A black and white logo

Description automatically generated

PROJECT REPORT BY GROUP B

**FORECASTING THE TRANSITION: ANALYSING THE FEASIBILITY FOR RENEWABLE ENERGY ADOPTION IN NSW**

Kiara Zentrich (z5382804)

Mathew Fraser(z5433663)

Santosh Ban (z5441817)

Karunya Sankar Kumar (z5339860)

April 2024

Link to GitHub Repo: <https://github.com/Karunya2655/GroupB_ZZSC9020.git>

SUBMITTED IN FULFILMENT OF THE REQUIREMENTS OF THE CAPSTONE COURSE ZZSC9020

Plagiarism statement

I declare that this thesis is my own work, except where acknowledged, and has not been submitted for academic credit elsewhere.

I acknowledge that the assessor of this thesis may, for the purpose of assessing it:

* Reproduce it and provide a copy to another member of the University; and/or,
* Communicate a copy of it to a plagiarism checking service (which may then retain a copy of it on its database for the future plagiarism checking).

I certify that I have read and understood the University Rules in respect of Student Academic Misconduct and am aware of any potential plagiarism penalties which may apply.

By signing this declaration, I am agreeing to the statements and conditions above.

Signed: Karunya Sankar Kumar Date: 13 April 2024

Signed: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Date: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Signed: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Date: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Signed: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Date: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# Abstract

Electricity serves as a cornerstone of modern civilisation, powering industries, homes, and critical infrastructure. However, the reliance on non-renewable sources of electricity generation poses significant environmental challenges, exacerbating climate change and endangering ecosystems. In response, the transition to renewable energy sources has emerged as a pressing priority worldwide. This project addresses the transition to a fully renewable energy-based electricity supply in New South Wales (NSW), Australia. The Ecological Defenders Office (EDO), a prominent environmental advocacy firm in NSW, seeks evidence-based forecasts to assess the feasibility of achieving this transition within the next decade. Specifically, the project aims to determine when renewable energy generation can meet forecasted electricity demand in the region. The significance of this endeavour cannot be overstated. Transitioning to renewable energy not only mitigates environmental harm but also fosters energy independence and resilience. The methodology for this project involves several key steps. Firstly, historical electricity demand data will undergo preprocessing to clean, detect outliers, and impute missing values. External factors such as weather data will be incorporated to enhance forecasting accuracy. Exploratory data analysis will identify relevant features, trends, and seasonal patterns. Supervised machine learning models, particularly regression and classification, will be employed for forecasting, considering the non-linear nature of the data. The project will utilise Python programming language and libraries such as scikit-learn and matplotlib for implementation. Google Colab will provide a collaborative development environment, ensuring reproducibility and transparency. Lastly, the data used for analysis will comprise historical electricity demand data, weather data, and additional external factors. This data, while comprehensive, may require normalisation and transformation to facilitate accurate modelling. By employing a combination of statistical and machine learning techniques, this project aims to deliver actionable insights into the feasibility of transitioning to renewable energy in NSW. The findings will inform advocacy efforts and policy decisions, driving progress towards a sustainable energy future. By providing the EDO with accurate forecasts, this project empowers them to advocate for policies and investments that accelerate the transition to renewable energy in NSW.

**Keywords:** Time series forecasting; linear regression; renewable energy; long term forecast; holt winter exponential smoothing; SARIMAX; Prophet

# Contents

Abstract 3

Contents 4

1 Introduction 6

2 Literature Review 8

3 Material and Methods 10

3.1 Software 10

3.2 Description of the Data 10

3.3 Pre-Processing Steps 11

3.4 Data Cleaning 11

3.5 Assumptions 11

3.6 Modelling Methods 11

3.7 Model Evaluation 11

4 Exploratory Data Analysis 13

4.1 Univariate Analysis – Electricity Demand 13

4.2 Bivariate Analysis – Electricity Demand and Temperature 18

4.3 Timeseries Analysis 19

5 Analysis and Results 21

5.1 Preliminary Linear Model – Forecasting Temperature 22

5.2 Linear Model to Forecast Electricity Demand 23

5.3 Other Models 25

6 Discussion 28

6.1 Predicted Demand in 2033 28

6.2 Renewable Energy Generation in 2033 29

6.3 Can RE Projection in 2033 meet NSW demand in 2033? 31

7 Conclusion 32

8 Further Work 33

References 34

Appendix 37

Detailed Data Source information 37

**List of Figures**

Figure 1: Total Demand Box Plot 14

Figure 2: Total Demand by Month (2010-2023) 15

Figure 3: Daily and Yearly Total Demand Means 15

Figure 4: Average Monthly Demand 16

Figure 5: Total Demand for 2022 16

Figure 6: Total and Mean Demand for 2022 by Quarter 17

Figure 7: Mean Total Demand by Weekday 18

Figure 8: Hourly Demand for a Four Week Period 18

Figure 9: Dickey-Fuller Test Results for Total Demand Timeseries 19

Figure 10: Decomposition of Total Demand Timeseries 19

Figure 11: PACF Plot for Total Demand 20

Figure 12: ACF Plot for Total Demand 21

Figure 13: Temperature Predictions for 2024-2033 23

Figure 14: Actual vs Predicted Values for Total Demand as Influenced by Temperature & Population 23

Figure 15: Linear Regression Predictions for Total Demand for 2024-2033 25

Figure 16: Predicted Electricity Demand for 2024-2033 26

Figure 17: Predicted Total Demand from Prophet Model 28

Figure 18: Average Demand Prediction from Prophet Model 29

# 1 Introduction

Electricity is a versatile and widely used source of power that has transformed industries, communication, and daily life, contributing to global progress, and shaping the way we live and work. Produced through various means, including fossil fuels, nuclear reactions, and renewable sources, electricity is transmitted through power grids to homes, businesses, and industries.

However, it is now globally recognised that non-renewable sources of electricity generation, particularly fossil fuels (namely coal, natural gas, and oil) have significant and sustained deleterious effects on the environment [1]

Continued use of these sources would further exacerbate climate change, jeopardising ecosystems, threatening the survival of species, and undermining the health and well-being of present and future generations [2]. As a result, renewable sources, such as wind, solar and hydro, have emerged as clean alternatives to generate electricity whilst reducing dependency on fossil fuels and greenhouse gas emissions.

Transforming the electricity network into a 100% renewable energy scenario is a crucial, but challenging priority in New South Wales (NSW). Achieving this target as soon as possible will result in a greater reduction of potential climate impacts within the region. However, a key concern is whether renewable sources can reliably provide enough electricity to meet NSW demand. The aim of this project is to provide a key environmental advocacy firm in NSW, The Ecological Defenders Office, with evidence-based forecasts to answer the following question: **Is the transition to a fully renewable energy-based electricity supply achievable within NSW in the next 10 years?**

The output from this analysis and modelling will empower the Ecological Defenders Office (EDO) to make informed advocacy and funding decisions. NSW currently has close to 13,500 megawatts of renewable energy generation capacity, which is 53% of total generation capacity in the state [3]. The EDO recognise this as strong progress, however achieving a fully clean energy state remains way off, with the current capacity representing just over halfway toward that goal. The EDO has commissioned this work to understand the earliest point in time at which achieving a 100% renewable electricity target is feasible. By determining this timeline, they aim to devise effective lobbying strategies for governments and electricity providers. If forecasts indicate that renewables can meet demand earlier in the 10-year period, the organisation can prioritize shorter-term, high-intensity campaigns. However, if the forecasts suggest feasibility closer to the 10-year mark, longer-term advocacy requiring extensive planning and organisation will be necessary [4]

This project aims to assess the viability of transitioning to a fully renewable electricity supply in NSW within a decade. Feasibility will be determined based on the analysis indicating whether electricity supply from renewable sources can meet predicted demand levels. The project methodology involves the use of data science techniques to

1. Forecast electricity demand in NSW for the next decade (2024-2033).
2. Forecast the supply of electricity solely from Renewable Energy Generation in NSW in 2033. These forecasts are derived from literature.
3. Finally, the discussion will then integrate the results from these models to determine whether RE generation from Step 2. Can match the Demand from Step 1.

# 2 Literature Review

There are many different approaches to forecasting electricity demand. The choice of approach depends on factors such as the availability of data, the forecasting horizon, the level of detail required, and the specific characteristics of the electricity system being modelled. In this paper, the scope is mainly influenced by the forecasting horizon. We want to forecast 10 years ahead from now therefore, that puts a constraint on the model choices we have.

[5] presents a simple and easy-to-understand method for the next decade of energy demand forecasting based on a nonlinear autoregressive (NAR) neural network. From its time series past values, NAR structurally is an optimal predictor for a future variable. This is useful as the we have the NSW demand data from 2016 and that could be used on its own to predict the future demand without depending on any other variables.

[6] introduces a comprehensive Long Short-Term Memory (LSTM) based model for forecasting electricity demand and price in large datasets within smart city contexts. The model operates as a sequence-to-sequence network, utilizing real electricity market data sourced from the Australian Energy Market Operator (AEMO) for validation. Through various simulations on actual data with different configurations, the model demonstrates its capability to generate reliable predictions. Validation results indicate that the proposed model outperforms existing methods, including Support Vector Machine (SVM), Regression Tree (RT), and Neural Nonlinear Autoregressive network with Exogenous variables (NARX), in terms of Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). Specifically, the proposed model shows improvements of 11.25%, 20%, and 33.5% in RMSE, and 14%, 22.5%, and 32.5% in MAE compared to SVM, RT, and NARX respectively, particularly in forecasting electricity demand. Furthermore, the proposed model demonstrates its ability to provide reliable forecasts even without extensive historical datasets.

In this project, we are dealing with a time series dataset as opposed to a static dataset. Time series is a special type of data set in which one or more variables are measured over time. There are numerous models for time series forecasting. The three main categories are Classical time series models (ARIMA family), Supervised models (Linear Regression, Random Forest and XGBoost), and Deep learning-based models (LSTM, Prophet, DeepAR). [7]

Linear regression serves as a solid baseline model choice, while random forest, XGBoost, and support vector machines (SVMs) excel in managing non-linear relationships and intricate data structures. While classical machine learning (ML) models can be employed for forecasting, they may fall short compared to specialized time-series forecasting techniques like autoregressive models, moving averages, and recurrent neural networks (RNNs). The primary reason for this discrepancy lies in classical ML models' assumption of independent and uniformly distributed data points, a feature often absent in time series data. Time series data is characterized by temporal dependencies, where values at one time point are influenced by those at preceding points. The independence assumption challenges classical ML models' ability to discern underlying patterns and trends. Additionally, classical ML models are ill-equipped to handle time-varying features or unevenly spaced time series data, common occurrences in forecasting tasks. For instance, classical ML models may struggle to capture long-term seasonality trends within the data [8].

There are multiple dependent variables such as Temperature, Population, Climate Change, Impact of AI, Technology, and other potential macroeconomic variables that could influence electricity demand a decade from now. An explanatory model proves advantageous as it integrates information from various variables, rather than solely relying on past values of the variable under consideration (Demand). Nonetheless, there are several reasons why a forecaster might opt for a time series model over an explanatory or hybrid model. Firstly, the system may lack full understanding, and even if it were comprehensible, measuring the assumed relationships governing its behavior might prove exceedingly challenging. Secondly, forecasting the future values of the multiple predictors necessary to forecast the variable of interest might present significant difficulty. Thirdly, the primary objective may solely be predicting future outcomes without delving into the reasons behind them. Lastly, a time series model devoid of explanatory variables may yield more accurate forecasts compared to an explanatory or hybrid model. [9]

# 3 Material and Methods

## 3.1 Software

**Software:**

This section outlines the software tools, libraries, and platforms that will be used for implementing the electricity demand forecasting methodology. It emphasizes the use of open-source software, reproducibility, and accessibility to promote transparency and collaboration.

**Development Environment:** The paper will be implemented using Python programming language, leveraging its rich ecosystem of libraries for data analysis, machine learning and timeseries forecasting.

**Libraries:** The following libraries will be implemented in the code developed for this project

* Pandas and NumPy will be used to manipulate the data and perform any necessary operations to examine or transform the data.
* Scikit-Learn will be used to generate the regression models that will be employed for the forecasting methods.
* The Matplotlib and Seaborn libraries will be used to create any visualisations needed for data analysis or for the explanation and discussion of results.

**Integration and Workflow:** A Jupyter notebook hosted in Google Colab will be used to develop and execute the code. Colab has been chosen as it is cloud based which allows for ease of access to all team members, as well as providing GPU and TPU resources which will allow complex models to be run more efficiently than on any of the team’s local resources. Colab will also enable easy version control as it can be connected to the GitHub repository being used by the team.

## 3.2 Description of the Data

**Data Description**

At this stage of the project, five datasets have been selected for use in the analysis and modelling involved in this forecasting electricity demand research. The details and relevance of each are below:

## 3.3 Pre-Processing Steps

**Data Pre-processing:** The historical electricity demand data provided for this project will be pre-processed to perform data cleaning, outlier detection and missing value imputation. Any relevant external factors such as weather data (provided by the course), public holidays and weekend calendar data will be incorporated into the data set.

## 3.4 Data Cleaning

## 3.5 Assumptions

## 3.6 Modelling Methods

We employed 4 distinct forecasting methods.

1. **Linear Regression with 2 independent variables**: This is our baseline model (needs more text)
2. **Holt Winter Model Exponential Smoothing (HWES):**The Holt-Winters model employs Exponential Smoothing to capture both the trend and seasonality present in the data. By analysing historical data, it discerns patterns and can effectively forecast future electricity demand with precision.
3. **SARIMAX Model:**

The SARIMAX model integrates the AutoRegressive Integrated Moving Average (ARIMA) framework with exogenous variables, enabling it to account for both the inherent temporal patterns in the data and external factors influencing electricity demand. Utilizing methods such as auto\_arima to identify optimal parameters, SARIMAX furnishes forecasts that incorporate seasonal adjustments.

1. **Prophet Model:**

The Prophet model, engineered by Facebook, excels in managing irregularities such as holidays and outliers within time series data. By automatically detecting and integrating seasonality, trends, and holiday effects, it furnishes resilient predictions for electricity demand, rendering it a potent asset for forecasting purposes.

## 3.7 Model Evaluation

**Training and Validation:** Splitting the dataset into training, validation and test sets is necessary. Training the selected models on training data and tuning hyperparameters using the validation set. Then the model performance can be monitored on the validation set to avoid overfitting.

**Evaluation metrics:** Metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) will be calculated to assess the accuracy of the forecasting models. The metrics from different models will then be compared to pick the best performing model.

The following sequence of steps explain the methodology for this research.

**Exploratory Data Analysis and Feature engineering:** Important features such as trends and seasonal indicators will be identified. If required, categorical variables will be transformed into proper formats and continuous features might be normalised. The model selection (e.g. type of regression used) will be done depending on whether the variables are linear or non-linear.

# 4 Exploratory Data Analysis

## 4.1 Univariate Analysis – Electricity Demand

The primary feature in this project is electricity demand. Hence, a substantive exploratory analysis of the NSW electricity demand data is essential to understand underlying patterns, trends and context.

We commenced with a review of the descriptive statistics for the raw electricity demand dataset. This dataset contains half hourly demand measures (in MW) for NSW from 2010 to 2023. Using Python pandas method.describe() and a seaborn boxplot we were able to understand the shape and central tendencies of the raw data.

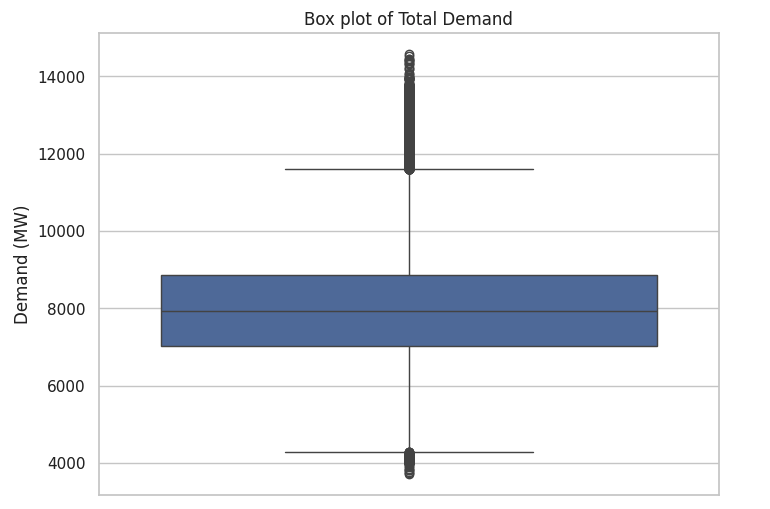
The median value of electricity demand for a 30-minute interval was ~7924MW. However, as indicated by the long whiskers of the boxplot in Figure 1, there is a large range of values in the sample. There are also a significant number of outliers, particularly in the upper range of values.

Figure 1: Total Demand Box Plot

Although descriptive analysis is useful for understanding the spread of the data, to identify patterns of demand we needed to view the data more completely, considering its time series nature.

Figure 2 below visualises daily mean electricity demand and effectively captures the seasonality of this utility. It is evident that there are strong peaks in electricity demand during the summer and winter periods in NSW. However, there are variations in how high demand peaks across the years, for instance mean daily demand was very high in the summer of 2011, but a much lower mean daily demand was observed for the summer of 2015.

The overlaid yearly mean trendline illustrates the fluctuating demand patterns spanning the past decade or so (Figure 3). Demand peaked in 2010, marking its highest point in the past 14 years, only to taper off until 2015. There was a slight uptick in electricity demand from 2015 to 2018, followed by a subsequent decline. By the end of 2023, demand reached its lowest levels for the time period within the scope of this analysis.

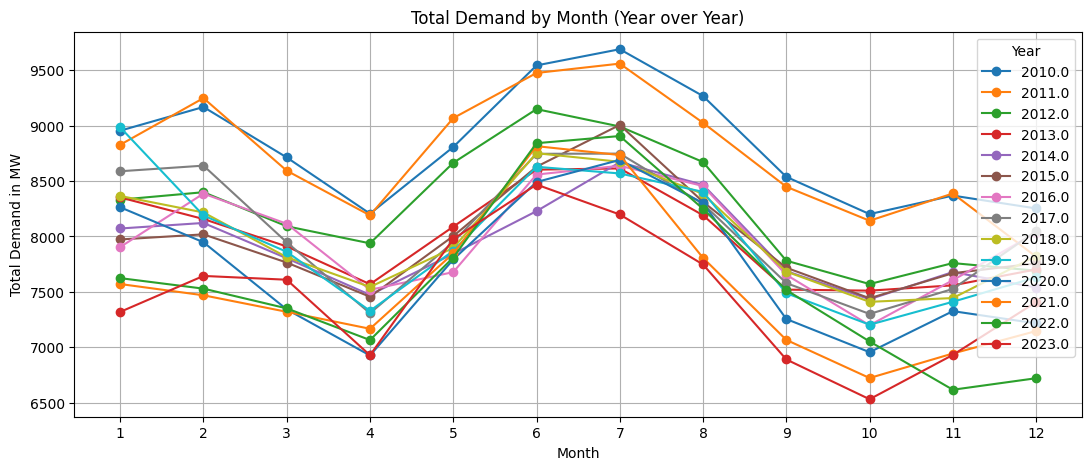


Figure 2: Total Demand by Month (2010-2023)

A graph showing a blue line

Description automatically generated

Figure 3: Daily and Yearly Total Demand Means

Figure 4 below, illustrates a notable trend. Over the past decade, there has been a significant increase in the monthly variation of electricity demand. Notably, there are more pronounced extremes observed during the middle of the year, which typically coincides with the winter season. There are also larger declines in demand during the off-season periods. For instance, the difference between mean daily demand in the highest month and demand in the lowest month of 2022 was 1840MW. Comparatively, the difference between the highest and lowest months of 2013 was 1102MW, a difference of ~50%.

A graph with blue dots and numbers

Description automatically generatedThis pattern suggests a shifting dynamic in energy demand patterns, possibly influenced by factors such as weather fluctuations, economic conditions, adoption of renewables and subsequent changes in heavy industry practices. The Australian Research Council reports that in terms of climate, the region will continue to face more intense extremes in the future, including periods of extreme heat and others of extreme cold [10]. These climate factors may have influenced the more dramatic variations in electricity demand that are being noted in the data across the last couple of years.

Figure 4: Average Monthly Demand

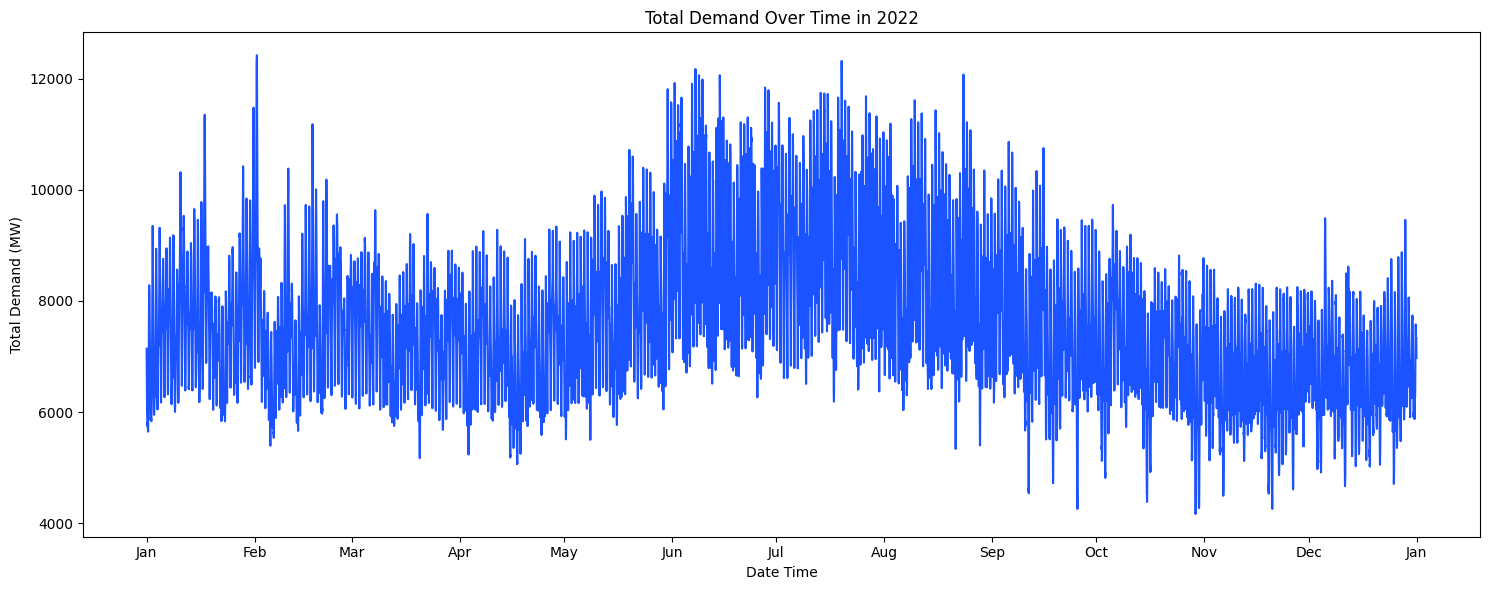
A further examination of seasonal trends in electricity demand across a one-year sample reveals the heightened levels of demand observed in the winter months (June – August). As seen Figure 5 below, although there are some spikes in daily electricity demand in the early, warmer months of the year, demand values are generally contained within the 6000 to 10000MW range. However, from June to the end of August the daily mean demand is mostly within the 7000 to 11000MW range.

Figure 5: Total Demand for 2022

A screenshot of a graph

Description automatically generatedThe quarterly plot shown below in Figure 6 indicates that months at the start and end of the year (Q1 and Q4), generally exhibit a more stable demand pattern, whereas as Q2 shows a trend of increasing demand that builds to a peak at the start of winter in June. Subsequently Q3 shows a trend of declining demand from the highs of the mid-year.

Figure 6: Total and Mean Demand for 2022 by Quarter

In this chart we can also see that within each month there are patterns of variation in the daily mean demand (indicated by the red line). A more granular examination of the data, this time at a week level, shown in Figure 7 below, shows that within a week there is a distinct pattern of changing demand. Weekdays have the highest mean daily electricity demand, before a sharp drop to lower levels on Saturday and an even further decline on Sunday.

A graph with a line going up

Description automatically generatedThis is likely caused by industrial and commercial activities, which are responsible for 42% of energy consumption in Australia, as opposed to residential usage at 11% [11]. Commercial and industrial activities typically have higher electricity demand during weekdays and conduct less energy intensive operations on weekends, hence supporting the trend we see in the chart above [12].

Figure 7: Mean Total Demand by Weekday

The heatmap in Figure 8 shows electricity demand broken down into a 24-hour period. Unsurprisingly, hourly total electricity demand in NSW is lower in the early hours of the morning, but not at zero, as many critical operations, such as hospitals, continue to consume electricity regardless of the hour of day. Electricity demand begins to increase more significantly from 6am onwards, generally reaching highest levels of demand around 3pm to 8pm.

A screenshot of a graph

Description automatically generated

Figure 8: Hourly Demand for a Four Week Period

## 4.2 Bivariate Analysis – Electricity Demand and Temperature

Please find correlation coefficients between (demand and temperature) & (demand and population) and talk about that in this section.’

## 4.3 Timeseries Analysis

To ensure the suitability of our data for time series analysis, we conduct the Dickey-Fuller test to check for stationarity. Additionally, we decompose the time series into its trend, seasonality, and residual components to gain a deeper understanding of its structure.

A screenshot of a test

Description automatically generated

Figure 9: Dickey-Fuller Test Results for Total Demand Timeseries

A graph of a trade

Description automatically generated with medium confidence

Figure 10: Decomposition of Total Demand Timeseries

The outcomes of the Dickey-Fuller Test (Figure 9) play a pivotal role in gauging the stationarity of time series data. With a notably high p-value of 0.289465, it implies that the null hypothesis—presuming the existence of a unit root denoting non-stationarity—cannot be dismissed at conventional significance levels. This suggests that the time series data for electricity demand in New South Wales likely lacks stationarity, implying the presence of trends or seasonality that may necessitate further scrutiny and adjustments before deploying certain time series forecasting models. Given this non-stationarity, we'll forgo delving into the Dickey-Fuller Test outcome for now and proceed directly to leveraging the data for forecasting purposes. The choice of models will account for the seasonality and trend present in the dataset [13].

Plotting the autocorrelation and partial autocorrelation will help decide how many lags to choose for the model. An autocorrelation (ACF) plot represents the autocorrelation of the series with lags of itself. A partial autocorrelation (PACF) plot represents the amount of correlation between a series and a lag of itself that is not explained by correlations at all lower-order lags. Ideally, we want no correlation between the series and lags of itself. Graphically speaking, we would like all the spikes to fall in the blue region. But we can see there are several spikes above the blue region at various lags meaning there are correlations at those lags.A graph with blue dots and lines

Description automatically generated

Figure 11: PACF Plot for Total Demand

A graph with blue lines and dots

Description automatically generated

Figure 12: ACF Plot for Total Demand

# 5 Analysis and Results

Table 1 below shows the summary of the evaluation metrics for each model implemented in this research.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model | Dependent Variable | Independent Variables | MAE | MAPE | R^2 | RMSE |
| 1. **Linear** | Electricity Demand NSW | Population, Temperature | 186.05 | 2.27% | 0.594 | 208.01 |
| 1. **HWES** | Electricity Demand NSW | None | 452.73 | 5.93% | 0.1562 | 512.50 |
| 1. **SARIMAX** | Electricity Demand NSW | None | 264.68 | 3.56% | 0.7079 | 301.51 |
| 1. **Prophet** | Electricity Demand NSW | None | 222.96 | 2.98% | 0.7454 | 281.45 |

Table 1 Summary of Results

## 5.1 Preliminary Linear Model – Forecasting Temperature

In this research, we embarked on an ambitious journey to forecast energy demand for the next decade (2024-2033), leveraging historical data spanning from 2010 to 2023, along with projections of temperature and population growth. We began by developing a Linear Regression Model, a decision driven by its suitability for capturing the direction and strength of relationships between energy demand and our chosen predictors. A simple model is generally best to commence with, allowing benchmarking and comparisons of performance with more elaborate models.

Our first challenge was ensuring consistency of the data prior to use as predictors. While we were able to source NSW population data with projections to 2033, we were not able to find a temperature dataset with annual projections extending to this same year. We decided that an important first step would be to forecast future NSW temperatures. With additional data sourced from the Bureau of Meteorology, we extended the provided temperature data set to include values reaching the end of 2023. Using NSW Government climate projections of a one degree increase in mean annual temperature by 2033, we built and fit a linear model that considered this rate of change.

The linear regression analysis yielded an intercept of -178.896 and a coefficient of 0.10011308. This coefficient means, for every one-year increase, the temperature is expected to increase by approximately 0.1001 degrees Celsius, hence fitting with the NSW Government’s projections. Thus, from the model we were able to derive temperature values for the years 2024 to 2033, which are shown in Figure 13 and would be used alongside the population projections in the electricity demand modelling work.

A graph with blue lines

Description automatically generated

Figure 13: Temperature Predictions for 2024-2033

## 5.2 Linear Model to Forecast Electricity Demand

During the pre-processing phase we divided our dataset into training and test sets, avoiding shuffling, ensuring a chronological order to preserve the temporal nature of our data. We allocated the first 80% of the data set for the model training, reserving the remaining 20% for test purposes.

A screenshot of a graph

Description automatically generatedThe outcome of our model's training phase was both enlightening and a testament to the complexity of forecasting energy demand. Our model achieved an RMSE of 208.012, an R^2 of 0.594, an MAE of 186.045, and a MAPE of 2.268%. In comparison to the range of electricity demand values, ~1339MW difference between the maximum and minimum values in the dataset, the RMSE is ~15.5% of this difference. The R^2 value, while indicating a reasonable fit, also suggested a significant portion of the variability in electricity demand remains unexplained. The metrics combined indicate the model performance to be fair at best, with moderate accuracy, highlighting the challenges inherent in predicting electricity demand with high precision, particularly with such a simple model.

Figure 14: Actual vs Predicted Values for Total Demand as Influenced by Temperature & Population

As evident in the scatter plots in Figure 14, the predicted values for demand as influenced by population are relatively close to the actuals in most cases. However, the demand and temperature relationship show greater variation between predicted and actuals.

The coefficients for the regression were ~ -52.14 (temperature) and ~ -0.00069 (population). The temperature coefficient indicates that for each one-degree Celsius increase, electricity demand decreases by approximately 52.14MW. Of note is that this is a negative coefficient, meaning there is an inverse relationship between temperature and demand. Put into the reverse perspective, this means that each one degree decreases in temperature, there is a 52.14MW increase in electricity demand. This relationship makes sense when reflecting upon the findings from the exploratory data analysis, which showed that demand peaked during the colder months in NSW.

The population coefficient as it stands is at a scale of one individual person. For instance, it indicates that for each one additional individual in the population, electricity demand decreases by approximately 0.00069MW. To bring this to a larger scale, we can calculate that for everyone million people added to the population, the model predicts electricity demand will decrease by 687.17MW. It is important to note here that the coefficient is negative, indicating a negative relationship between population and electricity demand. Realistically, this relationship does not make sense. It is a common planning assumption that when a population grows, the consumption of resources increases. However, the model does not capture this. As will be further expanded on in the following discussion section, this could indicate that there are other important factors exerting influence on a decreasing electricity demand pattern that have not been captured in our analysis, for instance the increasing uptake of renewable rooftop solar electricity generation.

Reflecting on the predictions from our linear regression modelling forecast, as shown in Figure 15, it is evident that the predicted electricity demand for the next decade (2024-2033) follows a downward trend, with demand gradually decreasing from 7639.074 in 2024 to 7007.161 in 2033. This outcome, initially unexpected, may reflect a significant shift towards the adoption of renewable energy sources and increased energy efficiency measures.

A graph with a line going up

Description automatically generated

Figure 15: Linear Regression Predictions for Total Demand for 2024-2033

Overall, we were unsatisfied with the performance of this model and recognised that the relationship between electricity demand, population and temperature may not have been as strong as we initially believed. Hence, a decision was made to explore further modelling that did not utilise any independent variables.

## 5.3 Other Models

Given the limited granularity of temperature and population projection data, with only one value per year, our dataset became notably small. Consequently, we opted to forecast electricity demand up to 2033 without incorporating any independent variables. To achieve this, we explored classical time series models such as SARIMAX, Exponential Smoothing (specifically, Holt-Winters Exponential Smoothing), and Prophet.

The electricity demand data exhibited clear seasonal patterns, as illustrated by the yearly seasonal decomposition plot in figure. This seasonality, along with any trends or cyclic behaviour in the data, was effectively captured by the models, particularly by the Prophet model, as reflected in its superior performance metrics.

A graph showing a graph of electricity demand

Description automatically generated with medium confidence

Figure 16: Predicted Electricity Demand for 2024-2033

Figure 16, above shows the predicted demand from the 3 timeseries forecasting models.

Results from the Holt-Winter's Exponential Smoothing yielded a MAPE of 5.93% and an RMSE of 512.49. This indicates that, on average, the model's predictions were off by 5.93% from the actual values, with a square root of the average squared differences between predicted and actual values being approximately 512.49. In the context of forecasting electricity demand, a MAPE below 10% is generally considered good, as demand can vary significantly due to various factors like weather, economic activity, and consumer behaviour [14]. RMSE values are scale-dependent, meaning their interpretative value is highly dependent on the magnitude of the data being forecasted. In large-scale systems, such as a major city's electricity grid, an RMSE of 512.49 might be relatively minor, whereas, for a smaller grid, it could be significant. Figure above shows that the Holt winter prediction curve closely follows the testing curve and captures the trends and seasonal fluctuations well.

SARIMAX, configured as (1, 1, 1)x(0, 1, [1, 2], 12), showed a MAPE of 3.56% and an RMSE of 301.51, demonstrating better accuracy in forecasting compared to the Holt-Winter model. The MAPE of 3.56% indicates that, on average, the forecasted values are only 3.56% away from the actual values. This suggests a higher level of accuracy compared to the Holt-Winters model's MAPE of 5.93%. A lower MAPE suggests that the SARIMAX model is better at capturing the underlying patterns and dynamics of electricity demand, resulting in more accurate forecasts. An RMSE of 301.51 indicates that, on average, the model's predictions deviate from the actual observed values by about 301.51 MW. This is substantially lower than the RMSE of the Holt-Winters model (512.49 MW). A lower RMSE suggests that the SARIMAX model's predictions are closer to the actual values, indicating higher precision and reduced variability in forecast errors compared to the Holt-Winters model. The SARIMAX model's improved performance might be attributed to its ability to capture complex temporal dynamics, including autoregressive (AR), differencing (I), and moving average (MA) components, as well as seasonal effects. However, SARIMAX models can be more complex to interpret compared to Holt-Winters, given their reliance on a larger number of parameters and their statistical nature. The inclusion of seasonal differencing ([1, 2]) in the SARIMAX model suggests that it captures more nuanced seasonal patterns compared to the Holt-Winters model. This is particularly relevant for electricity demand forecasting, where seasonal factors like weather and time of year play a significant role.

Prophet achieved the best performance among the three, with a MAPE of 2.98% and an RMSE of 281.45. The MAPE of 2.98% indicates that, on average, Prophet's forecasted values are only 2.98% away from the actual values. This represents a higher level of accuracy compared to both the SARIMAX (3.56%) and Holt-Winters (5.93%) models. A lower MAPE suggests that Prophet is better at capturing the underlying patterns and dynamics of electricity demand, resulting in more accurate forecasts. With an RMSE of 281.45, Prophet's predictions deviate from the actual observed values by about 281.45 MW. This is lower than both SARIMAX (301.51) and Holt-Winters (512.49) models. A lower RMSE indicates that Prophet's forecasts are closer to the actual values, suggesting higher precision and reduced variability in forecast errors compared to other models. The coefficient of determination (R²) of 0.71 indicates that approximately 71% of the variation in the predicted values is accounted for by the test values. In other words, Prophet's model explains 71% of the variability in electricity demand, which is a substantial portion. A higher R² value suggests that the model's predictions closely match the observed data, indicating a good fit of the model to the underlying patterns in electricity demand. One thing to watch out for is that Prophet is a complex model, and its inner workings might not be easily interpretable to users without a deep understanding of its underlying algorithms. This lack of transparency can make it challenging to understand why certain predictions are made, limiting the model's interpretability [15].

From the results of the three Timeseries forecasting models Prophet is chosen to the best option. The results from the Prophet model are plotted for a decade ahead until end of the year 2033 as shown in Figure 17.

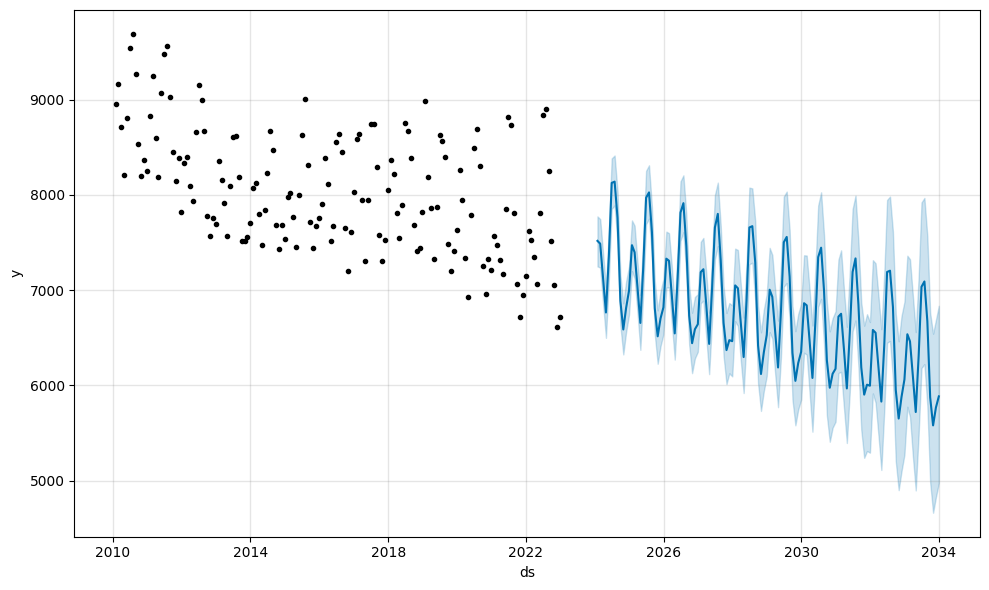


Figure 17: Predicted Total Demand from Prophet Model

# 6 Discussion

Answering the research question involves three main components.

1. Extract the predicted demand for the year 2033 from Section 5 Analysis and Results
2. Analyse the literature for Renewable Energy (RE) Generation in the year 2033
3. Combine results from the preceding steps and discuss whether the RE supply can meet demand.

## 6.1 Predicted Demand in 2033

The average of each year’s prediction from the Prophet model is consolidated as per figure below.

A screenshot of a computer

Description automatically generated

Figure 18: Average Demand Prediction from Prophet Model

The forecasted temperature for the year 2033 is **6252 MW**. An important aspect to note about this prediction is that the data source from the Market Management System database, National Energy Market, does not account for Rooftop PV in its fuel mix [12]. This omission represents a substantial portion of demand that is absent from our forecasted value. This can be confirmed by the Total Annual Demand value from Figure 1. To obtain the Total Annual Demand for the 3 remaining years, i.e. 2031, 2032 and 2033 a simple Straight Line Forecast through all the available points using Least Squares method was used [17]. The coefficients found for

y = ax + b was

a = 0.319,

b = 71.395

The forecasted Annual Demand for   
2031 = 78 TWh,

2032 = 78 TWh,

2033 = 79 TWh.

Converting 79 TWh to MW = 79\*1e6 / 365 \* 24 = 9018 MW.

The predicted Annual Demand including all fuel types is 9018 MW. There is a significant difference of 2766 MW per annum that isn’t covered by the data source and predictions from our work. This needs to be taken into consideration when answering the research question.

## 6.2 Renewable Energy Generation in 2033

Due to time constraints of this project, the forecast for electricity produced by renewable energy in the next decade is obtained from previous work instead of forecasting it by ourselves. Climate Energy Finance’s NSW Electricity model leverages the Open NEM database and models their estimates of Electricity Demand and Supply which includes Total Renewable Energy Generation (Annual in TWh). The full table is provided in Table 2 below [18]. To obtain the Total RE Generation for the 3 remaining years, i.e. 2031, 2032 and 2033 a simple Straight Line Forecast through all the available points using Least Squares method was used. The coefficients found for

y = ax + b was

a = 2.469,

b = -6.714.

The forecasted RE Generation for

2031 = 45 TWh,

2032 = 48 TWh,

2033 = 50 TWh.

A table of numbers and text

Description automatically generated with medium confidence

Table 2: CEF’s NSW Electricity Model of Demand and Supply (Annually, TWh) [18]

## 6.3 Can RE Projection in 2033 meet NSW demand in 2033?

From Section 6.1, The annual demand for electricity in 2033 from our results is **6252 MW** (without Rooftop PV in the Fuel mix).

From Section 6.2, The annual RE Generation for 2033 is 50TWh or 5708 MW.

(Convert annual value TWh to daily value in MW: 5x1e7 / 365 days \* 24 hours = 5708 MW)

As the demand is higher than Annual RE Generation, 6252 MW > 5708 MW, we can conclude that NSW’s electricity demand cannot be met by Renewable Energy sources in the next 10 years.

It is important to approach this conclusion with caution. While the current situation may indeed show that the demand for electricity in New South Wales (NSW) exceeds the annual renewable energy (RE) generation capacity, several factors need consideration:

**Demand Growth:** Electricity demand is not static and tends to increase over time due to factors such as population growth, economic development, and changes in consumer behavior. Therefore, projecting future demand accurately is crucial.

**Renewable Energy Expansion:** Governments and energy stakeholders often have plans to expand renewable energy capacity over time. These plans may include building new wind farms, solar installations, hydroelectric plants, and other renewable energy infrastructure. It's essential to consider these future developments when assessing the capacity to meet demand.

**Storage and Grid Integration:** Renewable energy sources such as solar and wind are intermittent, meaning they generate electricity when the sun is shining, or the wind is blowing. Effective energy storage solutions and grid integration technologies, such as battery storage and smart grid systems, can help manage fluctuations in renewable energy generation and ensure reliability of supply.

**Energy Efficiency and Demand-Side Management:** Implementing energy efficiency measures and demand-side management programs can help reduce electricity demand and alleviate pressure on the grid. These measures can include promoting energy-efficient appliances, implementing demand response initiatives, and incentivizing energy conservation practices.

**Policy and Regulatory Environment:** Government policies and regulations play a significant role in shaping the energy landscape. Policies that support renewable energy deployment, such as renewable energy targets, feed-in tariffs, and carbon pricing mechanisms, can accelerate the transition to a renewable energy future.

**Technological Innovation:** Advances in renewable energy technologies, energy storage solutions, and grid management techniques continue to evolve rapidly. These innovations have the potential to increase the efficiency, reliability, and cost-effectiveness of renewable energy systems over time.

# Conclusion

With non-renewable forms of electricity generation causing significant environmental challenges, there is a high priority globally to transition to renewable energy sources to meet current and future electricity demands.

The Ecological Defenders Office is an environmental advocacy group in New South Wales who aim to drive progress towards a sustainable energy future. To do this they require reliable estimates of future energy demands and renewable capacities to make informed decisions on policies and funding. As such, they have commissioned this piece of work to understand whether NSW can meet electricity demand solely from renewable sources within the next decade.

The conclusion is that the transition to a fully renewable energy-based electricity supply is not achievable within NSW in the next 10 years. The annual demand prediction for 2033 is not met by the Renewable energy generation in 2033. It is important for the Ecological Defenders Office to approach this conclusion with caution. While the current situation may indeed show that the demand for electricity in New South Wales (NSW) exceeds the annual renewable energy (RE) generation capacity, it's premature to conclude that renewable energy sources will be unable to meet demand in the next 10 years without considering some factors as discussed in Section 6.3. A comprehensive analysis that considers future developments, policy interventions, and technological advancements is necessary to make more accurate projections about the future energy landscape.

Although our research attempts to help The Ecological Defenders Office within the 6-week project timeframe, it is evident that the analysis is only one aspect of predicting whether RE Generation in 2033 can meet NSW’s electricity demand. The question of whether renewable energy can meet New South Wales (NSW) electricity demand in 2033 is complex and depends on various factors such as the rate of renewable energy development, technological advancements, energy policies, and energy demand forecasts. To address this question effectively, a thorough analysis of current renewable energy capacity, future renewable energy projects, energy consumption trends, and potential challenges in scaling up renewable energy generation would be necessary. Additionally, considerations such as grid infrastructure, energy storage capabilities, and the integration of intermittent renewable sources into the electricity grid would play significant roles in determining the feasibility of meeting NSW electricity demand solely through renewable energy by 2033.

# Further Work

There's a lot we can do to make this research even better. We can start by making our model better or trying out different ways to predict things. For example, in the Holt-Winter Model, we can adjust how we smooth out trends, seasons, and levels.

Adding more data could also help make our predictions more accurate. This could include things like weather conditions, economic factors, or changes in energy policies. It's really important to keep updating and adjusting our model as we get new information to make sure it stays accurate.

We need to keep testing our SARIMAX model with different sets of data and at different times to make sure it still works well. We have to be on the lookout for any changes in the data over time and adjust our model accordingly. If we find any unusual data points, we need to investigate them to see how they might affect our predictions and make changes to our model if needed.

When it comes to predicting future energy demand using data from CEF and Renewable Energy, it's a good idea to try out different ways of predicting and see which one works best. We need to test these methods really well to make sure they're accurate and better than what we're already using.

# References

|  |  |
| --- | --- |
| [1] | S. H. Chowdhury and A. M. T. Oo, “Study on electrical energy and prospective electricity generation from renewable sources in Australia,” *Renewable and Sustainable Energy Reviews,* vol. 16, no. 9, pp. 6879-6887, 2012. |
| [2] | J. P. McCarty, “Ecological Consequences of Recent Climate Change,” *Conservation Biology,* vol. 15, no. 2, pp. 320-331, 2001. |
| [3] | N. Government, “NSW Climate and Energy Action,” 2023. [Online]. Available: https://www.energy.nsw.gov.au/nsw-plans-and-progress/major-state-projects/shift-renewables/renewable-energy-nsw . [Accessed 16 March 2024]. |
| [4] | C. f. C. H. a. Development, “Chapter 33. Conducting a Direct Action Campaign | Section 20. Advocacy Over and For the Long Term,” [Online]. Available: https://ctb.ku.edu/en/table-of-contents/advocacy/direct-action/long-term-advocacy/main. [Accessed 16 March 2024]. |
| [5] | Q. A. Al-Haija, O. Mohamed and W. A. Elhaija, “Predicting global energy demand for the next decade: A time-series model using nonlinear autoregressive neural networks,” *Energy exploration & exploitation,* vol. 41, no. 6, pp. 1884-1898, 2023. |
| [6] | I. Fatema, X. Kong and G. Fang, “Electricity demand and price forecasting model for sustainable smart grid using comprehensive long short term memory,” *International Journal of Sustainable Engineering,* vol. 14, no. 6, pp. 1714-1732, 2021. |
| [7] | J. Korstanje, “How to Select a Model For Your Time Series Prediction Task [Guide],” neptune.ai, 14 12 2021. [Online]. Available: https://neptune.ai/blog/select-model-for-time-series-prediction-task. [Accessed 20 03 2024]. |
| [8] | N. E. Benti, M. D. Chaka and A. G. Semie, “Forecasting Renewable Energy Generation withMachine Learning,” *Sustainability,* vol. 15, 2023. |
| [9] | R. J. Hyndman and G. Athanasopoulos, Forecasting : Principles and Practice, Heathmont, Vic.: Otexts, 2018. |
| [10] | “The State of Weather and Climate Extremes 2022,” Australian Research Council, 2024. [Online]. Available: https://climateextremes.org.au/Stateof2022/. [Accessed 04 2024]. |
| [11] | NSW Environment Protection Authority (EPA), “Energy Consumption,” NSW Environment Protection Authority (EPA), 2024. [Online]. Available: https://www.soe.epa.nsw.gov.au/all-themes/human-settlement/energy-consumption#transport. [Accessed 04 2024]. |
| [12] | S. Cicala, “Powering Work from Home,” www.nber.org, 12 10 2020. [Online]. Available: https://www.nber.org/papers/w27937. [Accessed 04 2024]. |
| [13] | G. Dattaradon, “Time Series Forecasting: Anticipating Electricity Demand Patterns in New South Wales, Australia,” Medium, 21 03 2024. [Online]. Available: https://medium.com/@gigi.dattaradon/time-series-forecasting-anticipating-electricity-demand-patterns-in-new-south-wales-australia-1d1453d5e81c. [Accessed 13 04 2024]. |
| [14] | J. J. M. Moreno, A. P. Pol, A. S. Abad and B. C. Blasco, “Using the R-MAPE index as a resistant measure of forecast accuracy,” *Psicothema,* vol. 25, no. 4, pp. 500-506, 2013. |
| [15] | J. Korstanje, “How to Select a Model For Your Time Series Prediction Task [Guide],” 2023. [Online]. Available: https://neptune.ai/blog/select-model-for-time-series-prediction-task . [Accessed 16 March 2024]. |
| [16] | AEMO, “AEMO Data Dashboard - Mix Summary,” AEMO, 13 04 2024. [Online]. Available: https://aemo.com.au/en/energy-systems/electricity/national-electricity-market-nem/data-nem/data-dashboard-nem. [Accessed 13 04 2024]. |
| [17] | J. Stangroom, “Quick Linear Regression Calculator,” Socscistatistics.com, 2019. [Online]. Available: https://www.socscistatistics.com/tests/regression/default.aspx. [Accessed 13 04 2024]. |
| [18] | T. Buckley, “More coal subsidies to extend Eraring?,” https://climateenergyfinance.org/, Sydney, 2024. |

# Appendix

## Detailed Data Source information

|  |  |
| --- | --- |
| Dataset: **Total Electricity Demand (NSW)** | |
| Source | Market Management System database, National Energy Market |
| Format | CSV |
| Storage | 5.8MB |
| Variables | DATETIME: Date and time interval of each observation in the format (dd/mm/yyyy hh:mm). In 5 minute increments.  TOTALDEMAND: Total demand (MW)  REGIONID: Region Identifier (i.e. NSW1) |
| Messiness | NA: 0  Duplicates: 0  Inconsistencies:   * Format of DATETIME is object. In cleaning, should change to a consistent datetime data type for easier manipulation. |
| Size | 196513 rows x 3 columns |
| Relevance | This data shows historical electricity demand in NSW from 2010-01-01 to 2021-03-18. Historical demand is a powerful feature, which can be used as a predictor variable in electricity forecasting models. |

|  |  |
| --- | --- |
| Dataset: **Air Temperature (NSW)** | |
| Source | Australian Data Archive for Meteorology |
| Format | CSV |
| Storage | 6.9MB |
| Variables | DATETIME: Date time interval of each observation (dd/mm/yyyy hh:mm)  TEMPERATURE: Air temperature (°C)  LOCATION: Location of a weather station (i.e. Bankstown weather station) |
| Messiness | NA: 0  Duplicates: 13 duplicated rows  Inconsistencies:   * Format of DATETIME is object. In cleaning, should change to a consistent datetime data type for easier manipulation. * The smallest gap between temperature observations is 1 minute and the largest gap is 3 days, 18hrs and 30mins. |
| Size | 220326 rows x 3 columns |
| Relevance | Air temperature data covering 2010-01-01 to 2021-03-18 can also be considered as a potential predictor variable in forecasting. This data will be examined more deeply in the exploratory data analysis stage to identify if there is indeed a correlation between temperature and electricity usage. If this is proven, the data will form part of the modelling work undertaken in this project. |

|  |  |
| --- | --- |
| Dataset: **Population and Projected Population Totals (persons), 1971-2061** | |
| Source | 2022 NSW Common Planning Assumption Projections, NSW Government |
| Format | XLSX |
| Storage | 1.4MB |
| Variables | YEAR: The calendar year for each population projection (yyyy)  POPULATION: Number of total persons living in NSW in a given year |
| Messiness | NA: 0  Duplicates: 0  Inconsistencies: 0 |
| Size | 91 rows x 2 columns |
| Relevance | This data covers NSW population from 1971 to 2023, and then projected population to 2061. This data will be examined more deeply in the exploratory data analysis stage to identify if there is a relationship between population and electricity demand. If this is proven, the data will form part of the modelling work undertaken in this project. |

|  |  |
| --- | --- |
| Dataset: **Forecasted Demand (NSW)** | |
| Source | Market Management System database, National Energy Market |
| Format | CSV |
| Storage | 739.6MB |
| Variables | DATETIME: Date time interval of each observation (dd/mm/yyyy hh:mm). In half-hourly increments.  FORECASTDEMAND: Forecast demand (MW)  REGIONID: Region Identifier (i.e. NSW1)  PREDISTPATCHSEQNO: Unique identifier of predispatch run (YYYYMMDDPP)  PERIODID: Period count, starting from 1 for each predispatch run.  LASTCHANGE: Date time interval of each update of the observation (dd/mm/yyyy hh:mm) |
| Messiness | NA: 0  Duplicates: 284 duplicated rows  Inconsistencies:   * Format of DATETIME and LASTCHANGED are an object. In cleaning, should change to a consistent datetime data type for easier manipulation. |
| Size | 10906019 rows x 6 columns |
| Relevance | This data contains forecasted electricity demand in NSW from 2010-01-01 to 2021-03-18. This data can serve as an effective validation dataset, enabling the team to assess the accuracy of the predictive forecast models developed. |

|  |  |
| --- | --- |
| Dataset: **Electricity generation in New South Wales, by fuel type, physical units, financial year** | |
| Source | Department of Climate Change, Energy, the Environment and Water, Australian Energy Statistics, Australian Government |
| Format | XLSX |
| Storage | 150KMB |
| Variables | YEAR: The financial year corresponding to the amount of electricity generated in that time period (yyyy-yy)  FUEL SOURCE: The type of energy source (ie. Types of renewable and non-renewable fuels)  ELECTRICITY GENERATED: The amount of electricity generated by fuel type (GWh) |
| Messiness | NA: 0  Duplicates: 0  Inconsistencies:   * Years in this dataset are financial years, whereas other data to be used in this project is generally in the format of calendar year. |
| Size | 13 rows x 14 columns |
| Relevance | This data details the amount of electricity generated by each fuel source from 2008-09 to 2021-22 in NSW. This data is key to the analysis and modelling that will be conducted to assess the feasibility of a 100% renewable electricity scenario within 10 years. This historical data can serve as a predictor variable in the modelling. |