

# Analysis on Australian Energy Household – 30minute interval

[DOCUMENT SUBTITLE]

YAOW HUI CHONG

## Introduction

This dataset is named 'Australian Energy Household – 30minute interval'. I downloaded this dataset from Kaggle.com and it records the data from a digital energy meter in Australia from a home in Sydney with 21 solar panels (6.6kw system) on the roof with a maximum inverter output of 5kw. It is a 4-person household with a 13KW reverse-cycle ducted air-conditioning system as the largest electrical consumer in the home. This dataset ranges from July 2019 to June 2020 in 30-minute intervals.

## Dataset Description

There are 31484 records in this dataset with 12 columns in total. The record time started at 7/11/2019 12am and it ended at 6/2/2020 11.59pm. Each row records the meter read value and outdoor temperature, separated with solar (production) or general usage (consumption).

A	B	C	D	E	F	G	H	I	J	K	L
AccountNumber	NMI	DeviceNumber	DeviceType	RegisterCode	RateTypeDescription	StartDate	EndDate	ProfileReadValue	RegisterReadValue	QualityFlag	BOMTEMP
3456789012	23456789012	123456789	COMMS4D	73538#B1	Solar	11/7/2019 0:00	11/7/2019 0:29	0	0	E	13.5
3456789012	23456789012	123456789	COMMS4D	73538#E1	Generalusage	11/7/2019 0:00	11/7/2019 0:29	0	0	E	13.5
3456789012	23456789012	123456789	COMMS4D	73538#B1	Solar	11/7/2019 0:30	11/7/2019 0:59	0	0	E	13.2
3456789012	23456789012	123456789	COMMS4D	73538#E1	Generalusage	11/7/2019 0:30	11/7/2019 0:59	0	0	E	13.2
3456789012	23456789012	123456789	COMMS4D	73538#B1	Solar	11/7/2019 1:00	11/7/2019 1:29	0	0	E	13.3
3456789012	23456789012	123456789	COMMS4D	73538#E1	Generalusage	11/7/2019 1:00	11/7/2019 1:29	0	0	E	13.3
3456789012	23456789012	123456789	COMMS4D	73538#B1	Solar	11/7/2019 1:30	11/7/2019 1:59	0	0	E	12.7
3456789012	23456789012	123456789	COMMS4D	73538#E1	Generalusage	11/7/2019 1:30	11/7/2019 1:59	0	0	E	12.7
3456789012	23456789012	123456789	COMMS4D	73538#B1	Solar	11/7/2019 2:00	11/7/2019 2:29	0	0	E	13
3456789012	23456789012	123456789	COMMS4D	73538#E1	Generalusage	11/7/2019 2:00	11/7/2019 2:29	0	0	E	13
3456789012	23456789012	123456789	COMMS4D	73538#B1	Solar	11/7/2019 2:30	11/7/2019 2:59	0	0	E	12.9
3456789012	23456789012	123456789	COMMS4D	73538#E1	Generalusage	11/7/2019 2:30	11/7/2019 2:59	0	0	E	12.9
3456789012	23456789012	123456789	COMMS4D	73538#B1	Solar	11/7/2019 3:00	11/7/2019 3:29	0	0	E	13.2
3456789012	23456789012	123456789	COMMS4D	73538#E1	Generalusage	11/7/2019 3:00	11/7/2019 3:29	0	0	E	13.2
3456789012	23456789012	123456789	COMMS4D	73538#B1	Solar	11/7/2019 3:30	11/7/2019 3:59	0	0	E	13.2
3456789012	23456789012	123456789	COMMS4D	73538#E1	Generalusage	11/7/2019 3:30	11/7/2019 3:59	0	0	E	13.2
3456789012	23456789012	123456789	COMMS4D	73538#B1	Solar	11/7/2019 4:00	11/7/2019 4:29	0	0	E	13.2
3456789012	23456789012	123456789	COMMS4D	73538#E1	Generalusage	11/7/2019 4:00	11/7/2019 4:29	0	0	E	13.2

## Columns

1. AccountNumber: Energy retail account number
2. NMI: National measurement initiative digital meter number
3. DeviceNumber: Device number
4. DeviceType: Type of digital energy meter
5. RegisterCode: Meter register code
6. RateTypeDescription: Two-class text, either solar (production) or generalusage (consumption)
7. StartDate: Starting Recording Timestamp
8. EndDate: Ending Recording Timestamp
9. ProfileReadValue: Actual reading in KW

- 10. RegisterReadValue: Fault code values
- 11. QualityFlag: A for Actual, E for estimate
- 12. BOMTEMP: Bureau of Meterology Temperature in Sydney in Celsius

## Data Cleaning

No.	Action	Description
1	Remove AccountNumber	Only 1 unique value, not meaningful to analysis
2	Remove NMI	Only 1 unique value, not meaningful to analysis
3	Remove DeviceNumber	DeviceNumber doesn't seem useful to the analysis. It changed from 123456789 to 700373538 and has been using the same device since then.
4	Remove DeviceType	Only 1 unique value, not meaningful to analysis
5	Remove RegisterCode	RegisterCode doesn't seem useful to the analysis. '73538#B1' represents Solar and '73538#E1' represents Generalusage.
6	Remove RegisterReadValue	Only 1 unique value, not meaningful to analysis
7	Check null value and anomaly data	No null value or anomaly data found in rows.

## Dataset after data cleaning

A	B	C	D	E	F
RateTypeDescription ▼	StartDate ▼	EndDate ▼	ProfileReadValue ▼	QualityFlag ▼	BOMTEMP ▼
Solar	11/7/2019 0:00	11/7/2019 0:29	0 E		13.5
Generalusage	11/7/2019 0:00	11/7/2019 0:29	0 E		13.5
Solar	11/7/2019 0:30	11/7/2019 0:59	0 E		13.2
Generalusage	11/7/2019 0:30	11/7/2019 0:59	0 E		13.2
Solar	11/7/2019 1:00	11/7/2019 1:29	0 E		13.3
Generalusage	11/7/2019 1:00	11/7/2019 1:29	0 E		13.3
Solar	11/7/2019 1:30	11/7/2019 1:59	0 E		12.7
Generalusage	11/7/2019 1:30	11/7/2019 1:59	0 E		12.7
Solar	11/7/2019 2:00	11/7/2019 2:29	0 E		13
Generalusage	11/7/2019 2:00	11/7/2019 2:29	0 E		13
Solar	11/7/2019 2:30	11/7/2019 2:59	0 E		12.9
Generalusage	11/7/2019 2:30	11/7/2019 2:59	0 E		12.9
Solar	11/7/2019 3:00	11/7/2019 3:29	0 E		13.2
Generalusage	11/7/2019 3:00	11/7/2019 3:29	0 E		13.2
Solar	11/7/2019 3:30	11/7/2019 3:59	0 E		13.2
Generalusage	11/7/2019 3:30	11/7/2019 3:59	0 E		13.2
Solar	11/7/2019 4:00	11/7/2019 4:29	0 E		13.2
Generalusage	11/7/2019 4:00	11/7/2019 4:29	0 E		13.2
Solar	11/7/2019 4:30	11/7/2019 4:59	0 E		13.4
Generalusage	11/7/2019 4:30	11/7/2019 4:59	0 E		13.4
Solar	11/7/2019 5:00	11/7/2019 5:29	0 E		13.7
Generalusage	11/7/2019 5:00	11/7/2019 5:29	0 E		13.7
Solar	11/7/2019 5:30	11/7/2019 5:59	0 E		13.2
Generalusage	11/7/2019 5:30	11/7/2019 5:59	0 E		13.2
Solar	11/7/2019 6:00	11/7/2019 6:29	0 E		12.4
Generalusage	11/7/2019 6:00	11/7/2019 6:29	0 E		12.4
Solar	11/7/2019 6:30	11/7/2019 6:59	0 E		12.9
Generalusage	11/7/2019 6:30	11/7/2019 6:59	0 E		12.9
Solar	11/7/2019 7:00	11/7/2019 7:29	0 E		13
Generalusage	11/7/2019 7:00	11/7/2019 7:29	0 E		13

## Data transformation

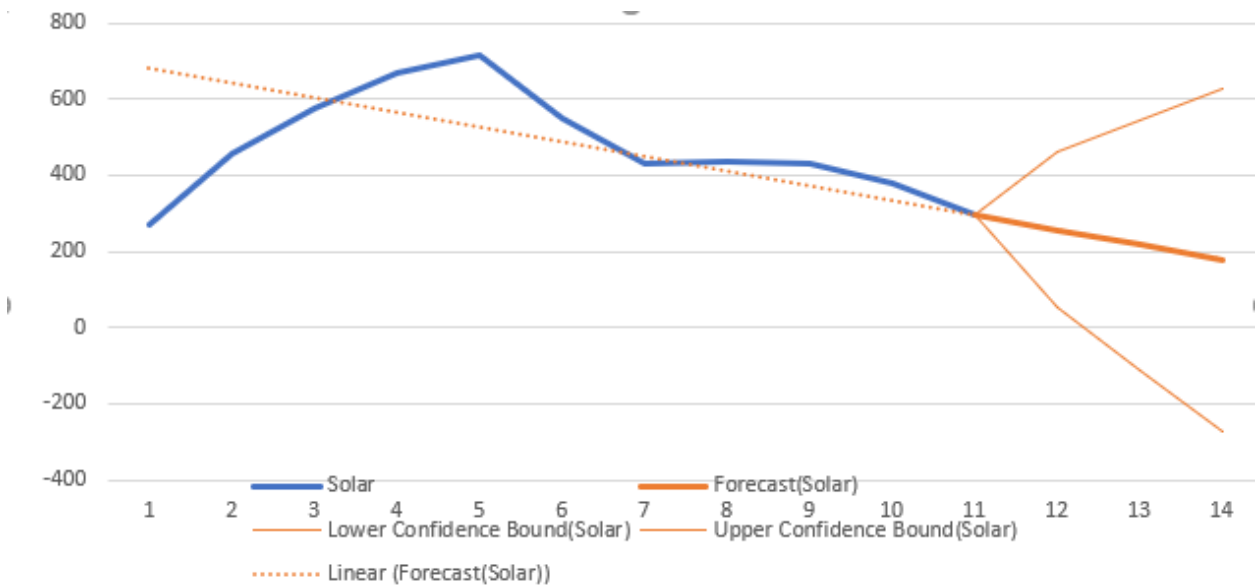
RateTypeDescription	Date	Day	MonthName	Year	StartTime	EndTime	Profile	Quality	BOMTEMP
Solar	11/7/2019	11	Jul	2019	12:00:00 AM	12:29:00 AM	0 E		13.5
Generalusage	11/7/2019	11	Jul	2019	12:00:00 AM	12:29:00 AM	0 E		13.5
Solar	11/7/2019	11	Jul	2019	12:30:00 AM	12:59:00 AM	0 E		13.2
Generalusage	11/7/2019	11	Jul	2019	12:30:00 AM	12:59:00 AM	0 E		13.2
Solar	11/7/2019	11	Jul	2019	1:00:00 AM	1:29:00 AM	0 E		13.3
Generalusage	11/7/2019	11	Jul	2019	1:00:00 AM	1:29:00 AM	0 E		13.3
Solar	11/7/2019	11	Jul	2019	1:30:00 AM	1:59:00 AM	0 E		12.7
Generalusage	11/7/2019	11	Jul	2019	1:30:00 AM	1:59:00 AM	0 E		12.7
Solar	11/7/2019	11	Jul	2019	2:00:00 AM	2:29:00 AM	0 E		13
Generalusage	11/7/2019	11	Jul	2019	2:00:00 AM	2:29:00 AM	0 E		13
Solar	11/7/2019	11	Jul	2019	2:30:00 AM	2:59:00 AM	0 E		12.9
Generalusage	11/7/2019	11	Jul	2019	2:30:00 AM	2:59:00 AM	0 E		12.9
Solar	11/7/2019	11	Jul	2019	3:00:00 AM	3:29:00 AM	0 E		13.2
Generalusage	11/7/2019	11	Jul	2019	3:00:00 AM	3:29:00 AM	0 E		13.2
Solar	11/7/2019	11	Jul	2019	3:30:00 AM	3:59:00 AM	0 E		13.2
Generalusage	11/7/2019	11	Jul	2019	3:30:00 AM	3:59:00 AM	0 E		13.2
Solar	11/7/2019	11	Jul	2019	4:00:00 AM	4:29:00 AM	0 E		13.2
Generalusage	11/7/2019	11	Jul	2019	4:00:00 AM	4:29:00 AM	0 E		13.2
Solar	11/7/2019	11	Jul	2019	4:30:00 AM	4:59:00 AM	0 E		13.4
Generalusage	11/7/2019	11	Jul	2019	4:30:00 AM	4:59:00 AM	0 E		13.4
Solar	11/7/2019	11	Jul	2019	5:00:00 AM	5:29:00 AM	0 E		13.7
Generalusage	11/7/2019	11	Jul	2019	5:00:00 AM	5:29:00 AM	0 E		13.7
Solar	11/7/2019	11	Jul	2019	5:30:00 AM	5:59:00 AM	0 E		13.2
Generalusage	11/7/2019	11	Jul	2019	5:30:00 AM	5:59:00 AM	0 E		13.2
Solar	11/7/2019	11	Jul	2019	6:00:00 AM	6:29:00 AM	0 E		12.4
Generalusage	11/7/2019	11	Jul	2019	6:00:00 AM	6:29:00 AM	0 E		12.4
Solar	11/7/2019	11	Jul	2019	6:30:00 AM	6:59:00 AM	0 E		12.9
Generalusage	11/7/2019	11	Jul	2019	6:30:00 AM	6:59:00 AM	0 E		12.9
Solar	11/7/2019	11	Jul	2019	7:00:00 AM	7:29:00 AM	0 E		13
Generalusage	11/7/2019	11	Jul	2019	7:00:00 AM	7:29:00 AM	0 E		13
Solar	11/7/2019	11	Jul	2019	7:30:00 AM	7:59:00 AM	0 E		13.6
Generalusage	11/7/2019	11	Jul	2019	7:30:00 AM	7:59:00 AM	0 E		13.6
Solar	11/7/2019	11	Jul	2019	8:00:00 AM	8:29:00 AM	0 E		14.1
Generalusage	11/7/2019	11	Jul	2019	8:00:00 AM	8:29:00 AM	0 E		14.1
Solar	11/7/2019	11	Jul	2019	8:30:00 AM	8:59:00 AM	0 E		14.9
Generalusage	11/7/2019	11	Jul	2019	8:30:00 AM	8:59:00 AM	0 E		14.9

Microsoft Power Query is used to adjust the date format, so we can do analyze the monthly read value

## Time series forecasting

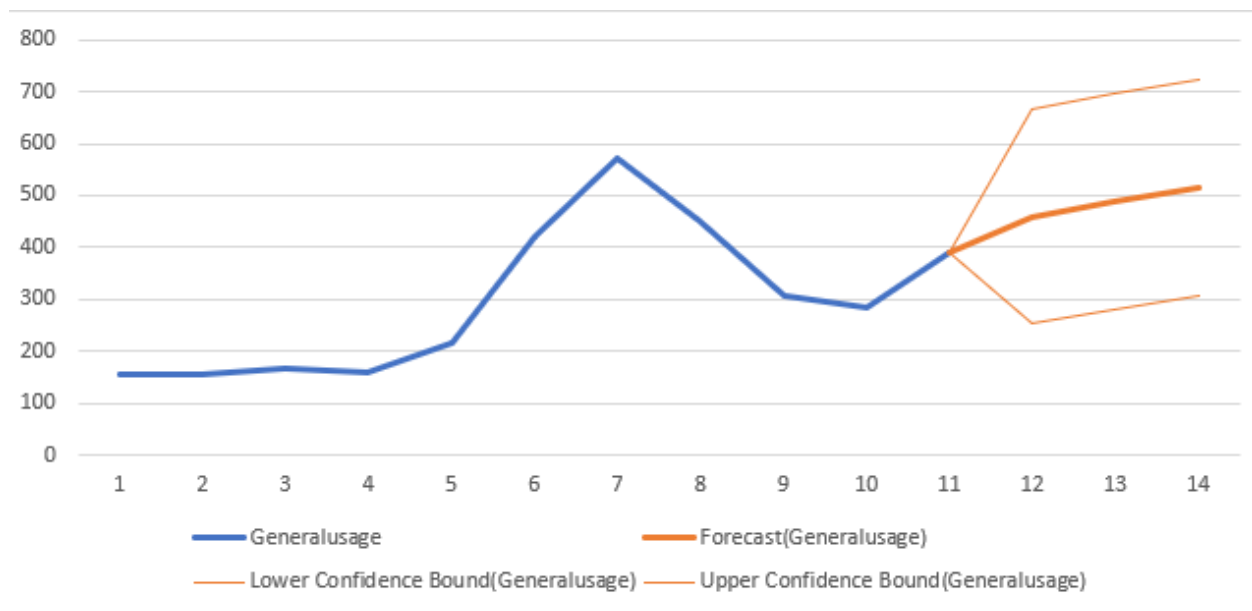
1. Solar forecasting for next 3 months. Note that  $t = 1$  indicates July,  $t = 2$  indicates August, and so on. For  $t = 12$ , it forecasts the solar production of 257kW in June 2020.

t	Solar	Forecast(Solar)	Lower Confidence Bound(Solar)	Upper Confidence Bound(Solar)	Statistic	Value
1	268.417				Alpha	1.00
2	457.252				Beta	0.25
3	574.48				Gamma	0.00
4	671.008				MASE	0.29
5	712.896				SMAPE	0.10
6	552.285				MAE	35.28
7	428.876				RMSE	38.59
8	434.612					
9	430.06					
10	377.746					
11	295.64	295.64	295.64	295.64		
12		257.0395814	51.48	462.60		
13		218.3234121	-110.25	546.89		
14		179.6072429	-270.42	629.63		



2. Generalusage forecasting for next 3 months. Note that  $t = 1$  indicates July,  $t = 2$  indicates August, and so on. For  $t = 12$ , it forecasts the solar production of 459.87kW in June 2020.

t	Generalusage	Forecast(Generalusage)	Lower Confidence Bound(Generalusage)	Upper Confidence Bound(Generalusage)	Statistic	Value
1	154.226				Alpha	0.00
2	153.745				Beta	0.00
3	165.259				Gamma	0.00
4	160.055				MASE	1.18
5	216.979				SMAPE	0.23
6	421.67				MAE	84.25
7	571.588				RMSE	89.33
8	450.994					
9	306.758					
10	283.656					
11	388.557	388.557	388.56	388.56		
12		459.874213	252.65	667.10		
13		487.6738572	280.45	694.90		
14		515.4735014	308.25	722.70		



# Linear regression analysis – Read value by average temperature

	Generalusage	Solar	Total Sum of ProfileReadValue	Total Average of BOMTEMP
Row Labels	Sum of ProfileReadValue	Sum of ProfileReadValue		
Jul	154.226	268.417	422.643	11.54702381
Aug	153.745	457.252	610.997	11.38373656
Sep	165.259	574.48	739.739	14.64743056
Oct	160.055	671.008	831.063	15.51049798
Nov	216.979	712.896	929.875	14.60854167
Dec	421.67	552.285	973.955	17.92399194
Jan	571.588	428.876	1000.464	20.36182796
Feb	450.994	434.612	885.606	21.83347701
Mar	306.758	430.06	736.818	18.82795699
Apr	283.656	377.746	661.402	17.62548611
May	388.557	295.64	684.197	14.7469086

## 1. General usage linear model

SUMMARY OUTPUT									
<i>Regression Statistics</i>									
Multiple R	0.775366439								
R Square	0.601193115								
Adjusted R Square	0.551342254								
Standard Error	95.80428094								
Observations	10								
<i>ANOVA</i>									
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>gnificance F</i>				
Regression	1	110690.7088	110690.7088	12.0598342	0.008409				
Residual	8	73427.68197	9178.460247						
Total	9	184118.3908							
<i>Coefficients</i>									
	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>
Intercept	-279.6711253	173.0281737	-1.616332874	0.14468475	-678.675	119.3326	-678.675	119.3326	
11.54702381	35.32559481	10.17229191	3.4727272	0.008408977	11.86825	58.78294	11.86825	58.78294	

## 2. Solar production linear model

SUMMARY OUTPUT									
<i>Regression Statistics</i>									
Multiple R	0.260415733								
R Square	0.067816354								
Adjusted R Square	-0.048706602								
Standard Error	134.472613								
Observations	10								
<i>ANOVA</i>									
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>gnificance F</i>				
Regression	1	10524.23729	10524.23729	0.581999945	0.467426				
Residual	8	144663.0693	18082.88366						
Total	9	155187.3065							
<i>Coefficients</i>									
	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>
Intercept	675.9027314	242.8654588	2.783033597	0.023812839	115.854	1235.951	115.854	1235.951	
11.54702381	-10.89254129	14.27801201	-0.762889209	0.467426253	-43.8177	22.03261	-43.8177	22.03261	



## Findings

Row Labels	Sum of ProfileReadValue	Average of BOMTEMP
Jan	1000.464	20.36182796
Feb	885.606	21.83347701
Mar	736.818	18.82795699
Apr	661.402	17.62548611
May	684.197	14.7469086
Jun	46.346	15.00104167
Jul	422.643	11.54702381
Aug	610.997	11.38373656
Sep	739.739	14.64743056
Oct	831.063	15.51049798
Nov	929.875	14.60854167
Dec	973.955	17.92399194
<b>Grand Total</b>	<b>8523.105</b>	<b>16.38275314</b>

Sum of ProfileReadValue	Column Labels		
Row Labels	Generalusage	Solar	Grand Total
Jan	571.588	428.876	1000.464
Feb	450.994	434.612	885.606
Mar	306.758	430.06	736.818
Apr	283.656	377.746	661.402
May	388.557	295.64	684.197
Jun	29.431	16.915	46.346
Jul	154.226	268.417	422.643
Aug	153.745	457.252	610.997
Sep	165.259	574.48	739.739
Oct	160.055	671.008	831.063
Nov	216.979	712.896	929.875
Dec	421.67	552.285	973.955
<b>Grand Total</b>	<b>3302.918</b>	<b>5220.187</b>	<b>8523.105</b>

1. Sum of Read Value by Month: Reason for June and July's low is that the dataset records time started at 7/11/2019 12am and it ended at 6/2/2020 11.59pm.
2. Temperate low in Australia is around July and August.
3. (Tableau Dashboard 1) Solar panels have highest production in 2019 Q4 with 1936KW. General usage consumption is highest in 2020 Q1 with 1329KW.
4. (Tableau Dashboard 1) The productions amount from solar panel are increasing from July to November, but it starts to decrease from December to May. It probably related to daytime and nighttime length in different months in Australia.

5. (Tableau Dashboard 1) General electrical usage has highest consumption in 2020 Q1 with 1329KW, while solar panel production is 1294KW at this period, therefore, the production from solar panel unable to cover the consumption needs. 2020 Q2 has the same situation.
6. (Tableau Dashboard 2) In dashboard 2, it shows the read value by hours in different quarters. Quarter 3 and quarter 4 have longer period time of solar production. I assume there are longer daytime in Q3 and Q4, but some data are needed to support this.
6. (Tableau Dashboard 3) Scatterplot of temperature with general usage and solar production. The trendline shows no strong relationship between them.
7. Time series model predicts a downtrend for solar production and an uptrend for general usage consumption. The time series model predicts a general usage of **459.87KW** in June 2020, and it forecasts the solar production of **257kW** in June 2020.
8. Linear regression model could be used to predict general usage. It has a R square value of 60.12%. Using the June average temperature of 15 celcius, we predict a general usage of **541.47KW** in June 2020.

## Recommendations

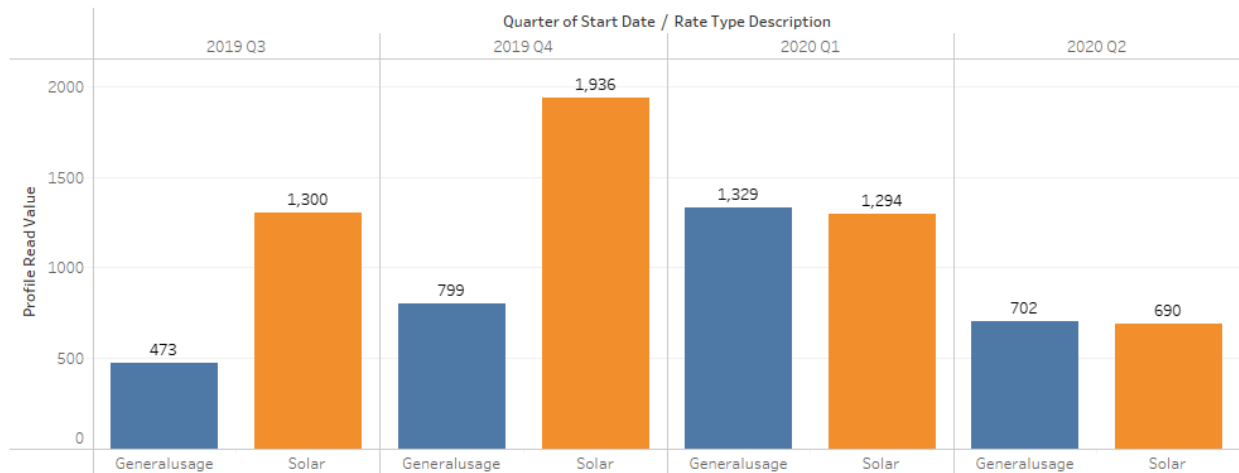
Based on the findings, it looks like the average temperature in city can be used to predict the general electrical usage. We predict a 541.47KW usage in June 2020 using the linear regression model. In addition, one assumption I made in the analysis is that the length of daytime might influence the solar production amount (refer to Tableau Dashboard 2). I assume there are longer daytime in Q3 and Q4. However, more data are needed to examine this. For the next step, we could look for the daytime data in Australia and examine if it can be used to predict the solar production amount.

## Reference

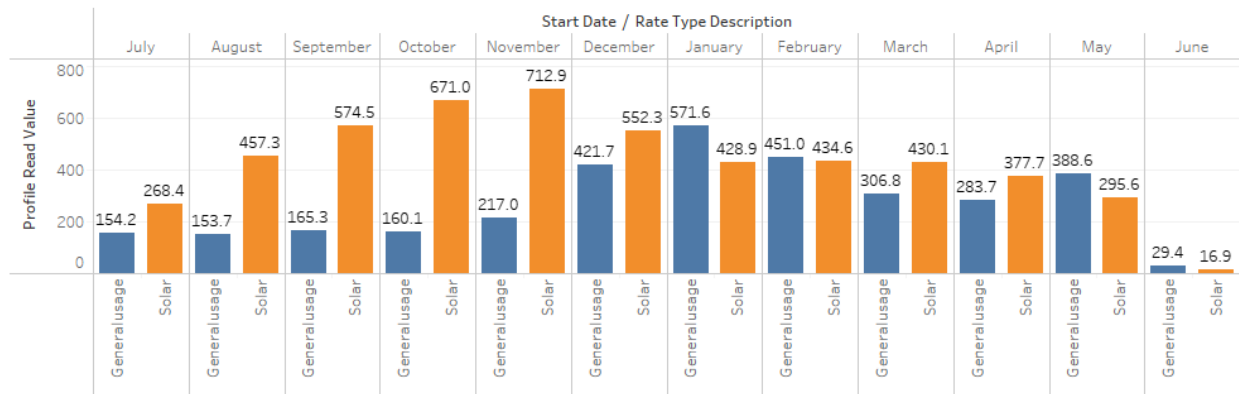
1. Dataset link: <https://www.kaggle.com/julianlee/australian-energy-household-dataset>

### 2. Tableau Dashboard 1

Read Value by Quarters

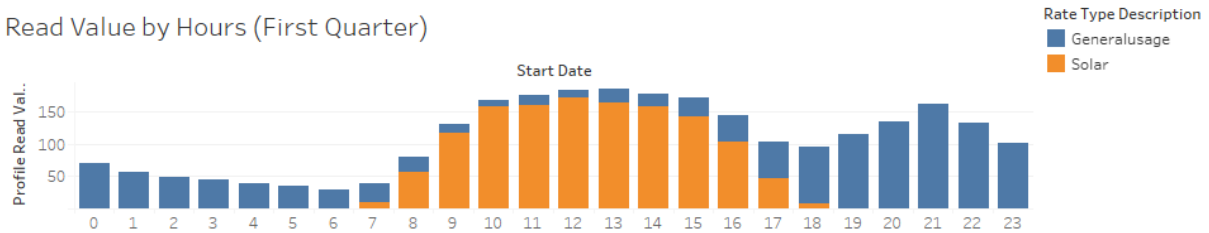


Read Value by Month

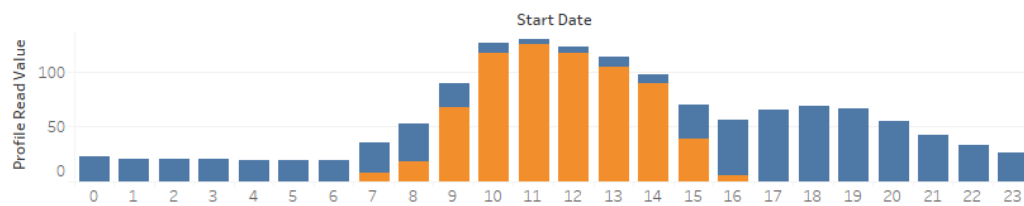


### 3. Tableau Dashboard 2

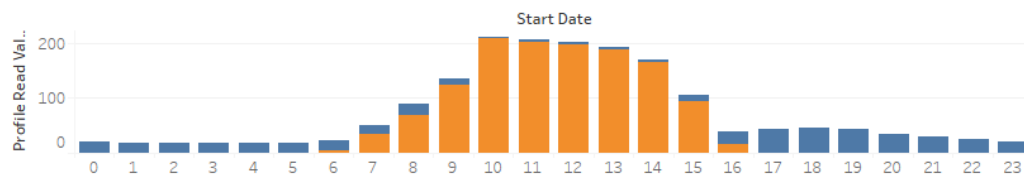
Read Value by Hours (First Quarter)



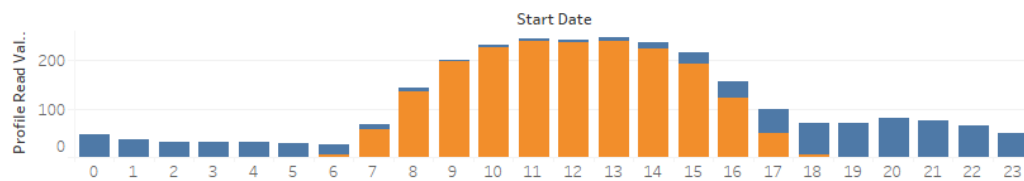
Read Value by Hours (Second Quarter)



Read Value by Hours (Third Quarter)

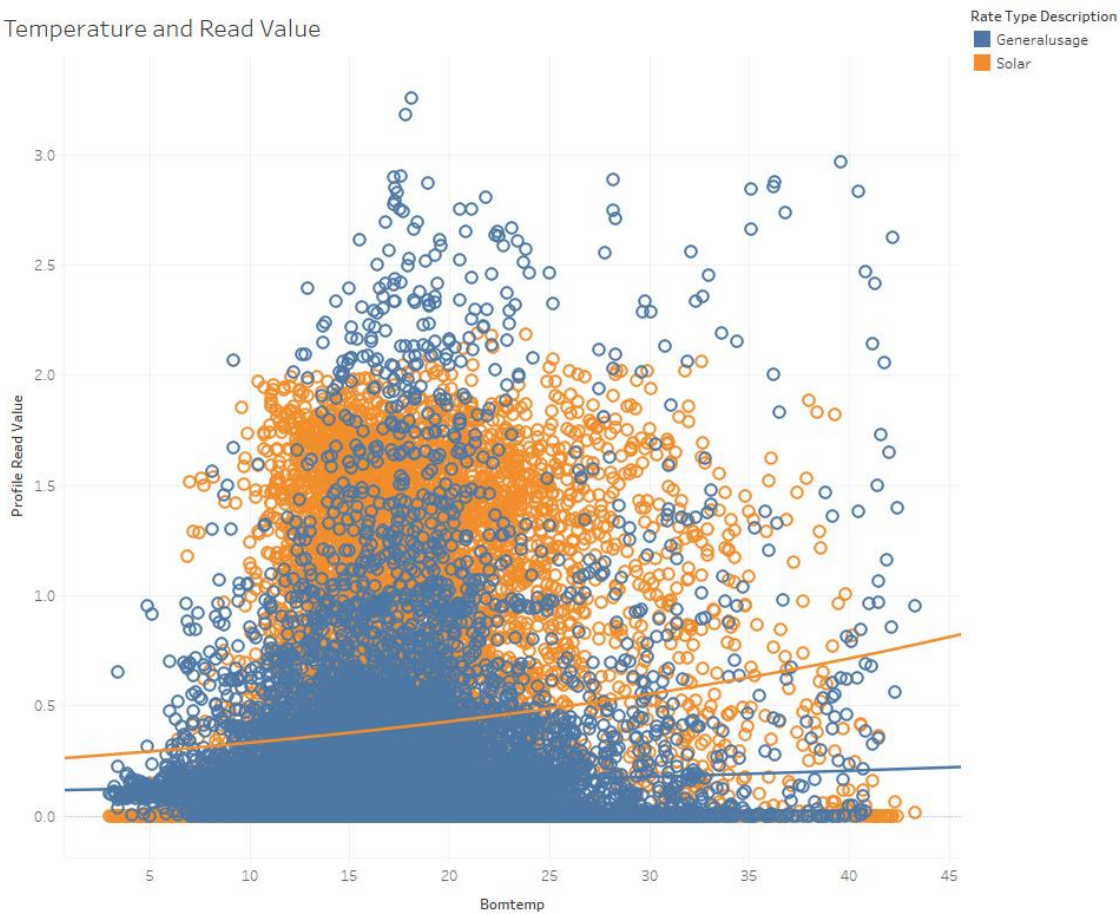


Read Value by Hours (Fourth Quarter)



#### 4. Tableau Dashboard 3

Temperature and Read Value



#### 5. Dashboard 3 Linear model output:

##### Trend Lines Model

A linear trend model is computed for natural log of Profile Read Value given Bomtemp. The model may be significant at  $p \leq 0.05$ . The factor Rate Type Description may be significant at  $p \leq 0.05$ .

**Model formula:** Rate Type Description\*( Bomtemp + intercept )  
**Number of modeled observations:** 17295  
**Number of filtered observations:** 631  
**Model degrees of freedom:** 4  
**Residual degrees of freedom (DF):** 17291  
**SSE (sum squared error):** 38041.8  
**MSE (mean squared error):** 2.20009  
**R-Squared:** 0.106725  
**Standard error:** 1.48327  
**p-value (significance):** < 0.0001

##### Analysis of Variance:

Field	DF	SSE	MSE	F	p-value
Rate Type Description	2	3903.6083	1951.8	887.147	< 0.0001

##### Individual trend lines:

Panels		Color	Line	Coefficients					
Row	Column	Rate Type Description	p-value	DF	Term	Value	StdErr	t-value	p-value
Profile Read Value	Bomtemp	Solar	< 0.0001	6416	Bomtemp	0.0255958	0.0037577	6.81151	< 0.0001
		Generalusage	< 0.0001	10875	intercept	-1.36308	0.0712799	-19.1229	< 0.0001
Profile Read Value	Bomtemp	Generalusage	< 0.0001	10875	Bomtemp	0.0143703	0.0024018	5.98308	< 0.0001
		Solar	< 0.0001	6416	intercept	-2.16691	0.0406559	-53.2988	< 0.0001

