**Data Analysis on Power Usage and Climate Change in South-East Queensland**

**[DATA7001] Introduction to Data Science G24**

by

Shang Gao(Student ID:45762402)

Weigao He(Student ID:45411728)

Yuchen Li(Student ID:4616340)

Jiaqi Yu(Student ID:46078678)

Fuquan Zhang(Student ID:46333117)

Under the guidance of Prof. Shazia Sadiq, Dr Nan Ye



[**1. Introduction**](#_g8uelebyoy1f) **3**

[1.1 Problem domain](#_5k534e9o7cb9) 3

[1.2 Problem defination](#_1w0ieqwcv3wl) 4

[**2. Getting the data we need**](#_dwolbujxljtm) **5**

[**3. System design**](#_31c4fdenm17b) **7**

[3.1 Workflow](#_h6q4xji8dwhp) 7

[3.2 Table Structures](#_msdwb2uvqz28) 7

[**4. Implementation Design**](#_ez7hbkevh5rx) **9**

[4.1 Data Collection and Cleaning](#_4o1fc72sj2l9) 9

[4.2 Data Exploration](#_yihaoxr1g2xr) 10

[**5. Conclusion**](#_ul6ci2s6wmtq) **15**

[5.1 Conclusion](#_m0a42ue1tlmi) 15

[5.2 Files and Repository](#_h1845ruz8o3n) 17

[**6. Feedback**](#_q6u7bslrlnlf) **18**

[**Reference**](#_mvxrheqroxy2) **19**

# **1. Introduction**

## **1.1 Problem domain**

Nowadays, energy conservation and emission reduction is an important topic. When people talk about their electrical power usage, the first thing that comes into their mind is what they should do to reduce power usage to save money. Seven out of ten survey respondents (69%) think air conditioning to be the most significant contributor to electricity bills(O'Neill, 2019). When people google the question about what is the energy of the use at home, the top communed answer shows that cooling and heating take over 47% of the energy that use at home(Desjardins, 2016). So, it is evident that the frequency and time of those power usage are relevant to climate change.

Utilizing our project, we aim to find out the relationship between residential power usage and climate change. During our analysis, we hope to help people save the money to get their best power usage plan, and power companies make a better strategy on the power to generate consumption.

## **1.2 Problem defination**

Founded in 1922 in Brisbane, QLD, Energex is an Australian electric power distribution company owned by the Government of Queensland. The company provides power distribution services to more than 1.4 million homes and businesses, and by May 2020, it will give elemental power to approximately 3.4 million people.(Who we are - Energex, 2020). For the Energex manger, find the relationship between residential power usage and climate change can improve the electric company to make a better strategy on the power to generate consumption. On the other hand, the residential user can choose the apparat bill plan to save money during the relationship.

In this project, we focus on residential power usage and climate change in south-east Queensland; this area is where we are living now.

# **2. Getting the data we need**

In order to find the relationship between the power usage and the climate change, we first need to get the data!

For the energy data, we use the data that provide from the Energex website.Energex is a subsidiary of Energy Queensland Limited, a company owned by the state government. Energex has established, operated, and maintained distribution networks in growth areas in southeast Queensland (including major urban areas such as Brisbane, Gold Coast, Sunshine Coast, Ipswich, etc.).(Who we are - Energex, 2020).

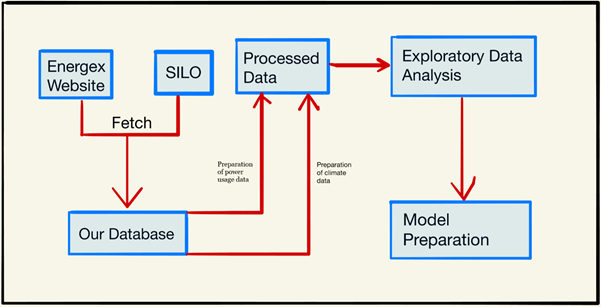
The data from the website contains the power usage of the business energy and residential energy since the business energy is more relevant for the purposes of the owner company. We take the residents' energy as the data set for our energy part. The data include the monthly residential energy usage sort by the area postcode over ten years, which from 2009 to 2018.

For the climate data, we find them in longpaddock.qld.gov.au/silo/. SILO is a database of Australian climate data from 1889 to the present. It is hosted by the Queensland Department of Environment and Science(SILO | LongPaddock | Queensland Government, 2020). For the data on the website, it comes with all kinds of daily climate records like the highest temp and the lowest temp or the rainfall records from the weather stations from all over Austria. All the data on the website are freely available to the public.

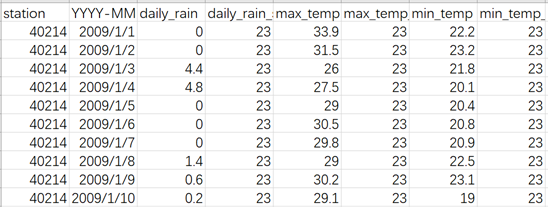
In order to connect the climate data with the power usage, we arrange the data of climate from weather station sort by the station belonging postcode and take one station from each postcode. Then we get the climate data from 101 weather stations in month degree. After that, we divided those weather stations into 7 areas to do the analysis.

# **3. System design**

## **3.1 Workflow**

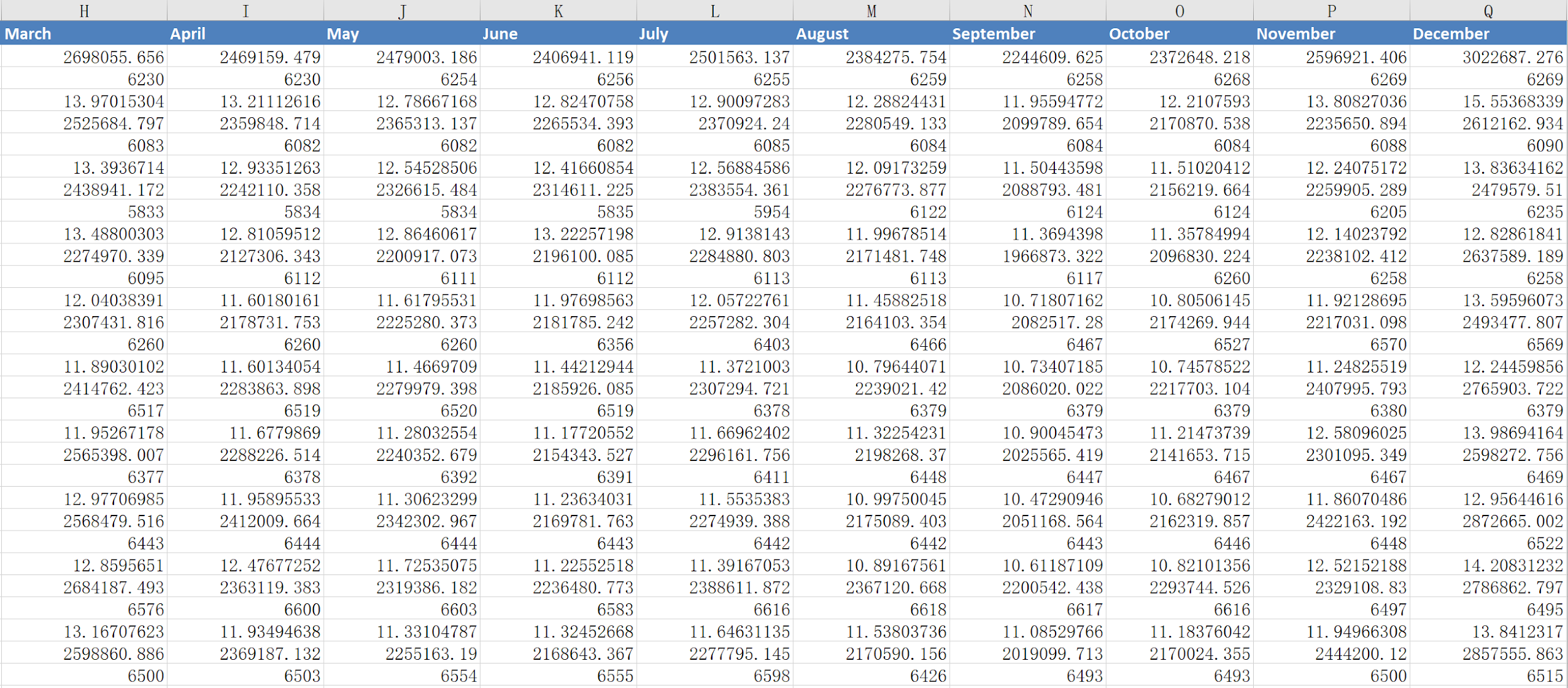
****

## **3.2** **Table Structures**

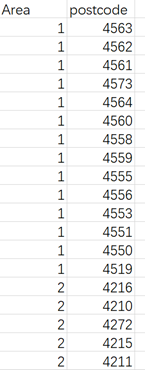
****

climate.csv





Power usage.csv

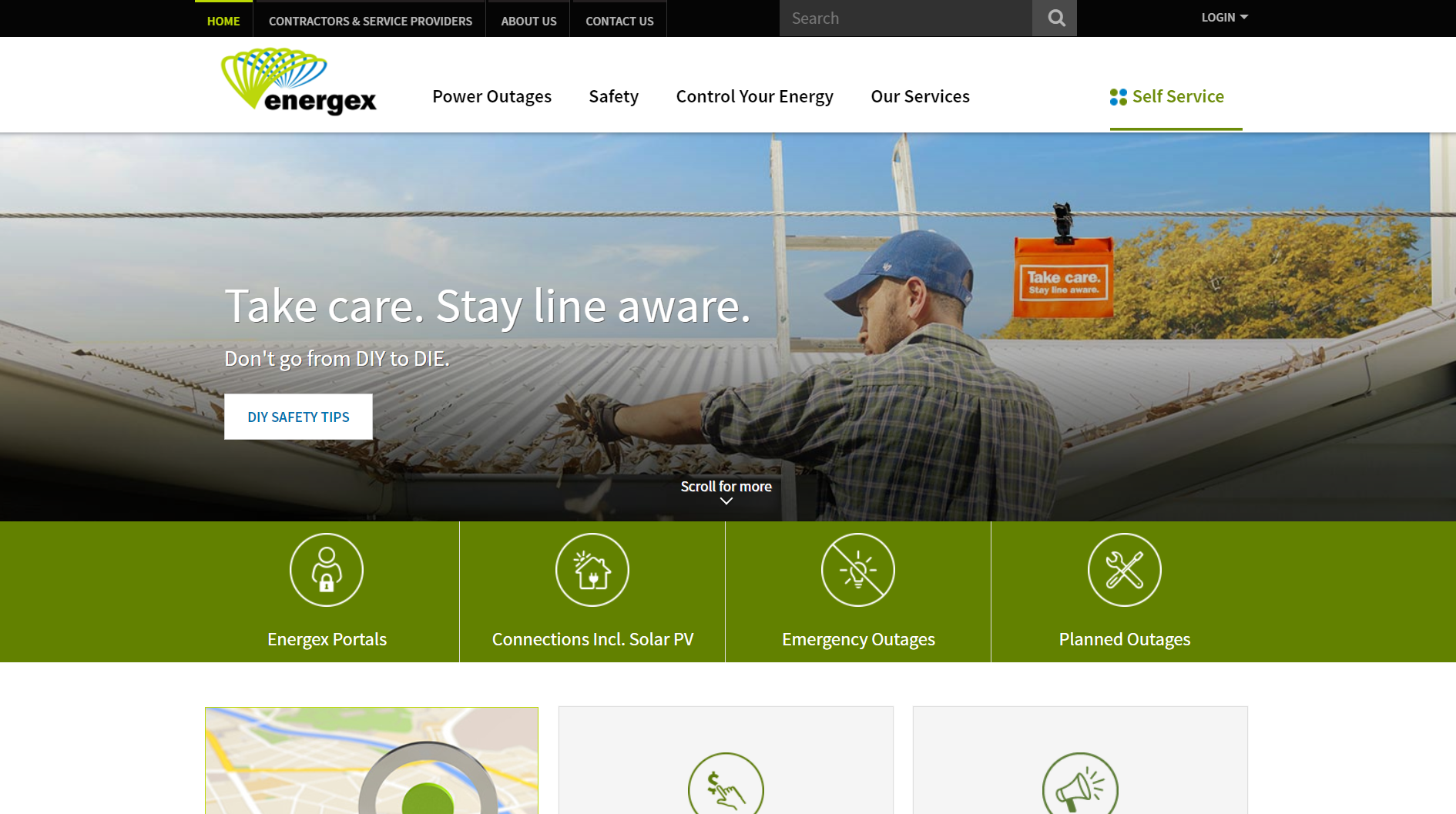


The Parts.csv

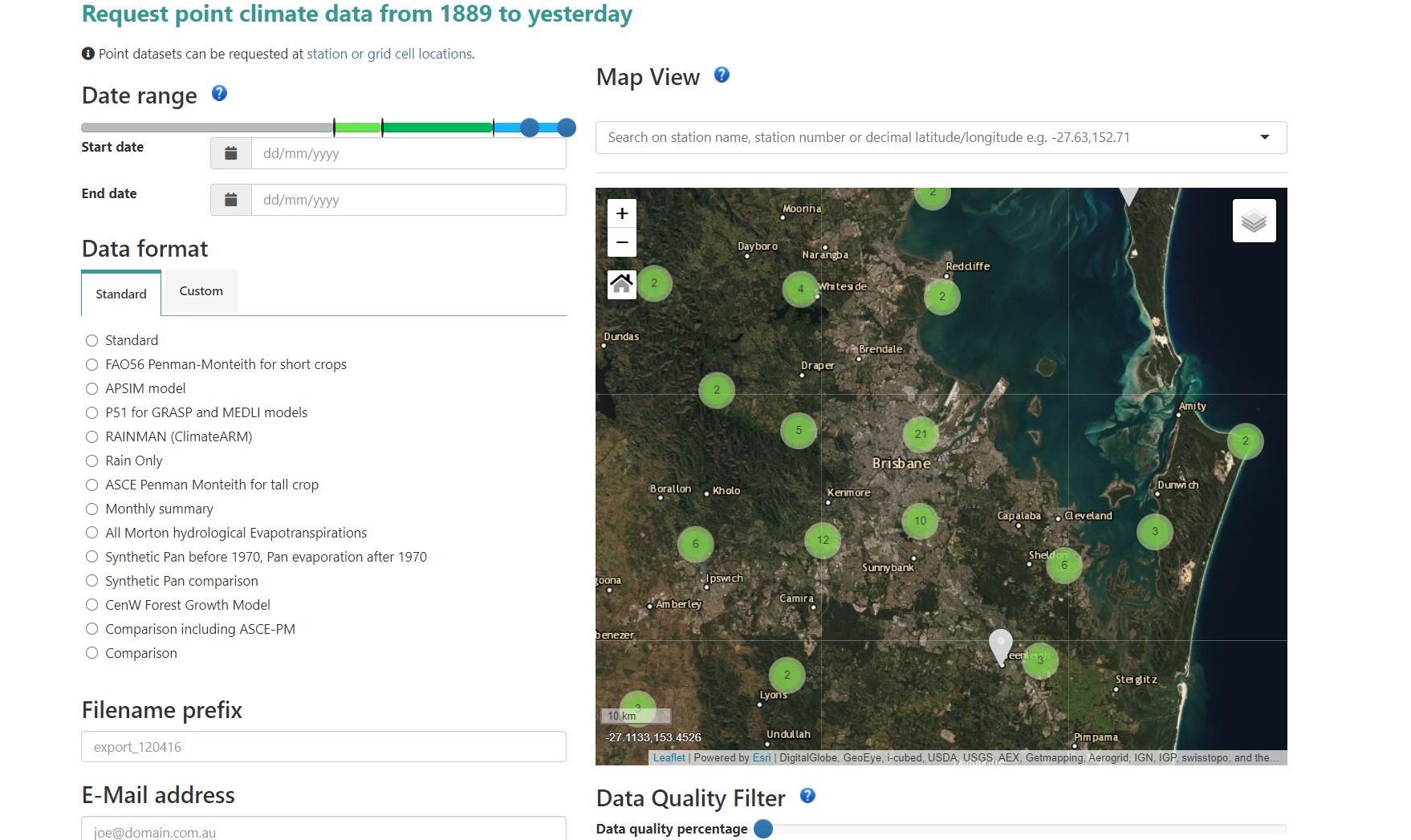
# **4.** **Implementation Design**

## **4.1 Data Collection and Cleaning**

Our data collection starts from downloading electricity usage data from the website <https://www.energex.com.au/home>, which provides the electricity usage each month of different postcode, and divided by business usage and residential usage. So for cleaning the data, we need to exact the postcode we want.



For climate data, we download it from <https://www.longpaddock.qld.gov.au/silo/point-data/>. Due to the data in this website is showed as each day, so we need to calculate the average temperature of a month for each postcode.

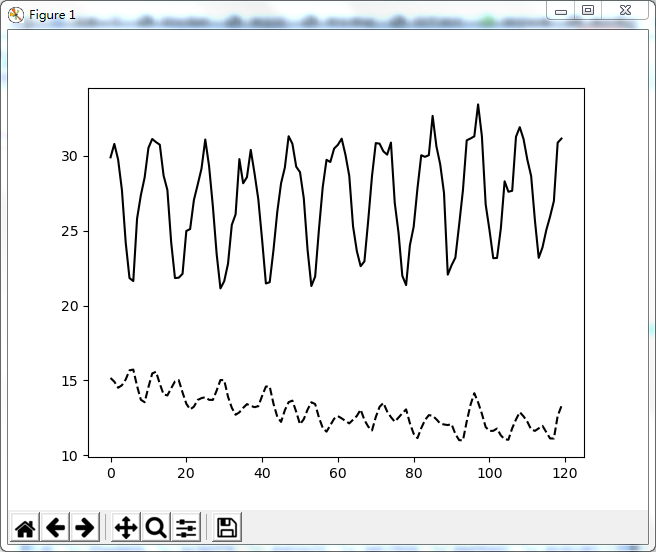


## **4.2 Data Exploration**

The primary goal of our project is to identify the degree to which temperature affects electricity consumption. To achieve this we decided to use Python, and some third-party libraries like numpy, pandas, sklearn and matplotlib.

To begin with, we imported data from 3 aspects of csv files (temperature, electricity consumption, area) and transformed them into 3 dictionaries of same keys, postcodes.

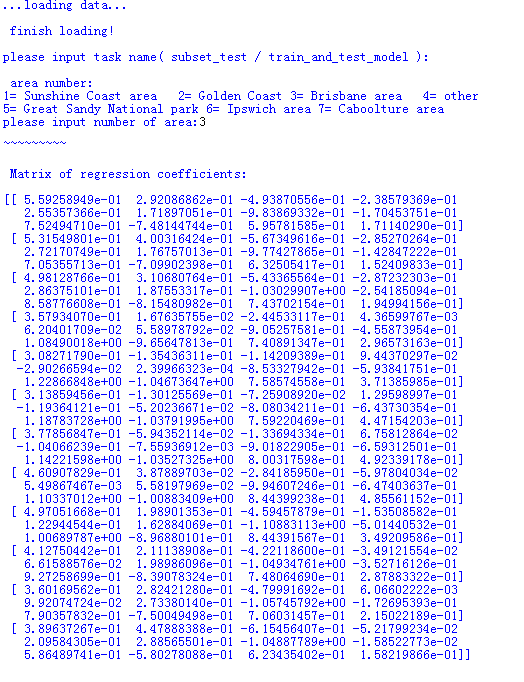
The imported data can be observed by our subset - observing module,the plotted image looks like this:



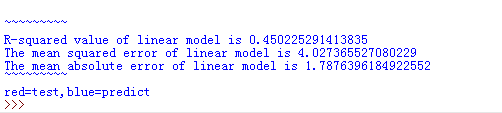
This is the plot of a subset (postcode=4066, area number=3), The solid line represents the temperature curve, while the other line represents the electricity consumption. It clearly illustrated that both temperature and electricity consumption have significant seasonal trend.

Obviously, in the hottest and coldest time of each year, residential electricity consumption will reasonably increase because of people’s usage of air-conditioner. However， how much energy consumption it exactly increased can not be directly infered by observing the temperature changes.What we need to do is to construct a Regression model.

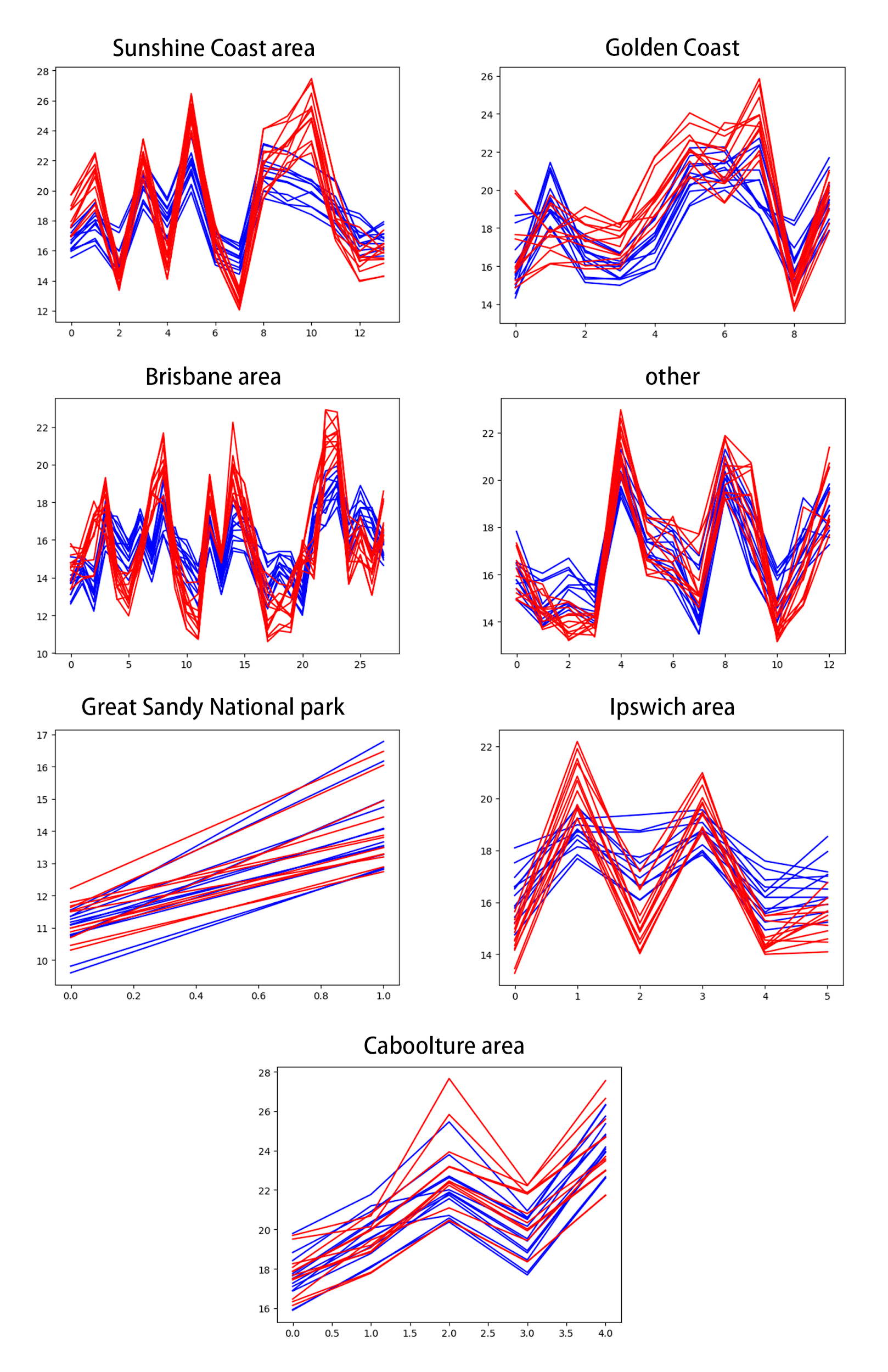
After selected the records from each area and divided them into train set and test set, we built (12 inputs/12 outputs)Linear Regression model based on the train set and fit the model into test set.



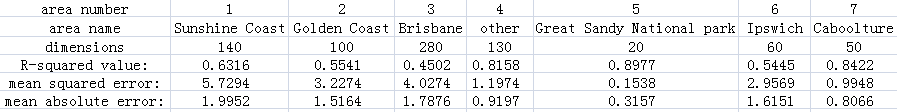
And to analyze the performance of our model, we calculated the R2 value, MSE value, and MAE value , then drew the plot of predicted value/ test value to show the comparison.



Therefore, we examine all 7 zones(manually divided by the geographical positions).As we can see from the plotted images below , the red line is test value, the blue is predicted value. The 12 different lines are corresponding to each month in a single year and the X-axis represents the different postcode sample in the selected city/area.



and get the following table of model performance:



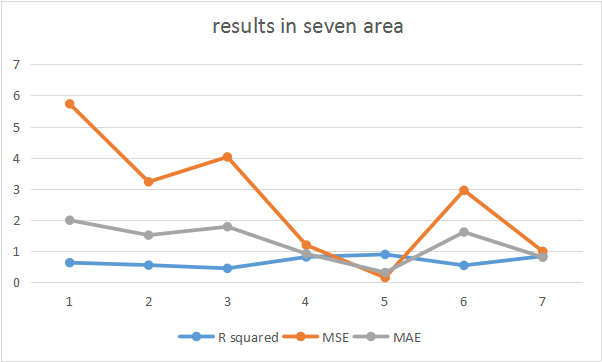
The R-squared value varies from 0.45 to 0.89, generally much higher than it was when we use samples without dividing into groups on the presentation (which value was 0.29).

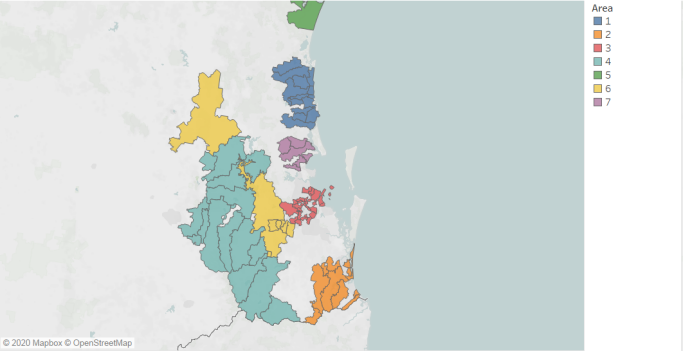
# **5.** **Conclusion**

## **5.1 Conclusion**

The original climate and energy consumption dataset of Queensland appears to be rich.In our data-preprocessing step,we clean data by extracting temperature data from climate dataset during past decade downloaded from silo.After that,we change”daily temperature” to “monthly temperature”(which means average daily temperature in this month).Next,we try to fit our data into multiple linear regression,using monthly temperature in January as X1,monthly temperature in February as X2......and electricity usage as response variable to build our prediction usage model.

To avoid the influence of region,we test this model in seven areas respectively,and then we calculate R squared value,MSE value and MAE value to evaluate fitness of this model.Results show like this:





From above discussion,we find that in area 5(Great Sandy National park),this multiple linear regression model fits best.R squared value is 0.89,which means in Great Sandy National park, 89% of electricity consumption can be attributed to changes in temperature our model.It performs also great in area 7 and area 4(0.84 and 0.82 respectively)

On the other hand,the model performs worst in Brisbane area(R squared value is 0.45).This means that apart from temperature,other factors have greater influence on energy usage.This may be determined by Brisbane's urban development.Brisbane is the most populous and most developed city in Queensland,thus there are more ways for residents to use electricity. Electricity consumption is not only affected by temperature, but also may be influenced more by other complicated factors such as living standard.

## **5.2 Files and Repository**

The repository has multiple files and folders as mentioned below.

1.Code

2.EnergyConsumption.csv

3.area.csv

4.4000.csv-4581.csv(climate data by zipcode downloaded from silo)

5.climate data of every station in certain area

# **6.** **Feedback**

In this case we mainly focus on relationship between energy usage and temperature.After previous presentation,we made some changes.We subdivide Queensland into seven regions according to latitude and longitude,performing it on different areas instead of entire Queensland to improve accuracy of our prediction.

However, we still have some technical problems in the process of data exploration.We need to improve the quality of our code, because some code is not very readable and may be too complex. Secondly, multiple linear regression is not necessarily the best choice, because the fitting in some areas is not good enough, which means that there may not be a strict linear relationship between temperature and power consumption in these areas.

Generally,we are very pleased to be able to cooperate on this project.In future studies, we will gradually improve this model and try to explore the impact of other factors such as geographic location on power consumption

# **Reference**

1. O'Neill, B., 2019. *How Much Electricity Does My Air Conditioner Use?*. [online] Canstar Blue. Available at: <https://www.canstarblue.com.au/appliances/how-much-electricity-does-aircon-use/> [Accessed 12 June 2020].

2. Desjardins, J., 2016. *What Uses The Most Energy In Your Home?*. [online] Visual Capitalist. Available at: <https://www.visualcapitalist.com/what-uses-the-most-energy-home/> [Accessed 12 June 2020].

3. Energex.com.au. 2020. *Who We Are - Energex*. [online] Available at: <https://www.energex.com.au/about-us/company-information/who-we-are> [Accessed 12 June 2020]

4. Longpaddock.qld.gov.au. 2020. SILO | Longpaddock | Queensland Government. [online] Available at: <https://www.longpaddock.qld.gov.au/silo/> [Accessed 12 June 2020].