000000	33.000000	16.90000	20.400000	2588.180000	210796.295000
75%	16.075000	26.275000	1.400000	41.000000	18.70000	24.275000	4233.325000	225354.227500
max	25.000000	40.500000	38.400000	59.000000	29.90000	39.600000	11033.550000	292888.140000

For consistency and ease, summary statistics for numerical columns were calculated with Pandas (Table 4) and provided general insights into the data. Based on basic domain knowledge, the weather data appeared as expected.

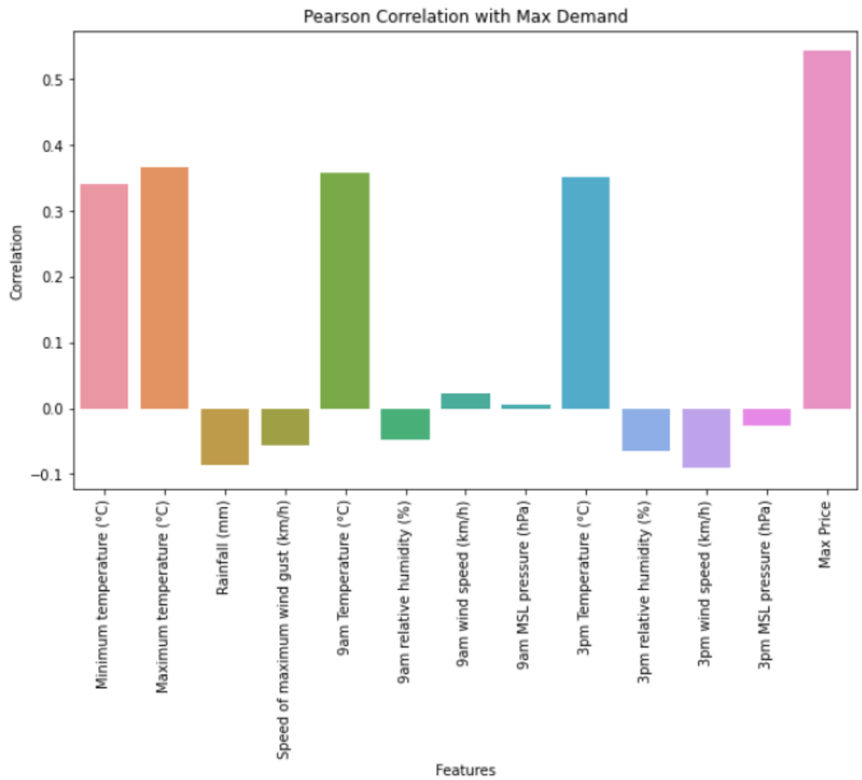
Figure 2 - Scatterplots of Numerical Variables vs Max Demand



Scatterplots of numerical variables vs Max Demand (Figure 2) revealed that several variables potentially had weak-moderately strong relationships with Max Demand, while some appear to have no relationship.

1.3.2 Pearson Correlation

Figure 3 - Pearson Correlation with Max Demand



To quantify the potential linear relationships between the variables and Max Demand, Pearson correlation was computed and visualised for each variable using Pandas and Seaborn (Figure 3). Max Price has a moderate to strong positive linear correlation with Max Demand with various temperature values displaying weak linear correlations, which implies it shifts with these variables, but importantly doesn't imply causation.

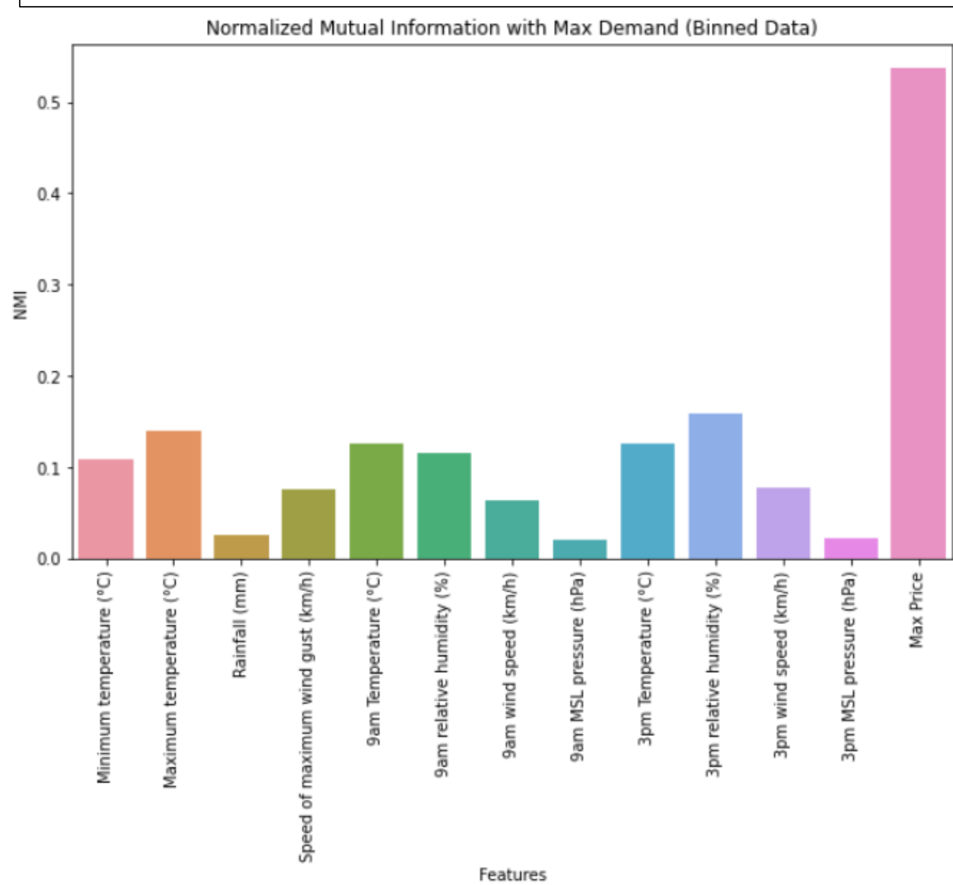
1.3.3 Normalised Mutual Information (NMI)

To analyse data for non-linear relationship of mutual information, Pandas was used to discretise the data into 'bins' based on domain knowledge. The .cut method was used to bin the continuous data into discrete intervals, and was chosen over other methods as it allowed for more flexibility in choosing specific bin parameters.

Figure 4 - Discretised Final Dataset - Snapshot

Date	Minimum temperature (°C)	Maximum temperature (°C)	Rainfall (mm)	Speed of maximum wind gust (km/h)	9am Temperature (°C)	9am relative humidity (%)	9am wind speed (km/h)	9am MSL pressure (hPa)	3pm Temperature (°C)	3pm relative humidity (%)	3pm wind speed (km/h)	3pm MSL pressure (hPa)	Max Price	Max Demand
01/01/2023	(15.0, 20.0]	[35, 40)	(-0.001, 10.0]	(30.0, 40.0]	(20.0, 25.0]	(60.0, 70.0]	(-0.001, 10.0]	(1000.0, 1050.0]	(35.0, 40.0]	(30.0, 40.0]	(10.0, 20.0]	(1000.0, 1050.0]	(2740.0, 2750.0]	(240000.0, 250000.0]
01/02/2023	(10.0, 15.0]	[20, 25)	(-0.001, 10.0]	(30.0, 40.0]	(15.0, 20.0]	(60.0, 70.0]	(-0.001, 10.0]	(1000.0, 1050.0]	(15.0, 20.0]	(60.0, 70.0]	(10.0, 20.0]	(1000.0, 1050.0]	(2600.0, 2610.0]	(210000.0, 220000.0]
01/03/2023	(10.0, 15.0]	[20, 25)	(-0.001, 10.0]	(20.0, 30.0]	(10.0, 15.0]	(90.0, 100.0]	(-0.001, 10.0]	(1000.0, 1050.0]	(20.0, 25.0]	(60.0, 70.0]	(-0.001, 10.0]	(1000.0, 1050.0]	(2010.0, 2020.0]	(210000.0, 220000.0]
01/04/2023	(10.0, 15.0]	[15, 20)	(-0.001, 10.0]	(30.0, 40.0]	(10.0, 15.0]	(70.0, 80.0]	(-0.001, 10.0]	(1000.0, 1050.0]	(10.0, 15.0]	(60.0, 70.0]	(10.0, 20.0]	(1000.0, 1050.0]	(4420.0, 4430.0]	(200000.0, 210000.0]
01/11/2022	(5.0, 10.0]	[10, 15)	(-0.001, 10.0]	(40.0, 50.0]	(10.0, 15.0]	(60.0, 70.0]	(10.0, 20.0]	(950.0, 1000.0]	(10.0, 15.0]	(50.0, 60.0]	(10.0, 20.0]	(950.0, 1000.0]	(-140.0, -130.0]	(210000.0, 220000.0]
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
30/11/2022	(10.0, 15.0]	[15, 20)	(-0.001, 10.0]	(30.0, 40.0]	(15.0, 20.0]	(50.0, 60.0]	(-0.001, 10.0]	(1000.0, 1050.0]	(15.0, 20.0]	(50.0, 60.0]	(-0.001, 10.0]	(1000.0, 1050.0]	(3680.0, 3690.0]	(200000.0, 210000.0]
30/12/2022	(10.0, 15.0]	[25, 30)	(-0.001, 10.0]	(20.0, 30.0]	(15.0, 20.0]	(60.0, 70.0]	(-0.001, 10.0]	(1000.0, 1050.0]	(20.0, 25.0]	(50.0, 60.0]	(10.0, 20.0]	(1000.0, 1050.0]	(2170.0, 2180.0]	(190000.0, 200000.0]
31/01/2023	(15.0, 20.0]	[20, 25)	(-0.001, 10.0]	(20.0, 30.0]	(15.0, 20.0]	(60.0, 70.0]	(-0.001, 10.0]	(1000.0, 1050.0]	(20.0, 25.0]	(50.0, 60.0]	(10.0, 20.0]	(1000.0, 1050.0]	(2830.0, 2840.0]	(210000.0, 220000.0]
31/03/2023	(10.0, 15.0]	[15, 20)	(-0.001, 10.0]	(30.0, 40.0]	(15.0, 20.0]	(70.0, 80.0]	(-0.001, 10.0]	(1000.0, 1050.0]	(15.0, 20.0]	(60.0, 70.0]	(10.0, 20.0]	(1000.0, 1050.0]	(4930.0, 4940.0]	(210000.0, 220000.0]
31/12/2022	(15.0, 20.0]	[30, 35)	(-0.001, 10.0]	(20.0, 30.0]	(20.0, 25.0]	(70.0, 80.0]	(-0.001, 10.0]	(1000.0, 1050.0]	(25.0, 30.0]	(50.0, 60.0]	(10.0, 20.0]	(1000.0, 1050.0]	(2690.0, 2700.0]	(200000.0, 210000.0]

Figure 5 - Normalized Mutual Information with Max Demand (Binned Data)



Max Price shared a high amount of mutual information with Max Demand, indicating similarities between the clusterings of data, possibly indicating a non-linear relationship. Other variables were only weakly related to Max Demand. An inherent limitation of this approach is its sensitivity to initial data preprocessing (binning), and that NMI doesn't imply causation.

1.4 Conclusions

This report identified that temperature measurements and Max Price are related to Max Demand and may prove to be strong predictors, corroborating previous literature that suggests energy demands increase as temperatures fluctuate beyond certain thresholds[7]. While Data Exploration provides initial insights into individual factors affecting energy demand, future projects are suggested to employ more comprehensive statistical and modeling analyses to delve deeper into the variable relationships.

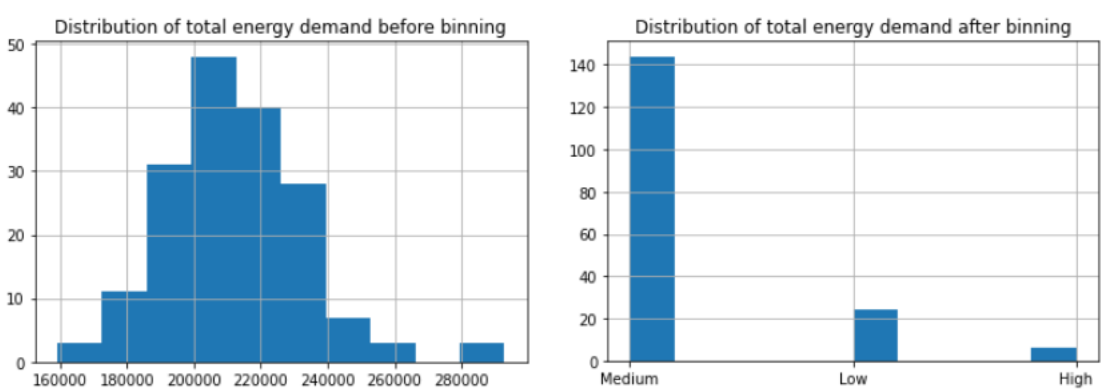
2 Predictive Modeling

This section identifies key predictive features and builds a machine learning model to predict energy demand using weather and price data. Various machine learning models and algorithms will be tested and assessed based on their accuracy.

2.1 Dataset

The dataset from 1.2.3, less the non-numerical columns, was used for the classification models, and Max Demand was re-discretised into three new bins to better capture their initial distribution (Figure 6) using domain knowledge, noting that the resulting uneven distributions may influence the accuracy of the model. As part of feature selection, the two pressure columns that appeared the least related to demand were removed to improve accuracy.

Figure 6 - Histograms - Distribution of Energy Demand before and after binning



2.2 Model Selection

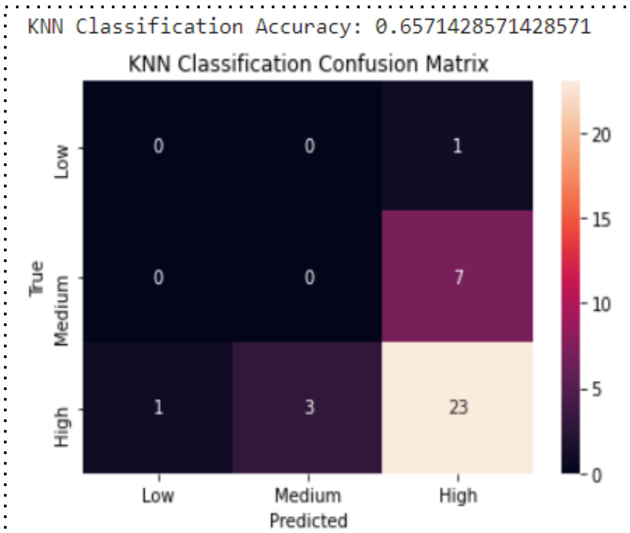
A classification model was explored due to absent strong linear correlations between variables and the target typically required by regression models, suiting the practical constraints of our pragmatic choice to forego outlier processing. Two models, ‘k Nearest Neighbour (kNN)’ and ‘Decision Tree Classification (DTC)’ models were considered.

2.2.1 Classification Model with KNN Algorithm

The algorithm was built using the ‘golden ratio’ of 80/20 train & test split and random_state of 42 for replicability, with scaling being applied to account for the algorithm’s sensitivity to extreme data. An arbitrary k value of 3 was selected, and an accuracy score and confusion matrix was used to evaluate effectiveness.

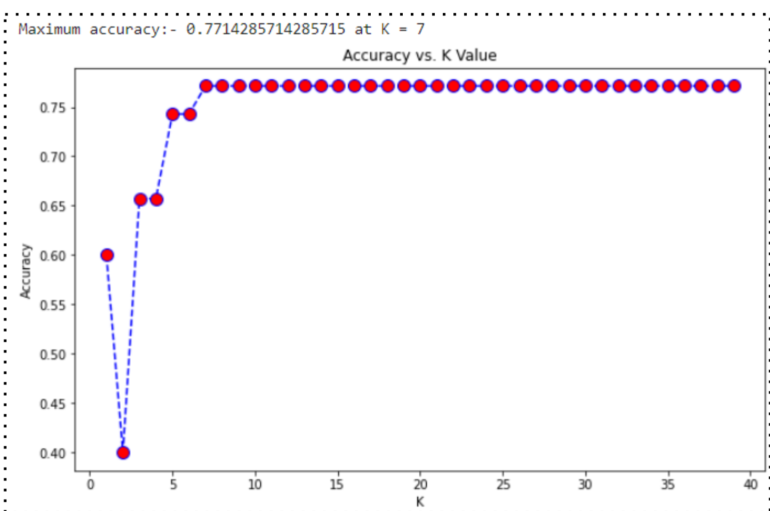
Accuracy score of ~0.66 is a fair result, however, not high enough to use this model for accurate predictions. Based on the confusion matrix, 23 values were predicted correctly as high energy demand, while 7 truly medium values were also classified as high (Figure 7). This result can indicate that the selected k value was not the best as a too small of a k value is largely affected by the noise.

Figure 7 - KNN Accuracy and Confusion Matrix



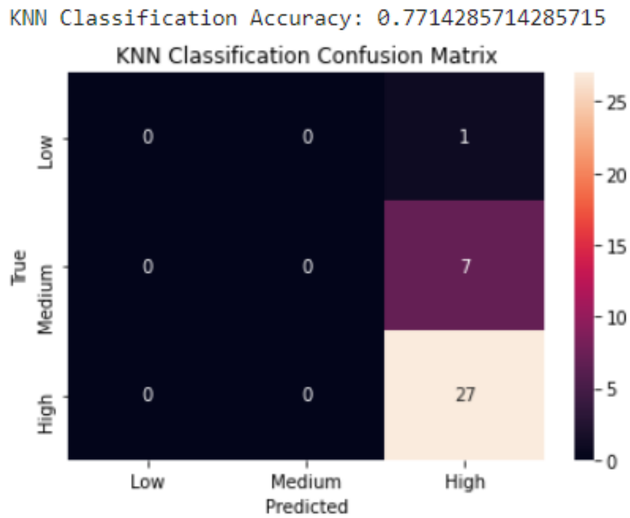
To select an optimal k, we derived a plot between accuracy score and k value which demonstrated that at k=7 and further accuracy is the highest = 0.77 (Figure 8).

Figure 8 - Accuracy vs K Value



With the optimal k = 7 a higher accuracy is observed in predicting 'High' class values, however, the model was still unable to predict 'Low' or 'Medium' values.

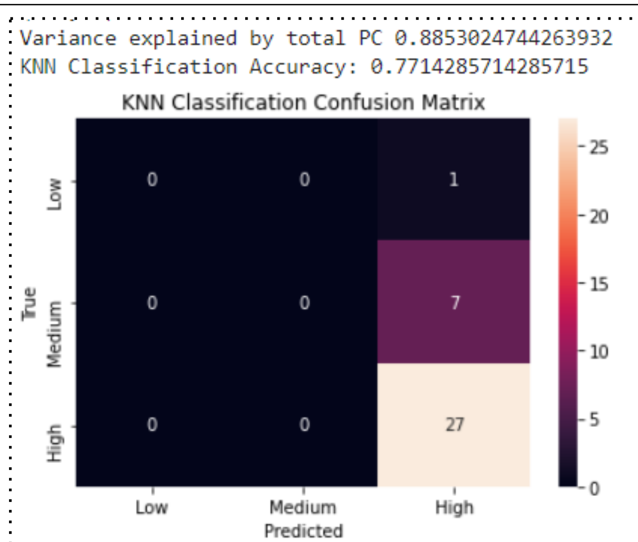
Figure 9 - KNN Accuracy and Confusion Matrix



KNN Model with Principal Component Analysis (PCA)

Scikit-learn was used to calculate PCA for feature engineering in favour of chi-square, due to the complicated nature of defining a threshold for the number of features to drop (as most of them have similar levels of correlation with the target variable and none of them have a clear linear correlation). We converted our current features into 5 new features using PCA, and this number of new features explained the variance of the new dataset by 88.5% indicating most of the original dataset was preserved. It was also observed that the accuracy score is not significantly affected by different combinations of features. Therefore, we decided to keep a number of 5 new features resulting in a final accuracy score of 77.14%.

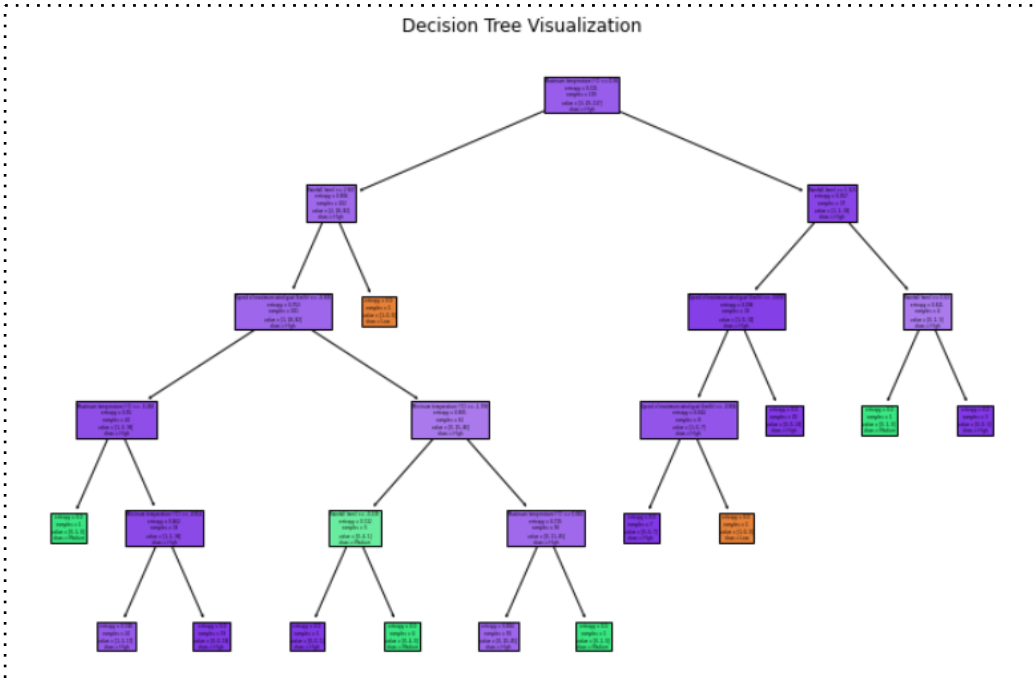
Figure 10 - Variance, KNN Accuracy and Confusion Matrix



2.2.2 Classification Model with DTC Algorithm

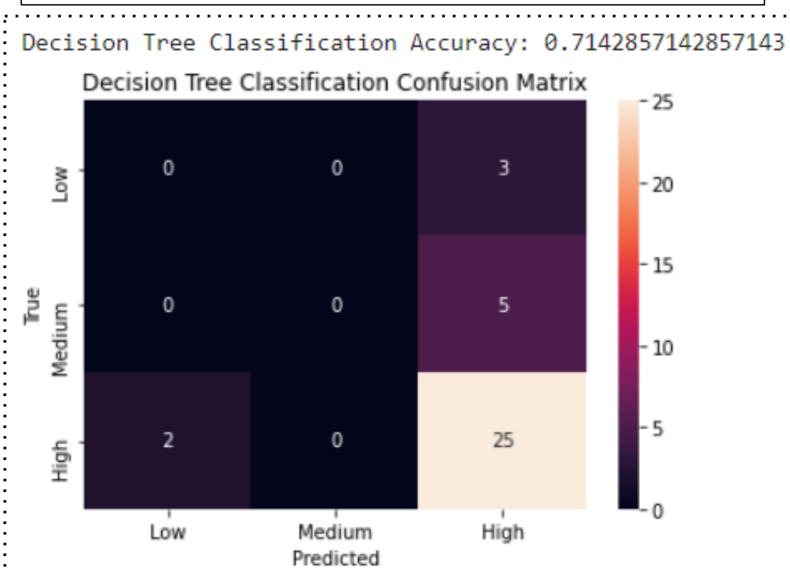
For consistency, a similar train & test split of 80/20, random state = 44, maximum depth of the DT = 5 were used, with scaling not required, as DTC is not a distance-based algorithm.

Figure 11 - Decision Tree Visualisation



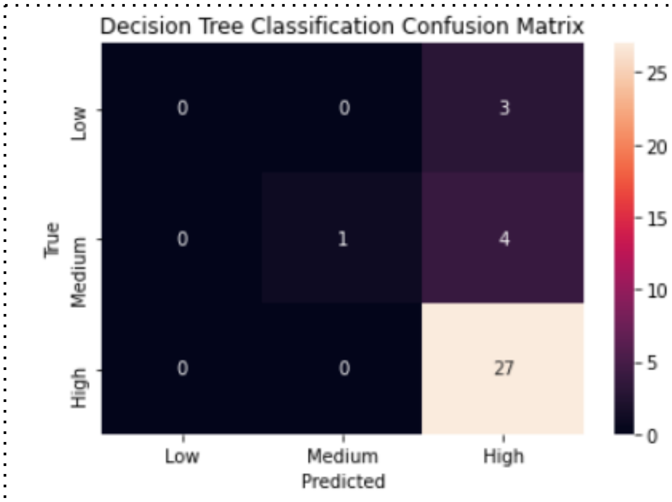
This classification model resulted in a 71% accuracy result with most of the originally high values predicted correctly, however, there was some confusion between Low and High class values.

Figure 12 - Decision Tree Classification Confusion Matrix



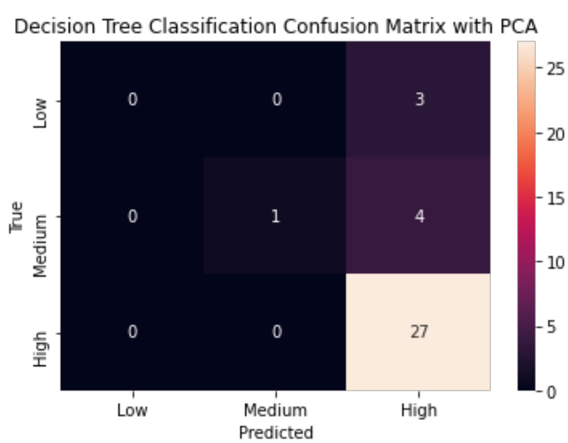
With a view to addressing this issue, PCA feature selection was applied. With 4 'new' features, and variance explained by total PC at 0.81, Decision Tree Classification accuracy increased to 80 %. It can now be observed that at least 1 value was correctly predicted as belonging to the 'Medium' class.

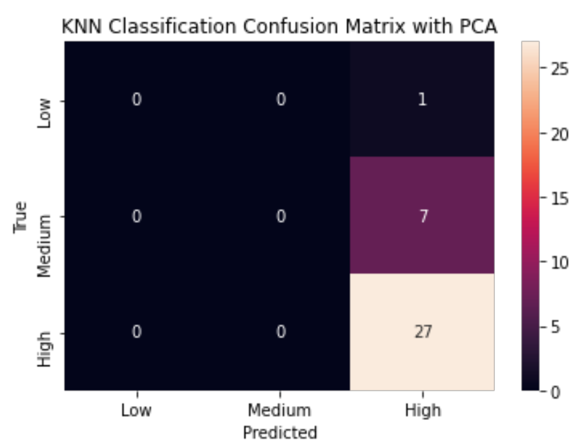
Figure 13 - Decision Tree Classification Confusion Matrix



2.3 Model Performance Metrics and Final Model Choice

- Models were compared on accuracy score, with DTC and kNN demonstrating similar effectiveness with high scores of 0.8 and 0.77 respectively.
- Confusion matrices showed both models tend to better predict values of a specific class - ‘High’, while failing to classify values of ‘Low’ and ‘Medium’ classes. In this regard, the decision tree classification model proved slightly more accurate, being able to predict one ‘Medium’ value after applying the PCA test.





- From the above examples it is clear that these metrics do not take into account the issue of class imbalance. K-fold cross-validation was used to prevent overfitting of the model and demonstrated that mean accuracy of kNN vs DTC was higher - 0.80 vs. 0.77.

Considering the effectiveness and limitations of both algorithms it was decided to use the kNN Classification model for the research. The model demonstrates the high overall accuracy and is easily interpretable.

2.4 Limitations and Conclusions

There are several limitations in the accuracy and reproducibility of using this model to forecast future trends of energy demand using weather patterns in Melbourne. Firstly, the analysis focused on limited months (November to April), and failed to account for potential unforeseen energy demands that occur in every season. General trends have shown a decrease in energy demand in Victoria during winter months[8]. Additionally, the weather data was limited to Greater Melbourne, representing 75.6% of the Victorian population[9], while the Price and Demand data covered the entire State of Victoria. Differences in energy needs stemming from diverse habitats[10] and agriculture[11] could exist. Addressing these limitations requires a year-round, region-specific dataset to enhance future models.

Climate patterns (La Niña and El Niño) and climate change influence weather in Melbourne, with La Niña resulting in fewer daily heat extremes but an increased frequency of long warm spells or heatwaves[12], while El Niño results in individual daily heat extremes to be hotter[13]. The data analysed was during a La Niña pattern[14] and the analysis may therefore be biased towards La Niña trends. El Niño is currently becoming predominant thus trends will change as consumers increase energy usage to maintain comfortable indoor temperatures[15]. Also, climate change has steadily increased the temperature in Victoria by ~0.8°C since the 1950s[16], coinciding with more severe and longer heatwaves, potentially placing more demand on energy that may eventually exceed power systems' capacity[15].

Nonetheless, energy demand and price from a state level may decrease as government proposals, such as the State Electricity Commission and initiatives for electrical energy over gas[17-18], look to promote renewable energy such as solar power, and decrease reliance on gas, potentially further alleviating statewide energy demands. Solar power units are increasing each year[19], and as the state adopts more renewable sources (wind and solar), they prove cost-effective compared to coal ($60-100/MWh vs. $130-250/MWh)[20]. While coal remains Victoria's primary energy source, it aims for 40% renewable energy by 2025 and 95% by 2035[21]. Thus, even if there is an increase in energy demand due to warmer climate, price may not proportionally increase due to more efficient energy production.

Global events also cause variability to predictability, potentially changing price and demand in surprising ways. Two geopolitical instances impacting energy demand are the Russia-Ukraine War and the COVID-19 pandemic. The model used post-event data, suggesting their potential influence. The Russia-Ukraine War, starting in February 2022, surged global energy prices by 20% in 5 months, driven by crude oil demand[22]. This unprecedented event could not be accurately predicted, and future instabilities may have similar effects. The impact of COVID-19 is still being understood, with conflicting data regarding its influence on energy demand and pricing. Initially, Australia experienced reduced energy price inflation due to COVID-19, which later transitioned to a period of higher inflation, encompassed within the analyzed dataset.[23-24]. Inflation directly impacts energy prices, potentially causing increases even without demand growth. The pandemic fuelled a drive for working from home arrangements due to lockdowns. One study suggested that the initial lockdowns, controlled for weather variations, showed an overall 1% energy demand[25]. However, looking at larger intervals (2020-2021) demonstrated a decrease in energy consumption due to the COVID-19 pandemic[26]. This dataset fails to acknowledge climate variations, and the pandemic coincided with a La Niña weather event[14], which could be a contributing factor to the decreased demand.

There are limitations to this model that could be incorporated through a larger dataset over various climate patterns and seasons. Global temperatures are expected to rise in the coming years, until potential effects from renewable energy start to make a difference. Worldwide events will remain a potential confounding factor, as they cannot be accurately predicted.

Overall, our data model provides insight into the relationship between weather and energy demand, and is able to accurately predict energy demand with 80% accuracy for the time period of November to April, with temperature and energy price being the strongest predictors of energy demand. The current model is able to effectively aid energy companies in understanding and planning for future usage.

References

1. Australian Bureau of Meteorology. (n.d.). Climate Data. http://www.bom.gov.au/climate/data/
2. Australian Bureau of Meteorology. (n.d.). Daily Weather Observations. http://www.bom.gov.au/climate/dwo/IDCJDW0000.shtml
3. Australian Energy Market Operator (AEMO). (n.d.). Aggregated Data - National Electricity Market (NEM). https://aemo.com.au/energy-systems/electricity/national-electricity-market-nem/data-nem/aggregated-data
4. Australian Energy Market Operator (AEMO). (2023). MMS Data Model Upgrade Report. https://visualisations.aemo.com.au/aemo/di-help/Content/Data_Model/MMS_Data_Model_Upgrade_Report_52.pdf
5. Australian Bureau of Meteorology. (n.d.). Wind. Marine Knowledge Centre - Reference. http://www.bom.gov.au/marine/knowledge-centre/reference/wind.shtml
6. AEMO. Aggregated price and demand data (n.d.). http://surl.li/jonpt
7. Guo, L., & Wallis, P. (2021). Heatwaves and Crime in Brisbane, Australia. Frontiers in Sociology, 6, 644789. https://doi.org/10.3389/frsc.2021.644789
8. Australian Energy Regulator. (n.d.). Seasonal Peak Demand Regions - Wholesale Statistics. https://www.aer.gov.au/wholesale-markets/wholesale-statistics/seasonal-peak-demand-regions
9. Australian Bureau of Statistics. (2021). 2021 Census Data Downloads. https://www.abs.gov.au/statistics/people/population/population-census/2021#data-downloads
10. Britannica. (n.d.). Victoria: Climate. https://www.britannica.com/place/Victoria-state-Australia/Climate
11. Australian Government Department of Industry, Science, Energy, and Resources. (n.d.). Agriculture - Industry Sector Guide. https://www.energy.gov.au/business/industry-sector-guides/agriculture
12. Australian Bureau of Meteorology. (2016). Climate Update - Record High Sea Surface Temperatures. http://www.bom.gov.au/climate/updates/articles/a020.shtml
13. Australian Bureau of Meteorology. (2021). El Niño and Australia. http://www.bom.gov.au/climate/updates/articles/a008-el-nino-and-australia.shtml
14. Australian Bureau of Meteorology. (2023). ENSO Outlook. http://www.bom.gov.au/climate/enso/outlook/
15. Colelli, F. P., Wing, I. S., & Cian, E. D. (2023). Air-conditioning adoption and electricity demand highlight climate change mitigation–adaptation tradeoffs. Scientific Reports, 13(1), 4413.

16. South West Climate Change Council. (2013). Climate Trends in Victoria. https://www.swclimatechange.com.au/cb_pages/climate_trends_vic.php
17. Victorian Government. (2023). State Electricity Commission of Victoria. https://www.vic.gov.au/state-electricity-commission-victoria
18. Premier of Victoria. (2023). New Victorian Homes to Go All-Electric by 2024. https://www.premier.vic.gov.au/new-victorian-homes-go-all-electric-2024
19. Clean Energy Regulator. (2023). Postcode Data for Small-Scale Installations. https://www.cleanenergyregulator.gov.au/RET/Forms-and-resources/Postcode-data-for-small-scale-installations
20. CSIRO. (2023). GenCost - Energy Data Modelling. https://www.csiro.au/en/research/technology-space/energy/Energy-data-modelling/GenCost
21. Energy Victoria. (2023). About the Electricity Sector. https://www.energy.vic.gov.au/about-energy/about-the-electricity-sector
22. Open Access Government. (2023). One Year Impact of Ukraine War on Global Energy Prices: Input-Output Analysis. https://www.openaccessgovernment.org/one-year-impact-of-ukraine-war-global-energy-prices-input-output-analysis/152599/
23. Reserve Bank of Australia. (2020). Inflation and the November 2020 Statement on Monetary Policy. https://www.rba.gov.au/publications/smp/2020/nov/inflation.html
24. Tsiaplias, S., & Wang, J. (2023). The Australian Economy in 2022–23: Inflation and Higher Interest Rates in a Post-COVID-19 World. Australian Economic Review, 56(1), 5-19.
25. Energy Networks Australia. (2020). Commercial Down, Residential Up: COVID-19's Electricity Impact. Energy Insider. https://www.energynetworks.com.au/news/energy-insider/2020-energy-insider/commercial-down-v-residential-up-covid-19s-electricity-impact/
26. Australian Government Department of Industry, Science, Energy, and Resources. (n.d.). Energy Statistics for States and Territories. https://www.energy.gov.au/data/states-and-territories