Investigation of Temperature and Energy Demand in Australia by Group 2

ADS1002 Semester 2 2023
Final project report

	Name, Surname	Monash ID	Contribution (description)
1.	Shesh Murali	33890293	- Some time series and rooftop pv - Report writing: Exploratory data analysis (rooftop pv section) modelling (roof top pv) - Imputation
2.	Emilia Zhang		- Imputation - Report writing: none
3.	Luis Tanag	33890323	- Preprocessing, demand anomalies, modelling - Report writing: preprocessing, processing, manipulation, modelling
4.	Ashlyn Tiwari	33890358	- Time series - Report writing: Exploratory data analysis (time series section), conclusion
5.	Alysha Sharp	32524528	- Exploratory data analysis - Report writing: description of project, exploratory data analysis (first section)

TEMPERATURE AND ENERGY DEMAND MODELLING

Contents

1. Description of Project ……...………………………………………………… 3.
2. Our Data………………………………………………………………. 3.
3. Aim ……………………………………………………… 3.
4. Preprocessing ……………………………………………………………….. 4.
5. Processing and Manipulation ……………………………………………………… 5.
6. Exploratory Analysis …………………………………………… 6.
7. Time Series ………………………………………………………………. 7.
8. Modelling ……………………………………… 12.
9. Conclusion …………………………………… 19.

Description of Project

The increase in global concern regarding climate change and energy sustainability has highlighted the importance and relevance of understanding the relationship between energy demand and temperature. The increase of energy demand rises, societies start to urbanise and population is on the rise, this places increased pressure on energy systems that provide society the demand for energy. Moreover, understanding its relationship is vital for energy supply reliability and improving its security as challenges such as drastic weather changes and global warming can pose a threat to energy resilience. However, by providing an accurate modelling solution and understanding the complex relationship between temperature and energy demand effective policies can be made to address energy sustainability and economic challenges, ensuring the public is provided with affordable and reliable energy, no matter the conditions.

Our Data

Our project, upon the basis of this, utilised Australian Energy Market demand data in excel spreadsheets and weather data collated by the Bureau of Meteorology (BoM) Automatic Weather Stations (AWS) provided in a text document. The energy demand data was measured in megawatts, recorded half-hourly for the east coast states (NSW, VIC, SA, QLD and TAS) over a period of years, from 2000-2019. The weather data included many observations from precipitation to wind speed to sea level pressure, however, the air temperature in degrees Celsius was determined to be of most use to us in this investigation. Similarly to the energy data, this was collected half-hourly, over the same period of time and at weather stations in the same states. We were also provided with rooftop photovoltaic (PV) data in an excel spreadsheet which was measured in kilowatts (kW) yearly from 2006-2021, for each state and territory in Australia. We were required to import these separate data files into the web-based program Jupyter, where the coding language Python was applied to process the data and create dataframes for the purpose of analysis and modelling.

Aim

Our aim was to uncover and investigate the relationship between temperature and energy demand, with the hope to develop a model which illustrated as such. Reproducing the plot provided *(Figure 1)* in the project outline:

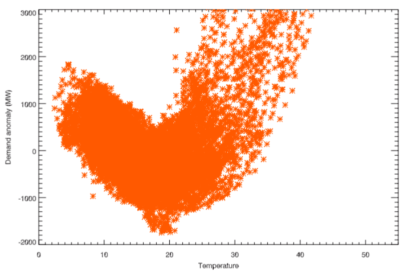


Figure 1: Demand anomalies versus temperature for 2019

Figure 1

and creating an accurate regression was one of a couple of our top priorities. We hypothesised that increasing temperature and decreasing temperature would increase energy demand, and that the energy demand would follow a cyclical behaviour as well. Likely to peak in the mornings and evenings, decrease during the middle of the day and increase on the weekends, following the stereotypical times that people are in and out of their homes. Despite our general hypothesis, we also expected to find data that would contradict this during our time-series plotting and further analysis. Through this project we strive to provide invaluable insights into the relationship between temperature and energy demand.

Preprocessing

The two largest datasets - these being the temperature and demand sets - were seen to be difficult to merge and/or analyse as a result of their features, with problems such as inconsistencies in timestamp columns, inconsistent data value types in the temperature dataset, and massive numbers of redundant columns proving to be major hurdles prior to our data analysis.

The first major problem we found and addressed was the lack of a consistent timestamp column from which we could merge the temperature and demand datasets on. The demand dataset had the most useful column for indicating time, with this being set out like so:

SETTLEMENTDATE
2000/01/01 00:30
2000/01/01 01:00
2000/01/01 01:30

This format was ideal for use as a timestamp, and as such, the goal was to use the existing time columns in the temperature dataset to create a new column with the same form as the above time column.

Year Month Day Hour Minutes in YYYY	MM	DD	HH24	MI format in Local time
2000	1	1	1	0
2000	1	1	1	30
2000	1	1	2	0

The above example is a sample of the time columns present in the temperature dataset, and in order to obtain a single time column with the correct form, these five columns had to be merged and the necessary separators inserted between each value. As a result, we could then merge both datasets together and convert the time column into a proper timestamp, which would allow the data to be grouped according to which date it corresponds to.

After this merge, we discarded every column that did not pertain to energy demand and temperature, with the format of our merged dataset being shown below *(Figure 2)*

	REGION	Air Temperature in degrees C	Energy generation (mW)	TOTALDEMAND	RRP
Timestamp					
2000-01-01 01:00:00	SA1	13.1	0.000000	1375.14833	38.54
2000-01-01 01:30:00	SA1	13.2	0.000000	1275.12667	14.37
2000-01-01 02:00:00	SA1	13.4	0.000000	1181.72167	14.49
2000-01-01 02:30:00	SA1	13.2	0.000000	1105.05000	13.98
2000-01-01 03:00:00	SA1	13.0	0.000000	1031.22500	10.35
...	...	...	...	...	...
2019-12-31 22:00:00	TAS1	15.6	157.061899	1032.56000	65.31
2019-12-31 22:30:00	TAS1	15.3	157.061899	1032.85000	80.97
2019-12-31 23:00:00	TAS1	14.9	157.061899	1024.17000	73.03
2019-12-31 23:30:00	TAS1	14.5	157.061899	1008.68000	81.76
2020-01-01 00:00:00	TAS1	14.2	183.860422	1006.70000	93.88

Figure 2

Our second hurdle was the inconsistency of data types in the air temperature column, with some values not being numerical data types despite representing numbers.This was rectified fairly quickly by iterating through the entire column and converting all data points into numbers. Additionally, missing values expressed as white spaces were present in this column, with approximately 1000 out of the 1.6 million total rows corresponding to white spaces. Due to the relatively small amount of these missing values, we initially settled on dropping these rows but later decided to impute these with the overall air temperature mean in order to prevent gaps in our time series graphs.

The final dataset - that of the average solar energy generation per year for each state - did not require as much preprocessing as the other two. This data, which was obtained from the Australian Energy Market Operator, sourced energy generation values from PV systems across each state in Australia from 2006 to 2021 in kilowatts. By using the df.drop function is used to manipulate and drop columns and rows that are considered irrelevant data for future manipulation and analysis.The major problem faced with the data was the duplicated 2021 and 2021.1 columns that were required to be removed to normalise the data. Moreover in regards to the data to be compared to demand data, the states ACT, NT and WA were not recorded, therefore they were removed from the data set. However, within the demand data in the year 2020 only 20 days were recorded, hence during exploratory analysis and modelling that year would drastically skew the data in the positive direction if compared with the rooftop pv data. Therefore all demand data were removed for the year 2020, similarly to provide an unbiased comparison, the year 2020 in the Rooftop PV data was removed. Ultimately, this data *(Figure 3)* set was relatively clean and required minimal change, thus data wrangling and manipulation was able to commence.

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
NSW	1044.984	2309.109	6516.811	25706.473	179092.635	374251.227	518840.662	652252.275	824216.341	1000627.317	1167297.424	1429882.577
VIC	857.842	1968.102	4697.987	15216.152	74175.361	214584.898	422168.723	553783.359	725483.148	875458.431	1020036.260	1225561.747
QLD	1402.912	2104.089	6261.171	30771.550	123136.808	348330.109	766735.362	1035607.185	1284399.867	1477160.876	1676699.605	1983375.713
SA	2288.091	3946.314	8826.973	21259.670	54494.226	218018.717	356059.266	497460.081	579995.125	646638.531	724311.269	839655.537
TAS	38.415	76.628	312.818	2035.142	5051.199	10963.607	31222.601	62724.701	81682.649	91208.965	103053.498	117640.319

Figure 3

Processing and Manipulation

After the preprocessing in the original data it can be observed that there are approximately 1080 missing columns in the Air Temperature in degree C by conducting the " isna().sum()" function. This refers to the columns that had dashes in them, thus to create a complete and defined dataset imputation was undertaken,specifically mean imputation where the missing values were systematically replaced with the mean value derived from the available data points in their proximity. This provided a simplistic and compatible solution to fill the missing values. This additionally altered the dimensions of the dataset by substantially reducing the amount of columns from (1630086, 4) → (1571909, 4). This ultimately enhanced the datasets suitability for subsequent analysis and interpretation.

```
REGION                          0
Air Temperature in degrees C    1080
TOTALDEMAND                     0
RRP                             0
dtype: int64
```

```
data_mean.isna().sum()

REGION                          0
Air Temperature in degrees C    0
TOTALDEMAND                     0
RRP                             0
dtype: int64
```

The demand data was initially filtered by region and year for a 30 minute interval within each year and region in Australia. To simplify this analysis the data was processed by computing the annual average demand for each state from the years 2006 to 2020 by using the “aggfunc=’mean’” function.

```
demand_table = grouped_data.pivot_table(index='REGION', columns='Year', values='mean', aggfunc='mean')
```

Moreover within the original dataset, an additional column was there indicating the air temperature in degrees Celsius. To help provide a comprehensive analysis a separate pivot table was created by using “pd.to_numeric”, manipulating the original dataset. Furthermore the table allowed for the identification of the average temperature of each region from years 2006 to 2020 by again employing the “aggfunc=’mean’” function.

```
data_temperature['Air Temperature in degrees C'] = pd.to_numeric(data_temperature['Air Temperature in degrees C'], error
```

```
rature by region year.pivot table(index='REGION', columns='Year', values='Air Temperature in degrees C', aggfunc='mean')
```

Thus for the analysis regarding rooftop PV with average demand and average temperature there was the construction of three distinct datasets;annual average demand *(Figure 5)*, Rooftop PV, and a statewide average temperature *(Figure 4)*

Region	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
REGION													
NSW	18.246953	18.455322	17.723053	18.558216	18.312125	18.075992	17.904466	18.743419	18.579726	18.409925	19.025723	18.857370	18.565004
QLD	20.562034	20.847191	20.075246	21.076383	20.644329	20.190401	20.406871	20.855309	21.105919	20.827973	21.482089	21.440197	21.031477
SA	17.076951	17.874194	17.154960	17.889700	17.032782	17.264367	17.163411	17.872289	17.765404	17.333307	17.105567	17.531799	17.716044
TAS	12.522913	13.063172	12.612180	12.779575	12.978539	12.696416	13.130180	13.324278	13.223495	12.747412	13.526983	13.305048	13.307168
VIC	15.620664	16.612567	15.803962	16.264828	15.976743	15.957221	15.954379	15.501363	16.282964	15.748454	15.790070	15.806698	15.910506

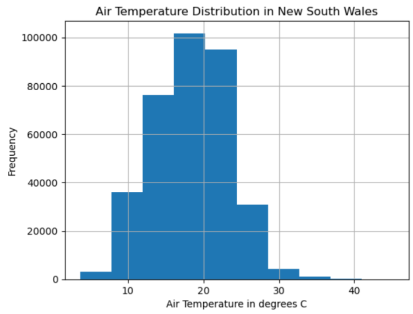
Figure 4

Year	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	201
REGION											
NSW1	8883.901099	8944.314705	8968.700636	8865.705398	8807.185635	8728.294505	8253.294400	7981.591276	7917.803791	7979.758276	7977.97716
QLD1	5829.411719	5886.078227	5940.722413	6017.281748	5973.122453	5834.093074	5826.641964	5703.642533	5745.697336	6035.463554	6271.78312
SA1	1495.004015	1524.099139	1527.397616	1538.416329	1547.245312	1494.639411	1477.396602	1426.560030	1403.301243	1398.484735	1342.53131
TAS1	1160.262642	1161.967189	1185.376847	1130.186189	1159.037467	1143.747100	1098.938829	1129.264159	1109.739691	1138.182191	1065.24045
VIC1	5845.859110	5917.086078	5956.350206	5856.753272	5842.902147	5723.975614	5657.281017	5511.769596	5324.419251	5194.532967	5082.89744

Figure 5

Exploratory Analysis

The exploratory data analysis involved determining data distribution, identifying potential relationships and visualising trends. First and foremost, we wanted to ensure the validity of the temperature and energy demand data in each state. This was done through finding the descriptive statistics and recognising that the minimum, maximum and average values made sense, for example maximum temperatures not exceeding unlikely values and similarly for energy demand. It was found that this data was also normally distributed, as shown. The main differences between each state was a slight skew, for example Queensland had a slight negative skew for temperature as Queensland generally experiences more warm weather than other states *(Figure 6 and 7)*.



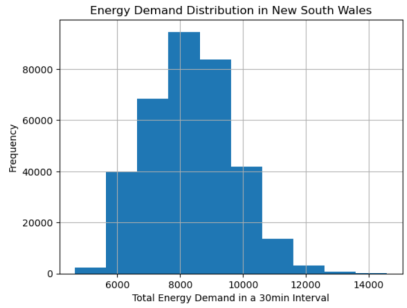


Figure 6 and 7

Following this, we looked into the relationship between the temperature and energy demand data. The correlation was found to be quite low at 0.27, as shown in the following heat map *(Figure 8)*.

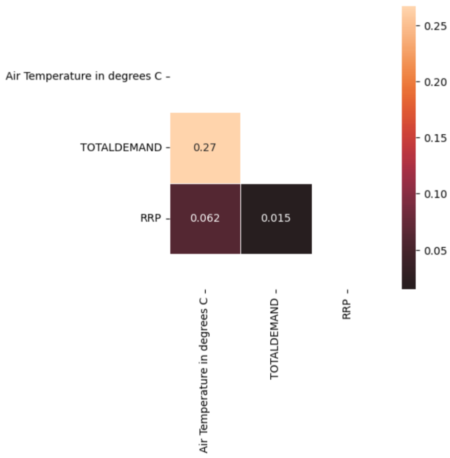


Figure 8

This was not totally unexpected, due to the fact that our data set is so large we believed it would be unlikely to find high correlations.

Then the relationship was represented in a scatterplot *(Figure 9).*

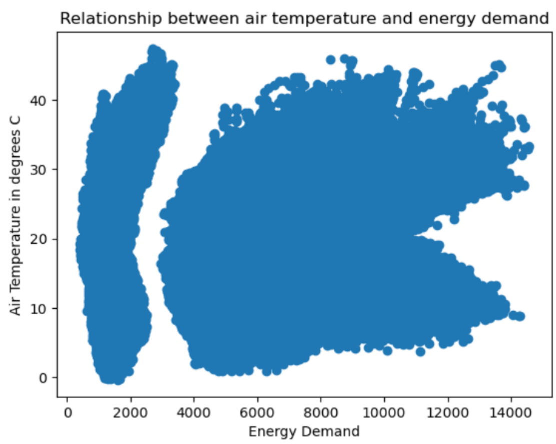


Figure 9

In this scatterplot, a separation in the data points can be recognised. To further investigate these outliers another plot was produced *(Figure 10)*, colour coding the points by state.

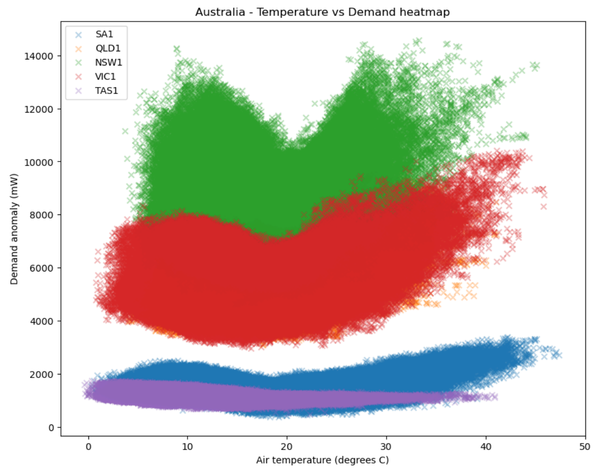


Figure 10

This plot was able to explain the separation in data points, by highlighting that the outliers were from Tasmania and South Australia. It was predicted that this could be due to these two states having the lowest population out of the five being analysed, as shown below:

```
regions = ['SA1', 'QLD1', 'NSW1', 'VIC1', 'TAS1']
populations = [1844600, 5418500, 8294000, 6766600, 575700]
```

Hence, as temperature increases, the total energy demand for Tasmania and South Australia is unlikely to experience as large of an increase due to the lower numbers of citizens that can increase the total demand.

To further visualise the relationship between temperature and energy demand, an additional plot was produced, shown below *(Figure 11)*. This demonstrates that total energy demand increases as air temperature reaches maximums and minimums, with the temperature at which demand does not spike much being around 13 degrees celsius.

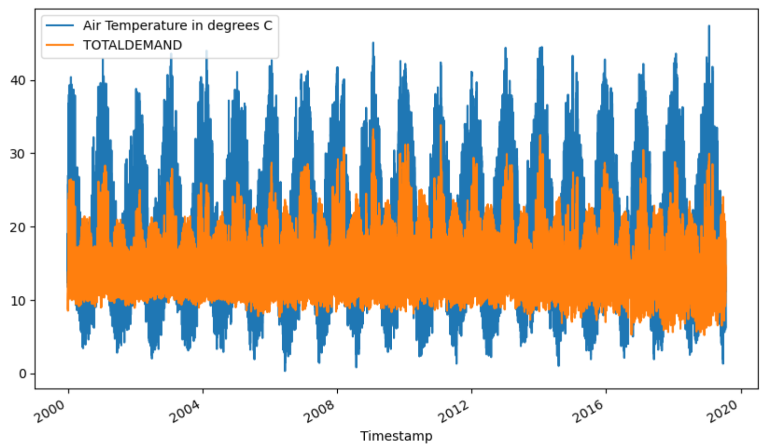


Figure 11

It can be seen that very high temperatures lead to higher peak demands than very low temperatures.

It was also found that past 2010, the graph of demand against temperature over a single day usually has a duck curve, with demand values decreasing significantly around midday even if temperatures are spiking around these times (*Figure 12)*.

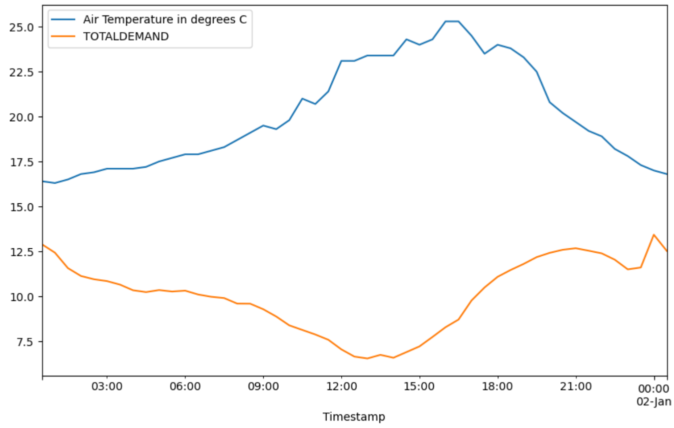


Figure 12

This graph is for the 1st of January, 2018, and showcases the demand drop during midday. Demand is scaled down by a factor of 100 in order to fit in the same graph as temperature.

The trend between region and energy demand was also investigated and a plot over a 7 day period was produced to represent this *(Figure 13)*. It was found that each state follows a similar trend.

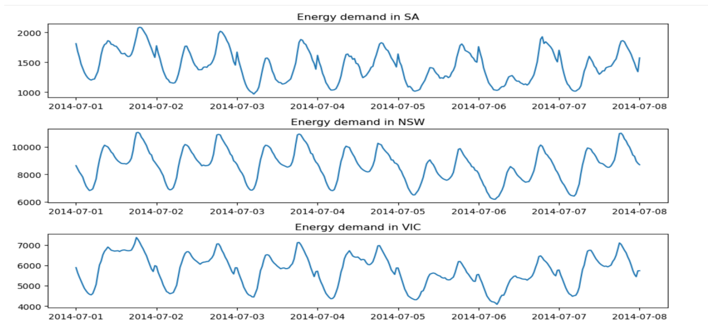


Figure 13

Overall, it was determined that there is most definitely a trend and non-linear relationship that can be identified between temperature and energy demand.

Time Series

By utilising the transpose() on the datasets   function where after even conducting the plot.line function these different time series plots were constructed that underlines the correlation between two separate variables.

We also investigated the change over time, over the course of 14 years (2006- 2020), in a number of variables given to us from the dataset. The most notable features included plotting the change in demand for energy (measured in MW). This was done by calculating the average of the total demand for energy and by adding columns to the dataframe to account for Time of Day, Time of Week and Month in a year. From this, we could see the general trend over a different period of time frames. For example, from plotting the overall mean demand (accumulating data from 2006-2020) over the months in a year, we could see that demand was the highest during February and July periods, as during this time, temperatures are likely to be at their most extremes as seen in the figure below *(Figure 14)*.

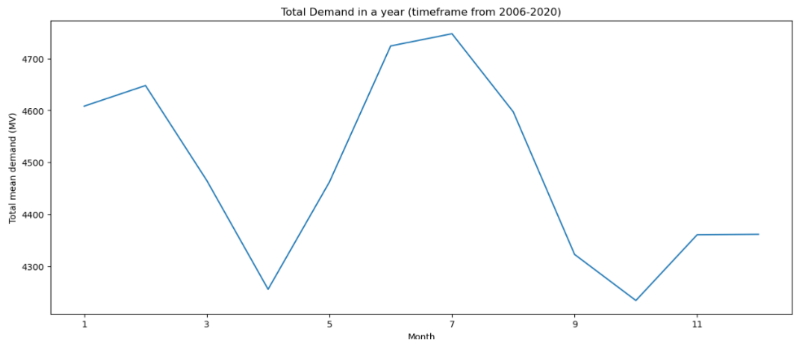


Figure 14

This is because, this is during peak summer and winter seasons in the year, and people are likely to be using more energy during this time. Further, the general trend observed from the different states over the course of the years from 2006 to 2020 in energy demand has overall remained relatively constant, with only slight decline from NSW, VIC and QLD. Demand seen from Tasmania and South Australia are significantly lower compared to other states.
We also saw that demand during the weekdays were higher compared to the average demand on a weekend as seen in the figure below *(Figure 15)*.

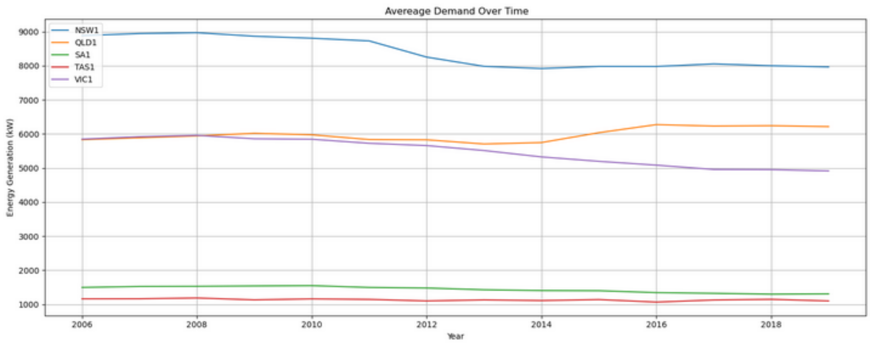


Figure 15

We also saw that demand during the weekdays were higher compared to the average demand on a weekend.

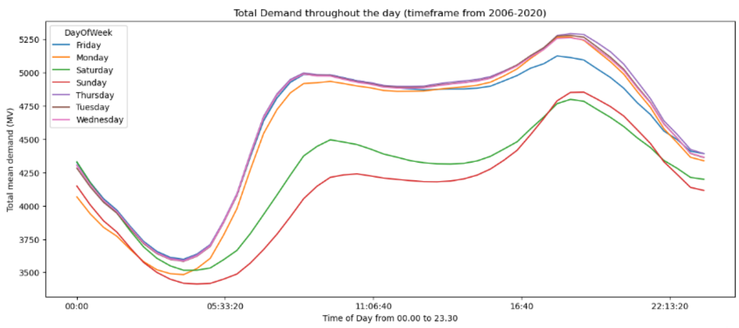


Figure 16

The figure above shows these findings *(Figure 16)*. From this, we can also see that demand is relatively lower past the 5am period in the day. This makes sense as we can determine that energy is not likely to be used heavily at this time of day (during nighttime) as people are not going to be active. During midday, we saw the highest demand for energy, typically from around 6am to 6pm time. The graph also depicts a duck curve which refers to a low energy demand during midday as a result of higher solar energy production.

Further, looking at the figure below we see the overall trend over a timeframe of relevant years (excluding 2020 as it only generated data for one month). Over the last 10 years we see an increase in usage of solar energy, and more so in recent years. This explains the more pronounced duck curve that we see. Comparing the earlier years to 2019 for example, we can see that the energy demand has decreased significantly during midday *(Figure 17)*.

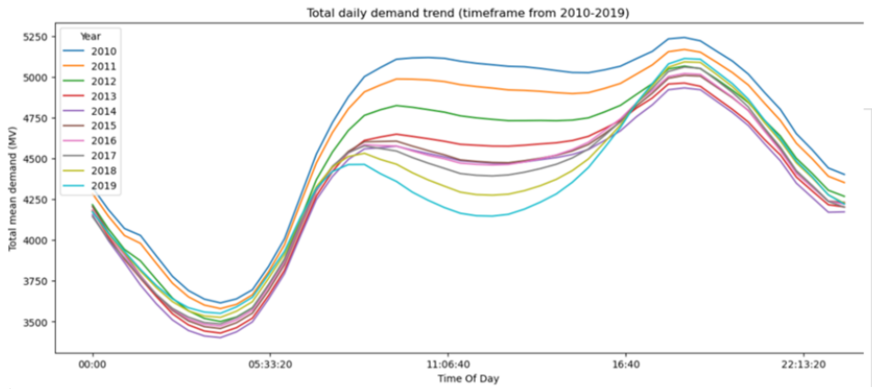


Figure 17

Similarly by conducting a time series analysis *(Figure 18)* to the Rooftop PV dataset a trend emerges, where there is a pronounced positive correlation. This correlation is notable for all regions, hence denoting a surge in rooftop PV adoption by the states.
Notably this positive trend is the result of the broader urbanisation that Australia is undergoing. As the nation undergoes a notable shift towards this, there is also an amplification in the demand for solar panels and various other renewable energy sources. This multifaceted relationship highlights the dynamic nature of this specific energy generation and its patterns. Therefore further analysis and modelling is required to understand the correlation with variables like temperature and demand.

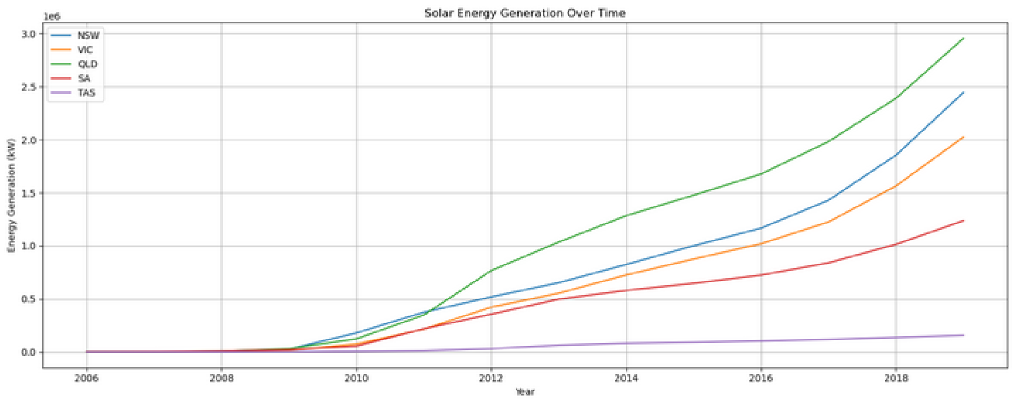


Figure 18

Modelling

The linear and polynomial regression analyses *(Figure 19 and 20)* on all three datasets between the dependent variables: Rooftop PV, Average Demand and Average Temperature. The independent variable chosen was the time spanning from 2006 to 2019. These modes were intended to make predictions regarding future values for Rooftop PV, Average Demand and Average Temperature.

The observed trends across all states is that Rooftop PV and Average Temperature exhibited a positive correlation. This could be attributed to the effects of increased urbanisation and global warming. However it was observed that Average Demand exhibited a decreasing trend. Potential explanations included advancements in energy efficiency , a shift towards renewable energy sources or economic downturns.

The main question addressed was the correlation between Rooftop PV and the other variables; Average Demand and Average Temperature. To explore these relationships, a correlation heatmap *(Figure 22)* was generated using the Seaborn library, employing the 'corrwith' and 'sns.heatmap' functions. The results indicated a strong positive correlation between Average Temperature and Rooftop PV, with the strongest correlation occurring in 2008. Conversely, the analysis suggested a minimal relationship between Rooftop PV and Average Demand. This outcome led to the speculation that small-scale PV systems primarily catered to self-consumption rather than contributing significantly to overall energy demand or being influenced by economic factors.

Subsequently, a polynomial regression and modelling exercise were conducted to provide a more flexible representation of the data, accommodating non-linear relationships. A polynomial degree of 5 was chosen, and the sklearn library was utilised, specifically the "PolynomialFeatures" and "poly.fit_transform" functions for all regions. This analysis reinforced the positive correlation in Rooftop PV and Average Temperature, while also highlighting the decreasing trend in Average Demand.

To ensure the accuracy and utility of these models, six different tests were performed across all regions, and the results were averaged. The tests included calculating R-squared (R2) *(Figure 23)* value, a statistical measure assessing the goodness of fit. The values obtained for both linear and polynomial models were notably high, implying that these models are effective in predicting future values. This aligned with the goodness of fit observed in the Average Temperature models. While the R2 value for Average Demand was not as high, it still indicated a reasonable level of accuracy.

However, the models began to reveal issues within the dataset when Random Forest Mean Absolute Error (MAE) and Mean Squared Error (MSE) tests were conducted *(Figure 23)*. The models exhibited exceptionally high MSE and MAE values for Average Demand and Rooftop PV. These results indicated poor model performance, likely due to high variance, the presence of outliers, or potential overfitting. This was corroborated when variance was calculated using the 'np.var' function and averaging across all states. Both Rooftop PV and Average Demand displayed high variance, suggesting overfitting and a lack of consistency in the data. In contrast, the models for Average Temperature indicated accurate predictions with minimal errors. Moreover, the models demonstrated consistency and lower sensitivity to outliers.

In conclusion, the models were found to be generally inaccurate. To enhance their accuracy, a deeper investigation of outliers and data preprocessing is recommended. Additionally, evaluating alternative algorithms or model types and addressing data quality issues that may be affecting predictions are vital steps in improving the overall performance and reliability of these models.

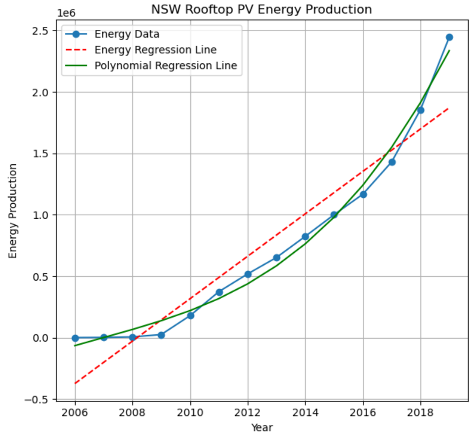


Figure 19

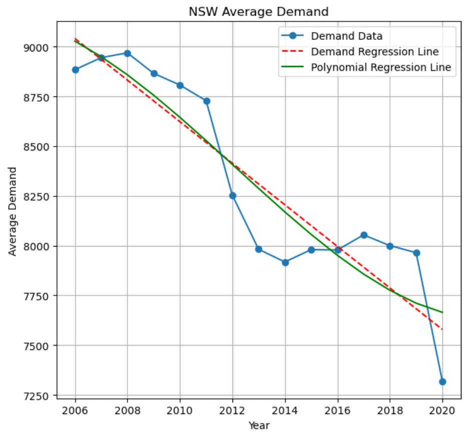


Figure 20

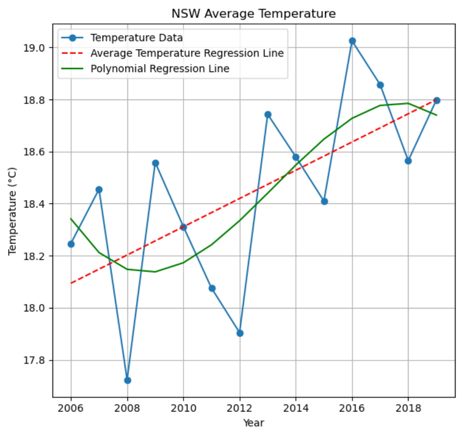


Figure 21

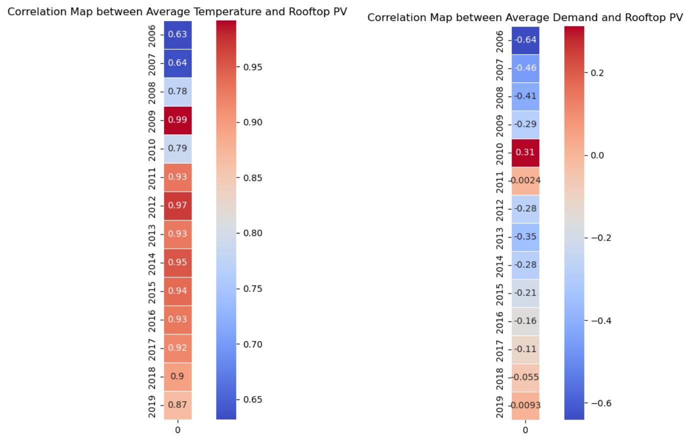
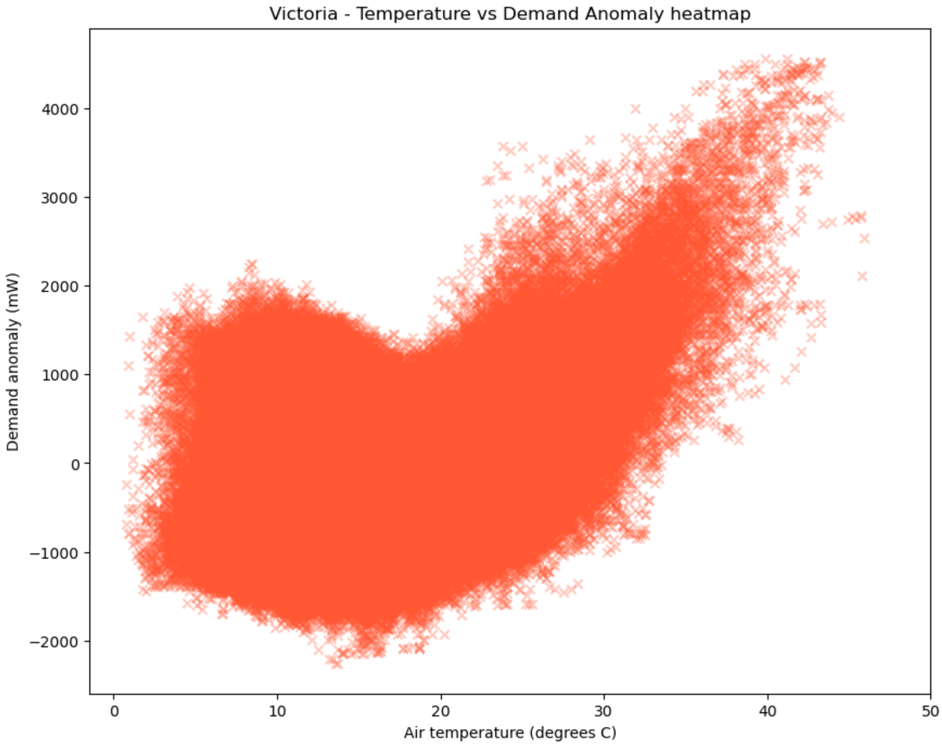


Figure 22

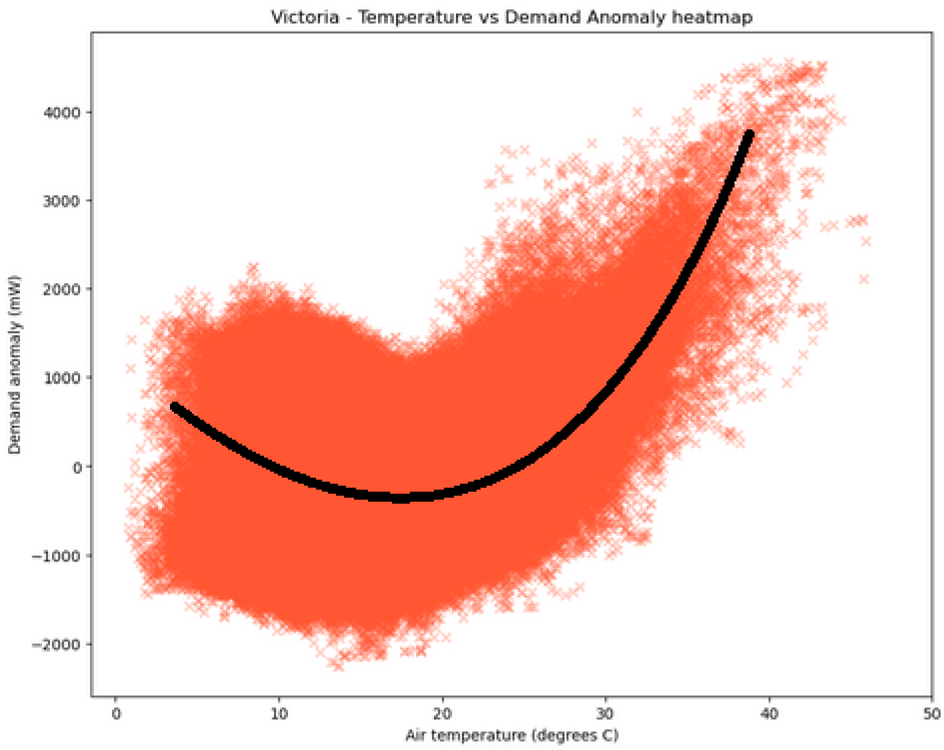
Variable	Linear R2	Polynomial R2	Random Forest MAE	Random Forest MSE	Variance
Rooftop Pv	0.9236	0.9886	44812.3874	4.3877*10^9	3.99073*10^11
Average Temperature	0.7952	0.8604	0.2774	0.1236	0.121661
Average demand	0.6796	0.7952	58.1704	5679.29	75039.8

Figure 23

Next, in order to model the relationship between air temperature and total energy demand, we first wanted to obtain an idea of what this relationship would appear as, and this could be done by plotting all of the available data for temperature against energy demand. One of our plots, this being from the data for Victoria, is shown below:



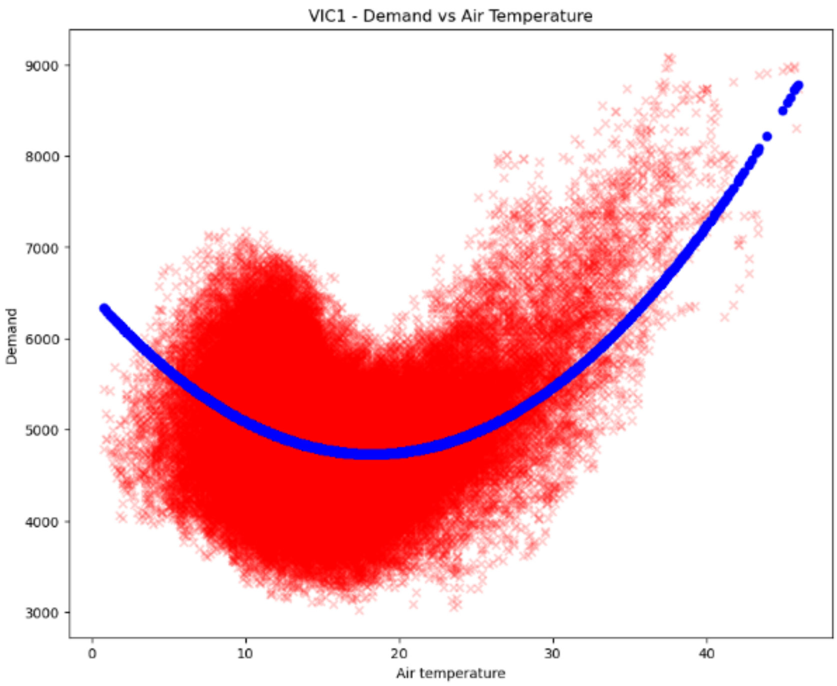
Looking at these, we concluded that the overall relationship between energy and demand would appear as a polynomial, such as this example line of best fit shown below for the Victoria plot:



Therefore, using this as a basis, we concluded that a non-linear model would be suitable for the data and began modelling sections of it via support vector machine regression, making sure to use non-linear regression methods such as polynomial modelling in order to fit our assumptions from the scatterplots above.

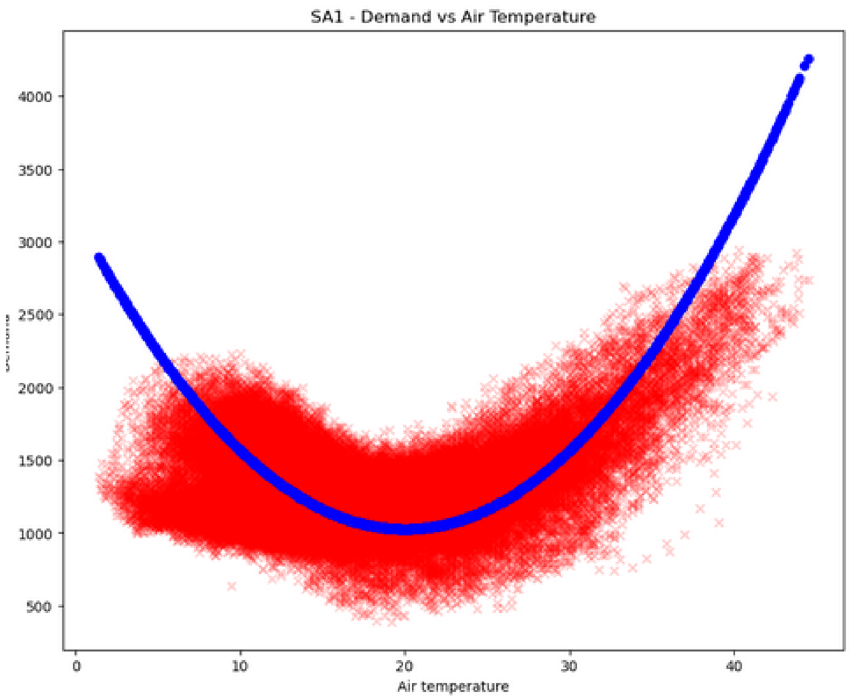
Due to the size of the dataset, the training data needed to be substantially decreased in order for models to be created in reasonable times, with subsets such as parts of the data corresponding to weekends, to different seasons, and to different times of day being used in order to both cut down on the amount of training data being used and to reduce any confounding variation due to how both temperature and demand can vary depending on the time periods being analysed.

Some of the notable models we produced are shown below:

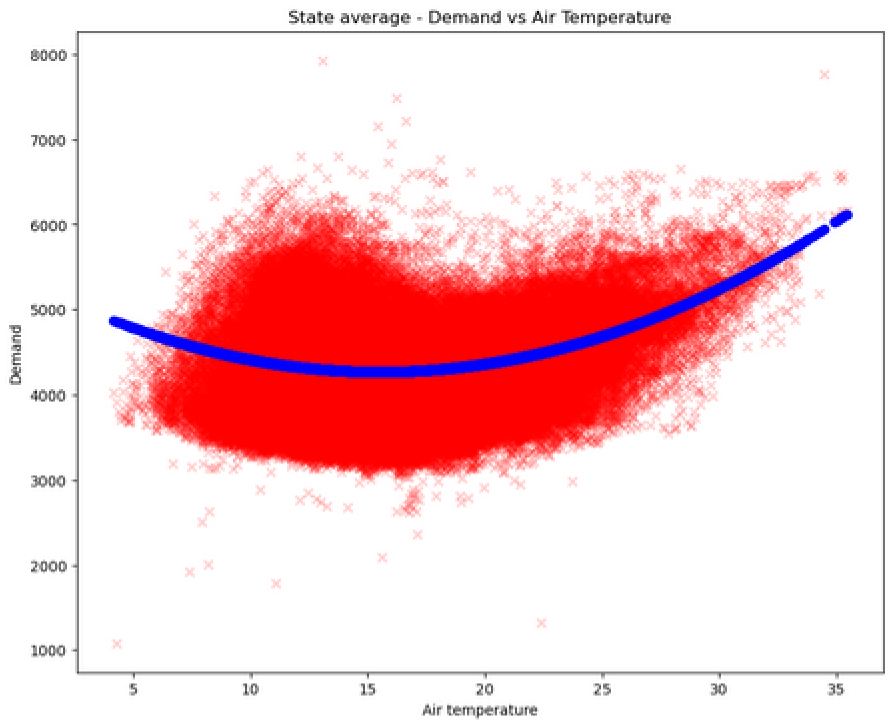



```
RMSE is 630.538
MAE is 507.391
R2 score is 0.120
```

(trained from 2000 to 2002)



R2 score is -0.312

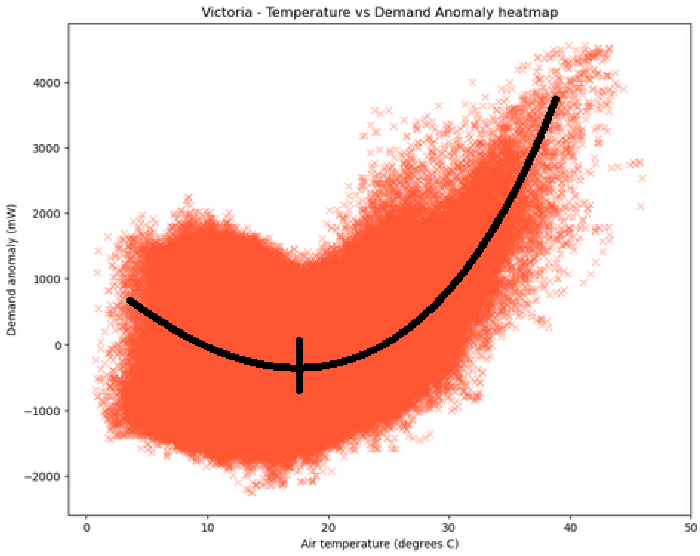


R2 score is 0.051

(Most of these models were filtered to only use weekend data and were trained on data from 2000 to 2002, then tested from 2018 to 2020.)

From our graphs, we concluded that the point of lowest average demand was seen to be around 15 to 20 degrees celsius, which corresponds to room temperature. Furthermore, there was supporting evidence to show that high temperatures will lead to higher demand increases than low temperatures, but a temperature change in either direction of the point of lowest demand will still increase overall energy demand.

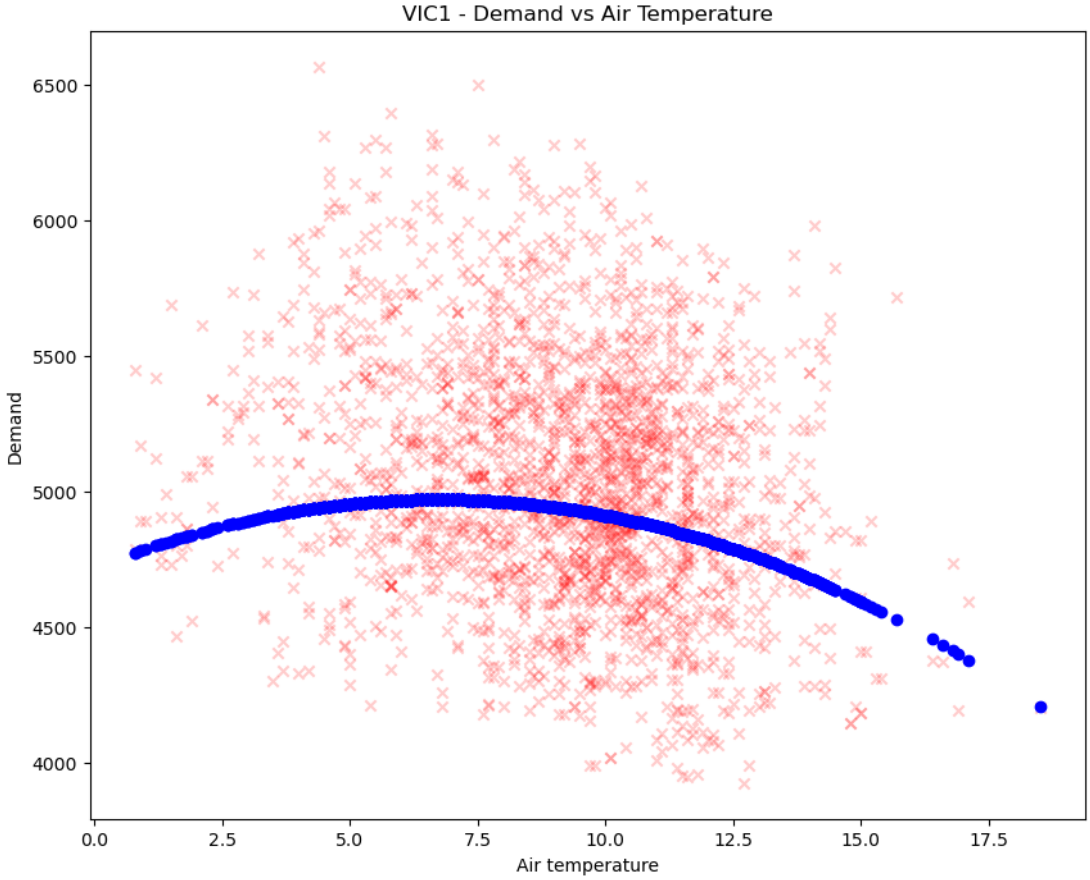
The primary model used was an SVM regression model with a degree 2 polynomial kernel - a model that aims to fit a quadratic graph onto given data points with the aim of minimising error between the graph and each value - with these generally appearing to fit the plotted data the best. However, kernels with a degree of at least 4 likely would have fit the plots better, as each plot was relatively asymmetric, which was unable to be properly represented by the symmetrical quadratic models. Below is an example of the asymmetry of a plot:



(notice the part of the estimated relationship to the left of the turning point increases slightly slower than the right side)

However, the relatively low R2 scores indicated that very little of the variance in the data could be explained by the model, and these scores persisted even with extremely specific subsets, such as with this model that used data filtered to weekends, winter, and the period between 7am and 9am:

```
RMSE is 470.809
MAE is 373.272
R2 score is -0.067
```



This variation was attributed to the overall unpredictability of energy demand per day, with potential confounding variables associated with factors such as work habits, the proportion of people who stay at home, and energy use for non-heating/cooling purposes likely being the cause of this variance. Further investigation is likely required to properly explain this unpredictability, as well as to allow the relationship between demand and temperature to be modelled with high accuracy.

Ultimately, our modelling for any trends in air temperature, energy demand, and photovoltaic energy generation showed that both air temperature and PV energy generation were increasing as of late while average demand is steadily dropping - a trend that may be a result of increased average solar energy generation in recent years. Despite this, our modelling for the relationship between average temperature and energy demand was not able to accurately model a clear relation due to confounding variation in the data, but we were still able to create polynomial models that roughly fit the estimated shape of the relationship - this being a polynomial that increases as temperature both reaches high and low values, with the point of lowest energy demand falling at approximately room temperature.

Conclusions

In conclusion, this project has provided valuable insights into the relationship between temperature and energy demand. Our two primary goals - these being to establish and model a relationship between temperature and demand both overall and over certain time periods, as well as to determine any significant trends in air temperature, energy demand, and photovoltaic energy generation - were met.

Exploratory analysis revealed that temperature and energy demand have a nonlinear relationship. Correlation analysis, scatterplots, and time series plots indicated that energy demand increases with higher and lower temperatures. The ‘duck curve’ phenomenon was observed through significant drops in demand during midday, which was attributed to increased solar energy production during this time. Furthermore, the project identified a positive correlation between rooftop PV adoption and average temperature, indicating the influence of urbanisation and global warming. However, the correlation between PV energy generation and average demand was minimal.

The modelling phase produced linear and polynomial regression models that provided insights into future predictions. While the models exhibited good fits and reasonably accurate predictions for average temperature, they encountered challenges related to outliers and overfitting in the case of average demand and rooftop PV. Data preprocessing and alternative modelling algorithms should be considered to enhance model accuracy. Modelling the temperature demand relationship proved to be challenging due to the dataset’s size. Non-linear models, including support vector machine regression with polynomial modelling were applied to capture the nature of the relationship. However, the models encountered difficulties in explaining the variance in the data, resulting in low R-squared scores. However, we were able to derive a fit in the form of a polynomial to model this relationship.

In summary, this project has enhanced our understanding of the relationship between temperature and energy demand. The findings contribute to the determining of factors affecting demand for energy.