**AIM**

The aim of this project is to investigate how different weather parameters (temperature, humidity etc..) of a region affect that region’s energy demand. This would be particularly useful for energy providers in helping them to meet consumer demand whilst preventing shortages, and allowing them to maximise profit. If they were to focus on wind and water as their source of energy and monitor price surges as well as regional demand, they can alter the prices they charge depending on the region’s demand.

Environmentalists can use this data to further analyse the effect of global warming on climate change and how it affects energy usage in different regions. This will also aid the Bureau of Meteorology in weather forecasting and better equipping them to handle any weather-related disasters.

**DATASETS**

Two datasets were used in our research. We used the weather dataset which consisted of numerous weather parameters in various cities in Australia, this data spanned over a year. Our second dataset was the price demand dataset, which consisted of a record of whether or not a region in Australia had experienced a price surge or not on a specific day. The day was split into half an hour intervals and this dataset also spanned more than a year.

Both datasets were in the CSV file format and we used regions and cities to link both datasets. Each region in our price demand dataset was represented by a major city from that region, in our weather dataset.

**PRE-PROCESSING AND WRANGLIING**

Before we used our datasets for analysis, they needed to be cleaned up and pre-processed due to there being numerous missing values and other inconsistencies. If a column crossed a threshold of null values (we took 350 as the threshold value) then the column was removed entirely. We removed the null values from object data type and changed the null values from float and int datatype to be the mean of the column. Next, we converted the datatype of wind speed and date to make them easier to work with.

For our price demand dataset, we needed to ensure that the date was in a consistent format with our weather dataset otherwise comparisons were made incorrectly.

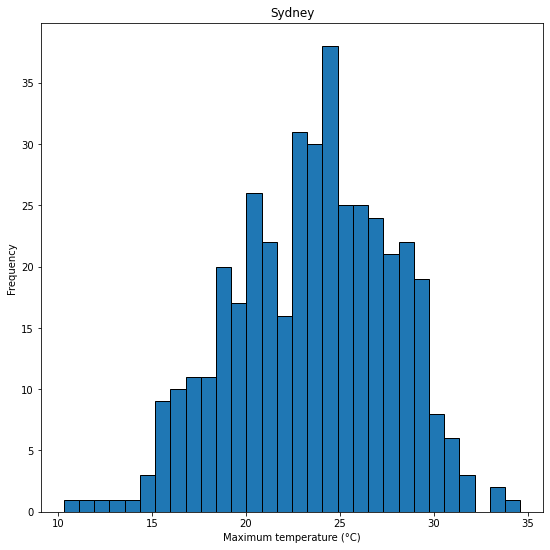
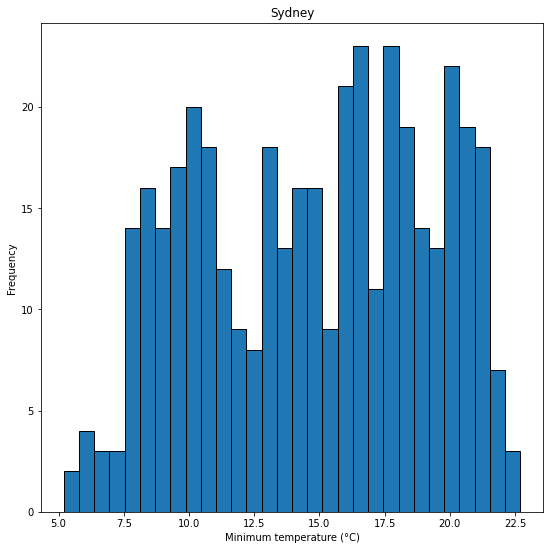
**ANALYSIS METHODS**

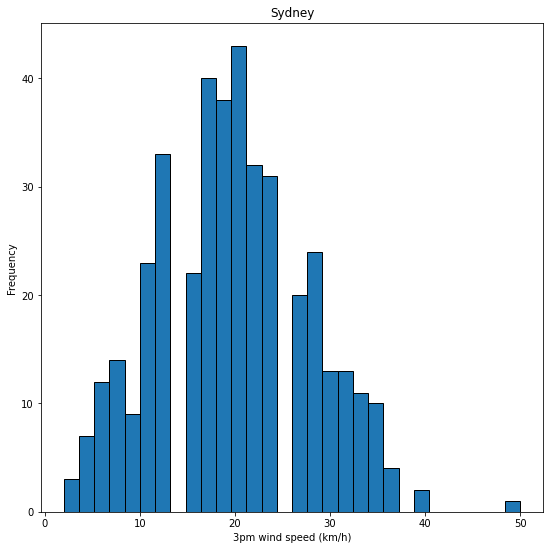
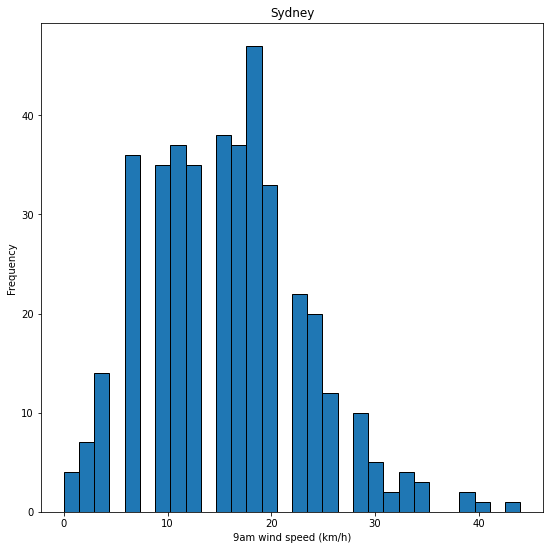
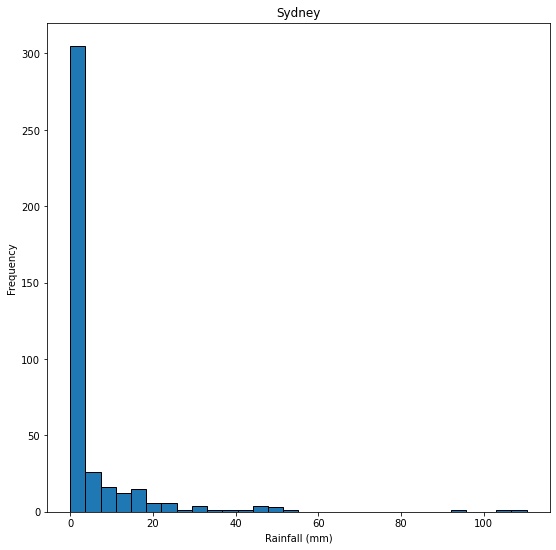
The weather parameters we used were both discrete and continuous. We used a correlation matrix for all four cities to find the pair-wise correlation between all of the columns in the dataframe. We removed highly correlated columns (threshold: abs(0.75)) to avoid overfitting our model due to multicollinearity. Then, we filtered the variables that are common for all 4 cities:

* Minimum temperature (°C)
* Rainfall (mm)
* Speed of maximum wind gust (km/h)
* 9am relative humidity (%)
* 9am wind speed (km/h)
* 9am MSL pressure (hPa)
* 3pm relative humidity (%)
* 3pm wind speed (km/h)
* 3pm MSL pressure (hPa)

We used histograms to see the distribution of variables in each city over the given time period (one year),time series enabled the visualisation of total energy demand. We used multiple linear regression to find the relationship between the city’s energy demand and its weather parameters.

The Rainfall and wind speed histograms are discussed below to provide a better understanding of data distribution,since most of the parameters are normally distributed, we use standardisation.





*Figure 1:* Distribution of variables for Sydney

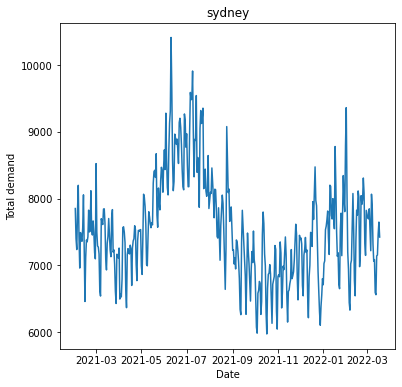
According to Sydney’s rainfall histogram, it can be inferred that the rainfall frequency is primarily light rain, with occasional heavy rain. It can be observed that the wind speed around 9am is, on average, 20 km/h. However, at about 3 pm, the wind speed increases to around 25 km/h, which is somewhat above the norm. As the histograms demonstrate, the average minimum temperature is approximately 17.5°C, whilst maximum temperature averages around 25℃.

Brisbane's rainfall histogram is almost identical to Sydney's, with generally light rain and rare heavy rain. Additionally, the 9am wind speed in Brisbane is lower than in Sydney, which averages 7.5 km/h and 12 km/h at 3 p.m. Furthermore, Brisbane has slightly higher temperatures than Sydney, with the minimum average temperature being 20℃ and the average maximum temperature being 30℃.

Adelaide's rainfall histogram is similar to both Sydney and Brisbane, with a majority of light rain and rare heavy rainfall. Adelaide’s 9am wind speed is around 12 km/h and approximately 17 km/h at 3pm, which is somewhat higher than Brisbane and marginally lower than Sydney. Additionally, Adelaide has a slightly lower minimum temperature than Sydney, around 13℃ and a maximum of approximately 17℃.

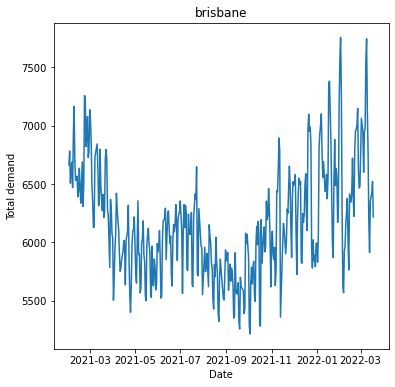
Lastly, Melbourne’s rainfall histogram is similar to Adelaide, Sydney, and Brisbane, with rainfall being primarily mild, and rarely heavy. Melbourne's wind speed between 9am and 3pm is somewhat greater than Brisbane's but far lower than Sydney and Adelaide (9km/h and 13 km/h, respectively). Further, Melbourne has the lowest minimum temperature of all of the cities, with a low of around 10℃ and a maximum temperature of approximately 17℃.

Each city’s time series chart shows the average total energy demand per day over the course of a year on the Y-axis, and the date on the X-axis.



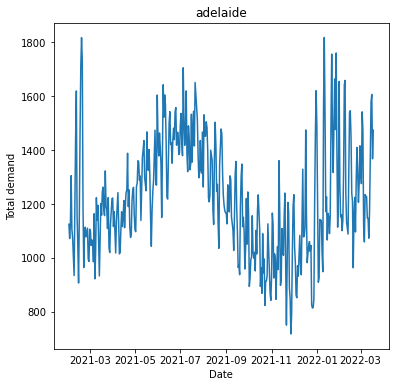
*Figure 2:* Time series of average energy demand vs date for Sydney (NSW)

Total energy consumption in Sydney peaks between June to August, which are the winter months. We can see another slight peak during January to February.



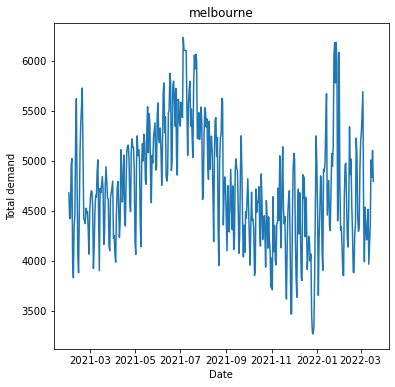
*Figure 3:* Time series of average energy demand vs date for Brisbane (QLD)

Total energy consumption in Brisbane takes a dip during August to October 2021 and peaks during January to March 2022. This differs from Sydney’s distribution.



*Figure 4:* Time series of average energy demand vs date for Adelaide (SA)

Adelaide’s energy consumption is very similar to Sydney with a rise in demand during June to August 2021 and January to March 2022.

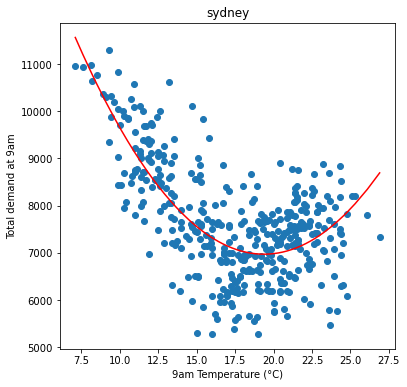
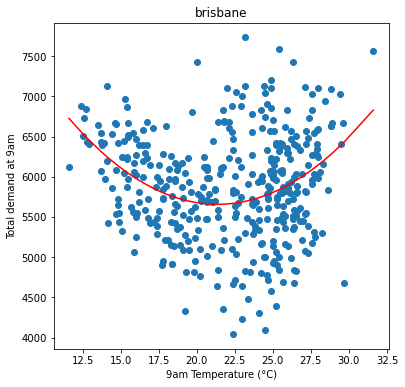


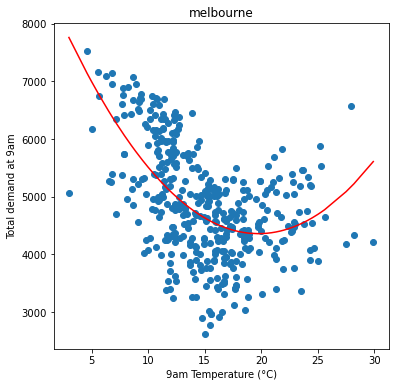
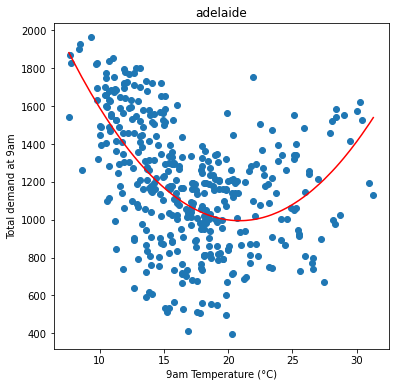
*Figure 5:* Time series of average energy demand vs date for Melbourne (VIC)

Melbourne’s energy demand is similar to that of Sydney and Adelaide. This may be due to all three cities being located in the south of Australia whereas Brisbane is not. This poses certain climatic differences in Brisbane compared to the other three cities which might be the reason for the different energy demand distribution.

We used scatter plots to visualise how the 9am and 3pm temperature affected energy demand.

**Distribution of 9AM Temperature of Sydney, Brisbane, Adelaide and Melbourne:**

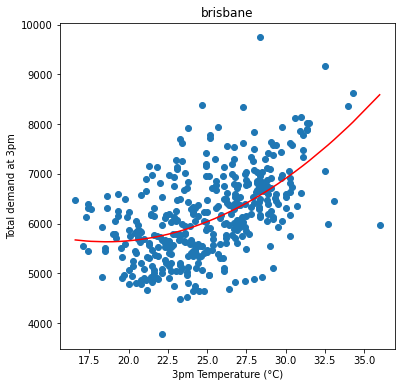
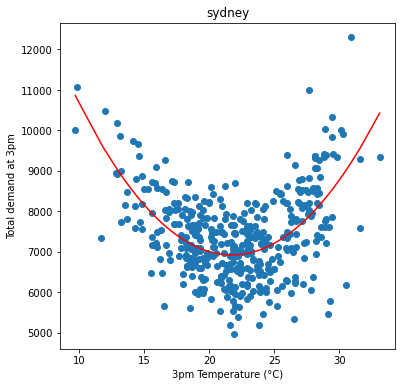
 

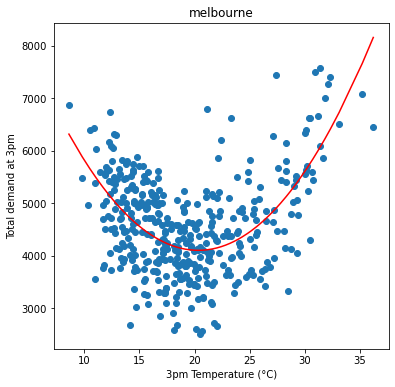
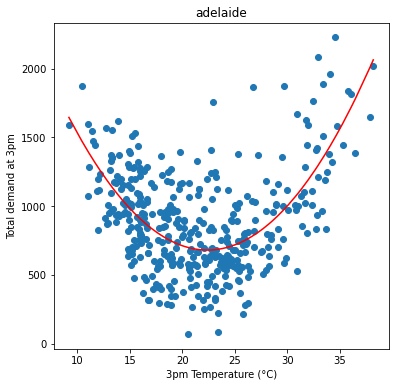


*Figure 6:* Distribution of 9 AM Temperature for all cities

We noticed that the relation between temperature and energy demand followed a common quadratic distribution. Energy demand is low when the temperature is between 17℃ - 22℃ in all four cities, and is high when temperature drops below 15℃ or goes above 25℃. This may be due to people using heaters when the temperature is low and air conditioning when the temperature is high.

**Distribution of 3PM Temperature of Sydney, Brisbane, Adelaide and Melbourne:**

****



*Figure 7:* Distribution of 3 PM Temperature for all cities

Similar to 9am temperature, there is a quadratic relationship between 3pm temperature and energy demand.

**MODELLING**

To see if weather affects energy demand for each region, we used multiple linear regression. Earlier, we saw that temperature and energy demand have a quadratic relation, so for temperature variables, new temperature squared columns were added.

We fit the model with all variables except date and time of maximum wind gust. Variables with high VIF score(>6) and high p-values(>0.05) were removed to get the most significant variables. We calculated the test, RMSE values, and trained R2 to see which variables are most significant and how they affect the energy demand. We also fit the model on temperature variables and then analysed the models to see which one gives the best results.

**MODEL ANALYSIS**

**Model 1: Using all variables(removing insignificant variables sequentially)**:

After fitting the model we calculate R2 and RMSE/Mean values for each city (Table 1).

|  | R2 test | R2 train | (RMSE/Mean) test | (RMSE/Mean) train |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
| Sydney | 0.13 | 0.45 | 0.098 | 0.075 |
| Brisbane | 0.32 | 0.54 | 0.064 | 0.050 |
| Adelaide | 0.32 | 0.35 | 0.147 | 0.134 |
| Melbourne | 0.17 | 0.30 | 0.108 | 0.100 |

*Table 1*: Goodness of fit

Table 1 illustrates that aside from Adelaide, there are huge discrepancies between R2 train and R2 test, indicating we need to remove some independent variables. For all data, the R2 train is significant indicating that significant variables are a better predictor than simply using average.

For Sydney we find that variables **9am wind direction\_NW, 9am wind direction\_WNW, 3pm wind direction\_NW, 3pm wind direction\_S, 3pm wind direction\_W, 3pm wind direction\_WSW, Maximum temperature, Rainfall, 9am relative humidity** play significant roles in determining energy demand

We get the following coefficients:

9am wind direction\_NW -0.696500

9am wind direction\_WNW -0.316372

3pm wind direction\_NW 0.862070

3pm wind direction\_S -0.544256

3pm wind direction\_W 0.534101

3pm wind direction\_WSW -0.601888

Maximum temperature (°C) -0.357295

Rainfall (mm) -0.110161

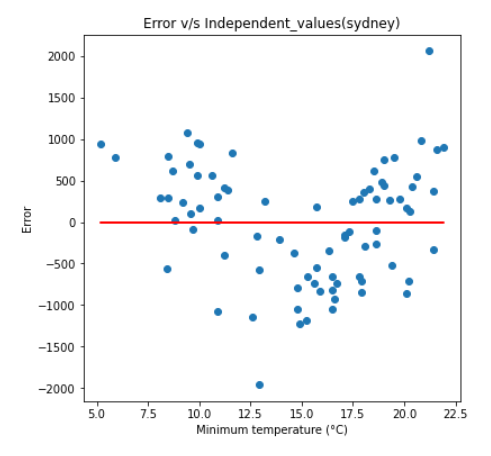
9am relative humidity (%) 0.367335

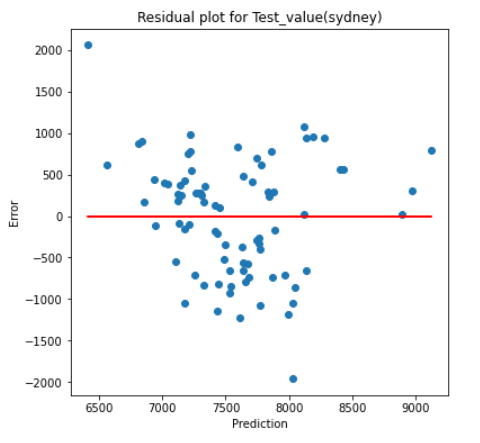
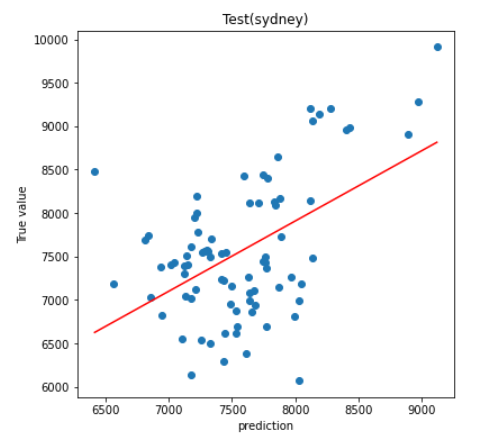
Positive coefficients mean the variable positively affects energy demand and vice-versa.

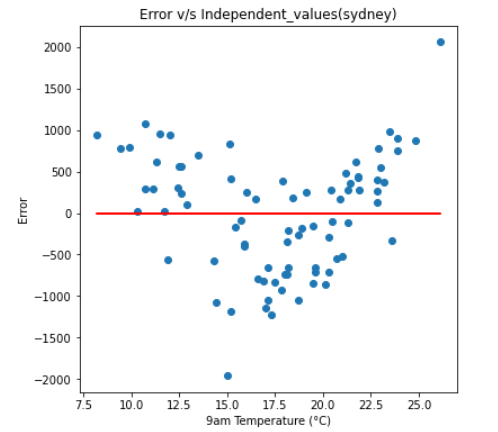
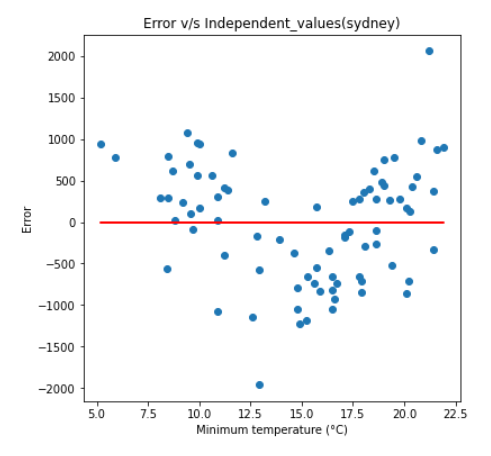
We see that the 9am wind direction in NW or WNW decreases average energy demanded for the day. Other cities are not explained for brevity.

**Model 2: Using all non-categorical variables:**

|  | R2 test | R2 train | (RMSE/Mean) test | (RMSE/Mean) train |
| --- | --- | --- | --- | --- |
| Sydney | 0.23 | 0.36 | 0.092 | 0.081 |
| Brisbane | 0.38 | 0.55 | 0.061 | 0.050 |
| Adelaide | 0.18 | 0.17 | 0.162 | 0.152 |
| Melbourne | 0.06 | 0.24 | 0.115 | 0.104 |

*Table 2*: Goodness of fit





The above residual plots found that errors are not random with respect to temperature parameters. Using the 9am and 3pm energy demand v/s temperature graphs we see a quadratic relation. Thus, squares of temperatures will be also used for model training.

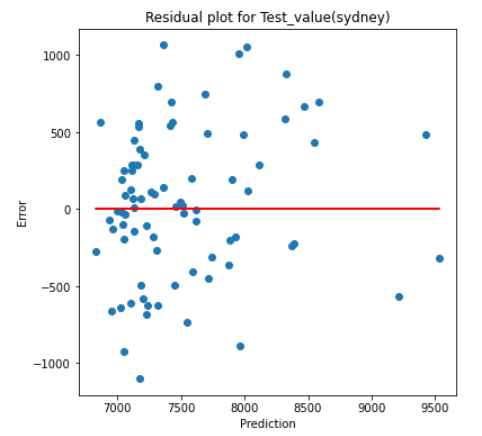
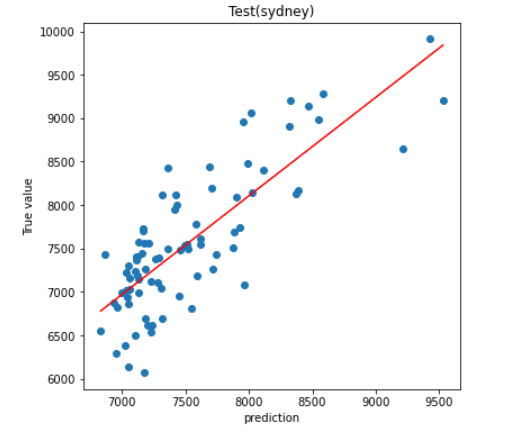
**Model 3: Using all temperature and temperature squared variables:**

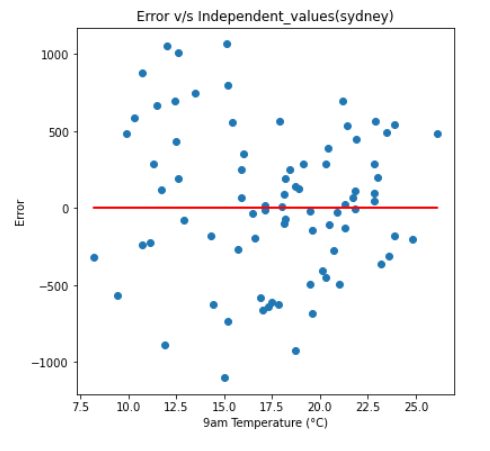
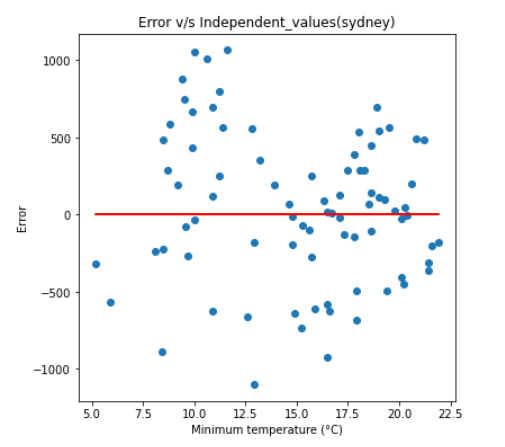
Since we are looking for a simple model, we go with temperature variables only and test the model.

|  | R2 test | R2 train | (RMSE/Mean) test | (RMSE/Mean) train |
| --- | --- | --- | --- | --- |
| Sydney | 0.63 | 0.62 | 0.063 | 0.062 |
| Brisbane | 0.69 | 0.68 | 0.043 | 0.042 |
| Adelaide | 0.62 | 0.54 | 0.111 | 0.113 |
| Melbourne | 0.34 | 0.48 | 0.096 | 0.086 |

*Table 3*: Goodness of fit

Table 3’s R2 train and R2 test values are close and high. RMSE values are lower than the previous model. Residual plot points are also randomly distributed, indicating that temperature indeed has a quadratic relation.

****



**Model 4: Using only one temperature parameter with its squares:**

Using Minimum temperature

|  | R2 test | R2 train | (RMSE/Mean) test | (RMSE/Mean) train |
| --- | --- | --- | --- | --- |
| Sydney | 0.41 | 0.47 | 0.08 | 0.074 |
| Brisbane | 0.65 | 0.59 | 0.045 | 0.047 |
| Adelaide | 0.37 | 0.25 | 0.142 | 0.145 |
| Melbourne | 0.33 | 0.2 | 0.097 | 0.106 |

*Table 4*: Goodness of fit

Sydney and Brisbane have high R2 train and R2 test values and a low RMSE to mean ratio, indicating that minimum temperature is a really good predictor of energy demand for these cities. But we see higher RMSE values compared to using all the temperatures.

Similar tests were done with maximum temperature, 9am temperature and 3pm temperature individually. We found different temperature parameters are significant for different cities:

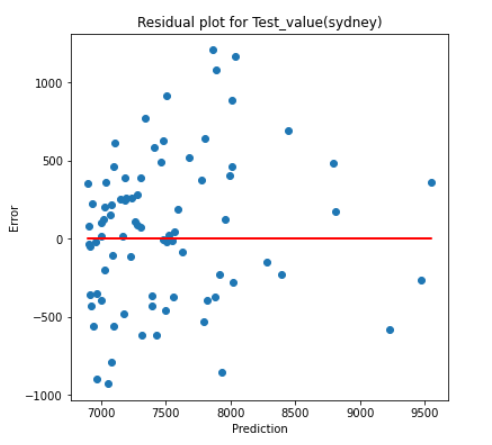
* Sydney: 9am temperature- highest R2 values and lowest RMSE.
* Brisbane: minimum temperature- highest R2 values and lowest RMSE.
* Adelaide: maximum temperature- highest R2 values and lowest RMSE.
* Melbourne: 3pm temperature- highest R2 values and lowest RMSE.

**Model 5: Using minimum and maximum temperature parameters with squares:**

|  | R2 test | R2 train | (RMSE/Mean) test | (RMSE/Mean) train |
| --- | --- | --- | --- | --- |
| Sydney | 0.64 | 0.58 | 0.062 | 0.065 |
| Brisbane | 0.66 | 0.68 | 0.044 | 0.043 |
| Adelaide | 0.62 | 0.53 | 0.111 | 0.115 |
| Melbourne | 0.39 | 0.47 | 0.093 | 0.087 |

*Table 5*: Goodness of fit

Comparing Model 3 and Model 5, we find very little change in RMSE. Thus, we can use Model 5 and get similar results.



**ANALYSIS CONCLUSION:**

In Model 1, we tried to find significant parameters that affect energy demand. Though our current model is not very good, we can see there are certain wind directions that affect the energy demand significantly. If in future we can get a better model, we can analyse the windflow pattern affecting energy demand and this may help city planners to build in a certain way that blocks or allows certain wind flow direction to reduce energy demand.

We found that using only Minimum and Maximum temperature gives us good predictions, though we need to use a squared term also. We will use this model since it is simpler with less parameters, it is better for explaining to our audience.

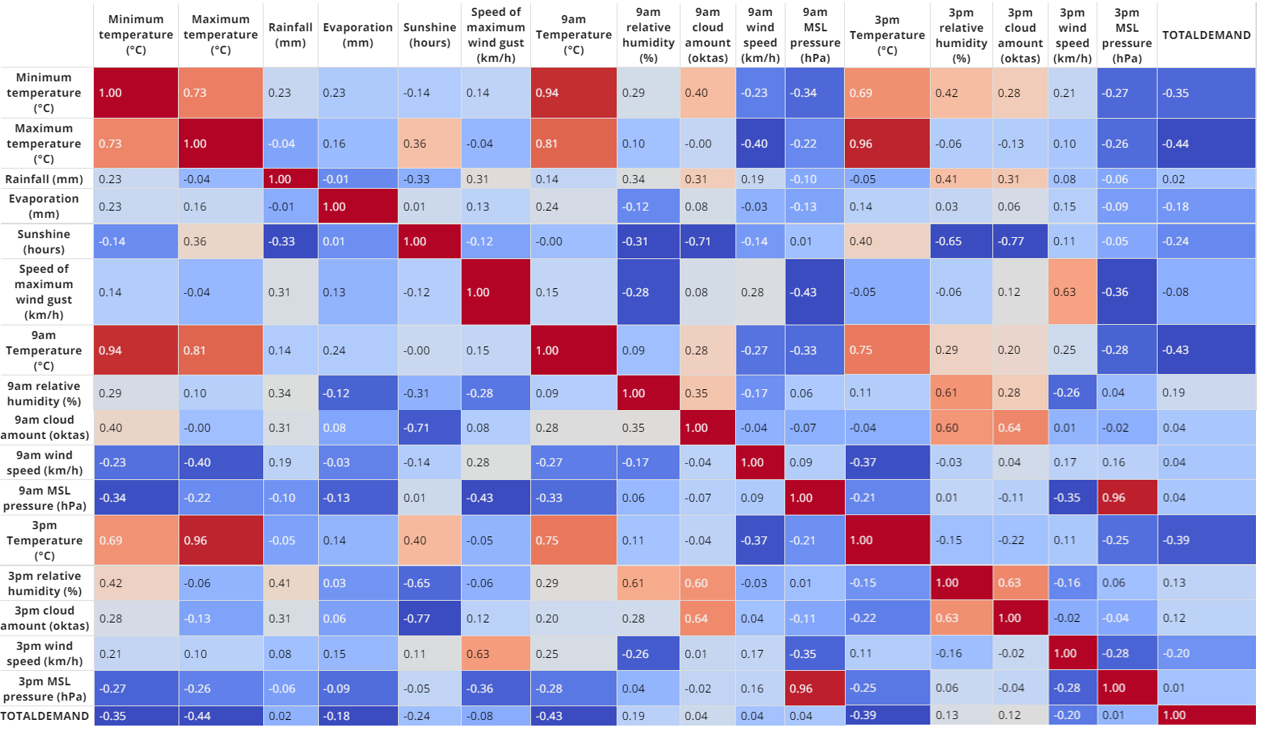
**EVALUATION**

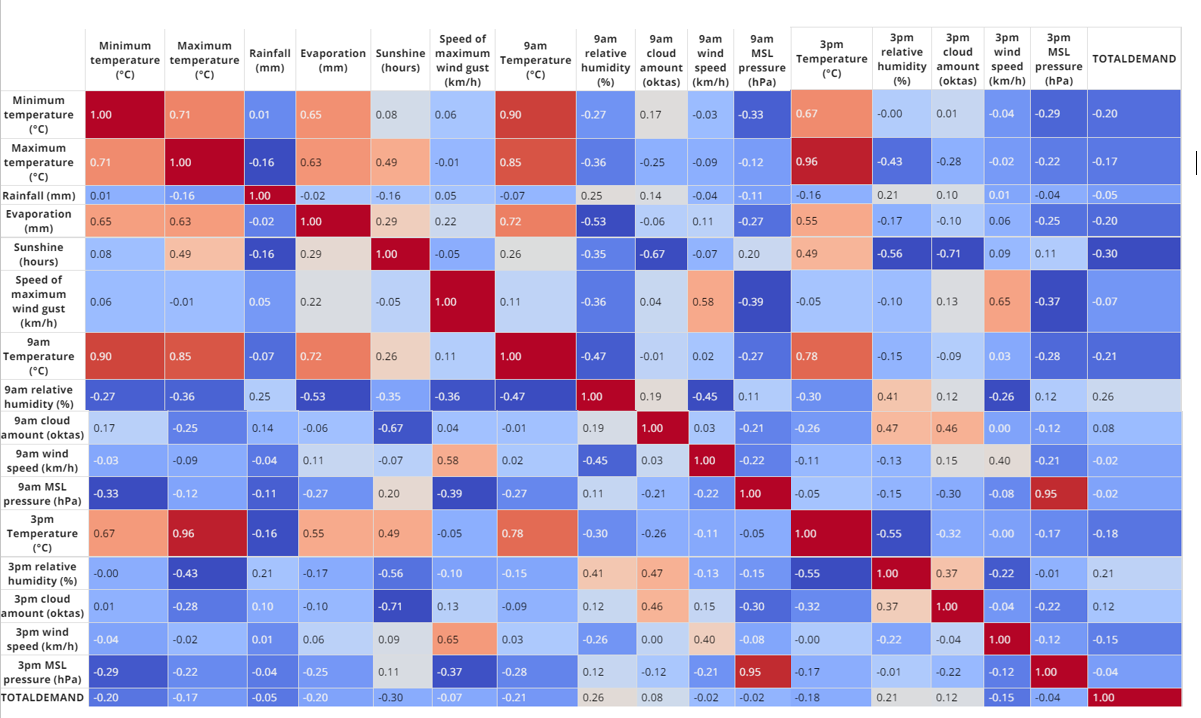
In this section we examine the limitations of our results and propose possible improvements for future research. Throughout all the models, Melbourne seems to have bad performance, it may be due its energy demand having low correlation with all its weather parameters(refer fig. 15 in appendix). We may need more weather parameters for Melbourne data.

We got energy data at 30 min intervals each day but the weather parameters were given at 1 day interval, getting weather values at 30 min may help in better analysis.

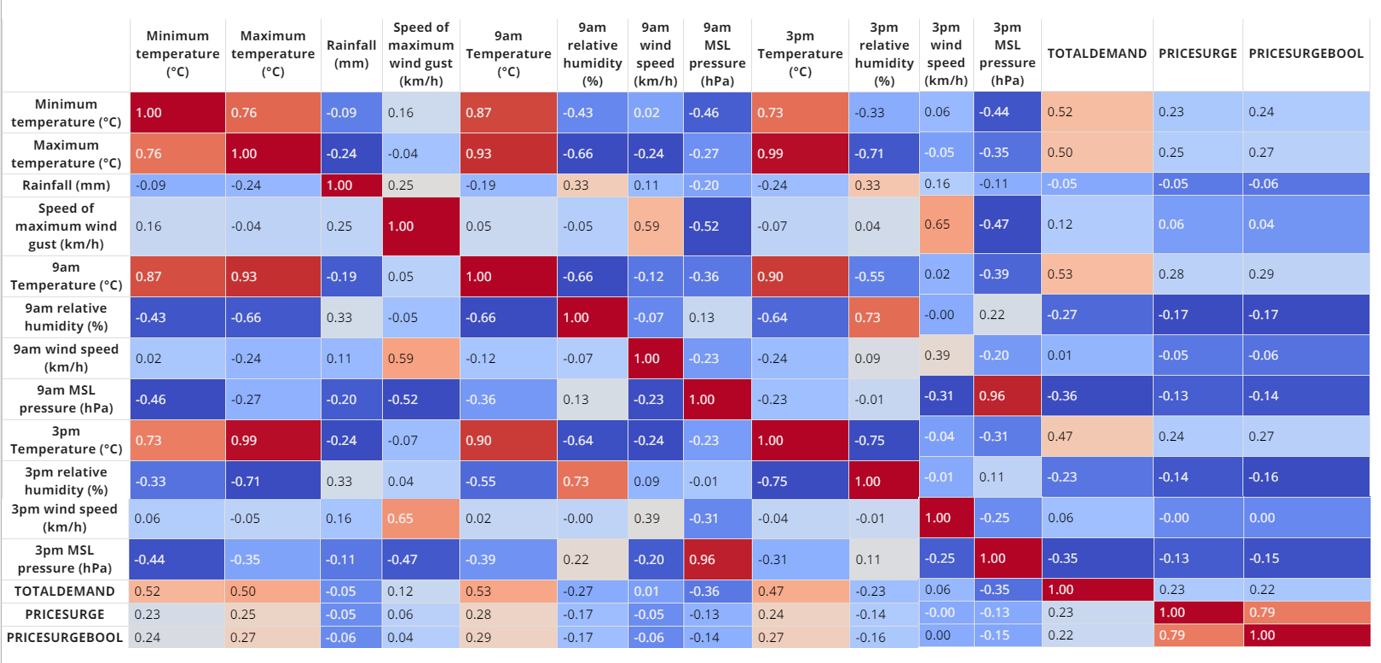
We considered the model to be linear for simplicity which introduced a lot of bias. New methods like Neural Networks can be used for better predictions.

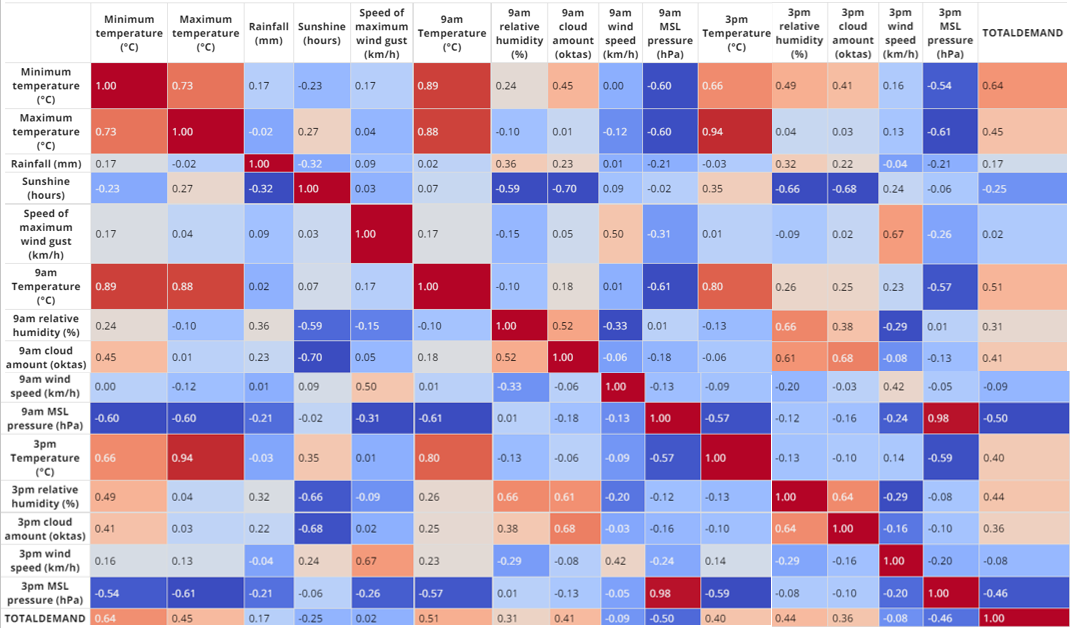
**APPENDIX:**

*Figure 14:* Correlation matrix showing the level of correlation between different weather parameters in Sydney



*Figure 15:* Correlation matrix showing the level of correlation between different weather parameters in Melbourne

*Figure 16:* Correlation matrix showing the level of correlation between different weather parameters in Adelaide



*Figure 17:* Correlation matrix showing the level of correlation between different weather parameters in Brisbane