

# Energy Usage Analysis

## An Analysis of Solar Generation Effectiveness

Student Number: 2710017

18 November 2024

### Introduction (20 marks)

#### Objectives

About six years ago, I decided to remove my reliance on oil and attempt to reduce my use of electricity imported from the National Grid. This was to try to reduce my energy costs and also as part of my small contribution towards reducing the use of fossil fuels. So I obtained installation costs and received estimates on how much energy I could generate by using solar panels. The existing setup of oil powered central heating was replaced by an installation of a solar panel array, battery storage and ground source heat pump. The aim was to be able to export surplus electricity and only import when generation and storage failed to meet the demands of the house.

My main objective is to see if the installation meets the predictions made during the sales process and to what extent I no longer have to rely on imported electricity, ie how effective the overall installation is and would, for example, additional battery storage improve it. I used a sample of data for the month of September 2024 as a hopefully representative month (ie not the height of summer with peak solar generation and not winter with low solar generation and peak consumption).

In the final report, R code and outputs have been shown to demonstrate working, however to reduce the size of the document this is suppressed if repeating previous workings. It is obviously all available in the submitted .rmd file.

#### Summary of The Data

The data was collated from three original sources and combined into a single .txt file, then imported as an R dataframe.

```
# Import all data from the tab-separated data file which is held in the data sub-folder
file_path <- './Data/Energy_September_2024.txt'
energy_df <- read.delim((file_path))
# Convert the string date to a valid date format
energy_df$Date <- as.Date(energy_df$Date, '%d/%m/%Y')
```

The data analysed comprises four parts, all daily data, 30 observations, for each day in September 2024:

- Weather: Temperature and solar irradiance readings
- Energy Use: Electricity consumption
- Energy Source: The source of electricity: solar, battery or import from the grid

- *Occupied*: The approximate number of hours the house is occupied each day

All data and supporting files can be found online at Github<sup>1</sup>.

## Weather

Weather data is sourced from the Balquhiddy Weather Station<sup>2</sup> and consists of:

- *Temp* - the mean daily temperature in °C which is derived from 6 readings taken at 4 hourly intervals over a 24 hour period
- *Irrdnce* - irradiance, a measure of the solar energy experienced over a specified area, units are kW/m<sup>2</sup> or W/m<sup>2</sup> and this is used to calculate the theoretical power generated from an array of solar panels<sup>3</sup>

## Energy Use & Source

The distribution of power for the house is managed by a Tesla Powerwall and Controller and an iPhone app is used to monitor this, see Figure 1. All electricity data was downloaded via this app. The imported electricity data was in Wh but is usually reported in kWh.

Electricity used and where it is sourced from:

- *Home\_Total* - total energy used by the house
- *From\_Solar* - solar power generated by an array of 36 solar panels
- *From\_PWall* - battery storage
- *From\_Grid* - the national power grid

Electricity generated by the solar panels and where it is used (the controller intelligently makes the routing decisions):

- *Solar\_Total* - total energy generated by the solar panels
- *To\_Home* - consumption by the house
- *To\_PWall* - for battery storage
- *To\_Grid* - export to the national power grid

**TO DO:** Tidy up table columns display . . . look at Pandoc options etc to change layout for table and whole document? <https://pandoc.org/MANUAL.html#synopsis>

Once the sources of data have been collated and loaded the dataset consists of 30 observations and 12 columns. The first 6 rows are shown below:

```
# Display the first 6 rows of the data
#head(energy_df)
kable(head(energy_df), caption = 'First 6 Rows of the source dataset')
```

<sup>1</sup><https://github.com/StuartG24/Home-Solar-Usage-Analysis>

<sup>2</sup><https://www.blsc.org/weather>

<sup>3</sup>Wikipedia: [https://en.wikipedia.org/wiki/Solar\\_irradiance](https://en.wikipedia.org/wiki/Solar_irradiance)

Table 1: First 6 Rows of the source dataset

Date	Home_Total	From_Pwall	From_Solar	From_Grid	Solar_Total	To_Home	To_Pwall	To_Grid	Temp	Irrdnce	Occupied
2024-09-01	19048	11652	2180	5217	2360	2180	172	9	12.9	73.18	24
2024-09-02	11304	5224	1171	4909	1232	1171	50	11	13.4	67.08	6
2024-09-03	13867	7062	3690	3115	6372	3690	2664	18	9.9	269.36	0
2024-09-04	16241	9315	4380	2546	6768	4380	2378	10	9.5	344.50	0
2024-09-05	17960	9127	5744	3090	11036	5744	5204	88	14.5	336.36	12
2024-09-06	16015	10998	4617	400	14726	4617	9152	957	16.6	297.88	24

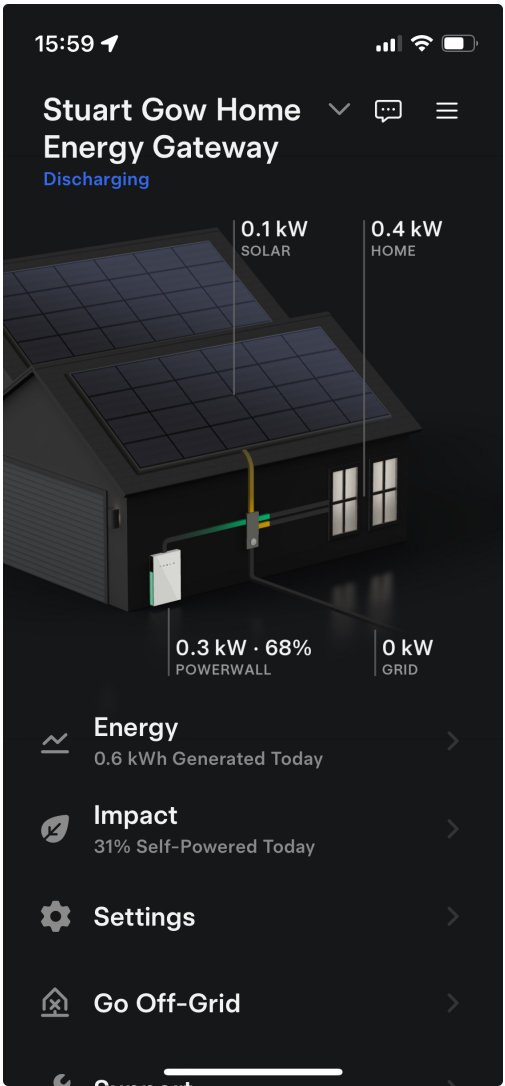


Figure 1: Tesla Powerwall

## Methods and Results - (40 marks)

The data was analysed in several groups and themes, in summary:

- Energy Consumption - What drives energy consumption?
- Solar Generation Effectiveness - How well does the installation meet sales promises?
- Solar Energy Sufficiency - How well does solar meet energy demand?

?? Predict annual generation using non September ?? monthly irradiance data: [https://re.jrc.ec.europa.eu/pvg\\_tools/en/#MR](https://re.jrc.ec.europa.eu/pvg_tools/en/#MR)

- !! Fishers Exact Test
- ?? Contingency tables
- ?? box plots

### *WIP - Analysis, expected conclusions:*

- 1) See ideas in notebook, 17/11
- 2) Fishers exact test
- 3) Full year projections, observations
- 4) Consumption is covered by solar .. or not? how much? how much grid still needed
- 5) Consumption is linked to temperature and house occupancy
- 6) Solar generation is linked to irradiance .. but how much?
- 7) Any other links such as temperature? .. probably not usage or occupancy though?
- 8) ?? Cannot account for £ cost and different costs at times of day .. battery importing then for example
- 9) ?? Battery timing in and out complicates the analysis?
- 10) Increased battery will smooth out across days? forecast storage/impact .. but can't see the intra-day detail to better analyse
- 11) Solar generation meets the forecasts at purchase .. need original data/estimates!?

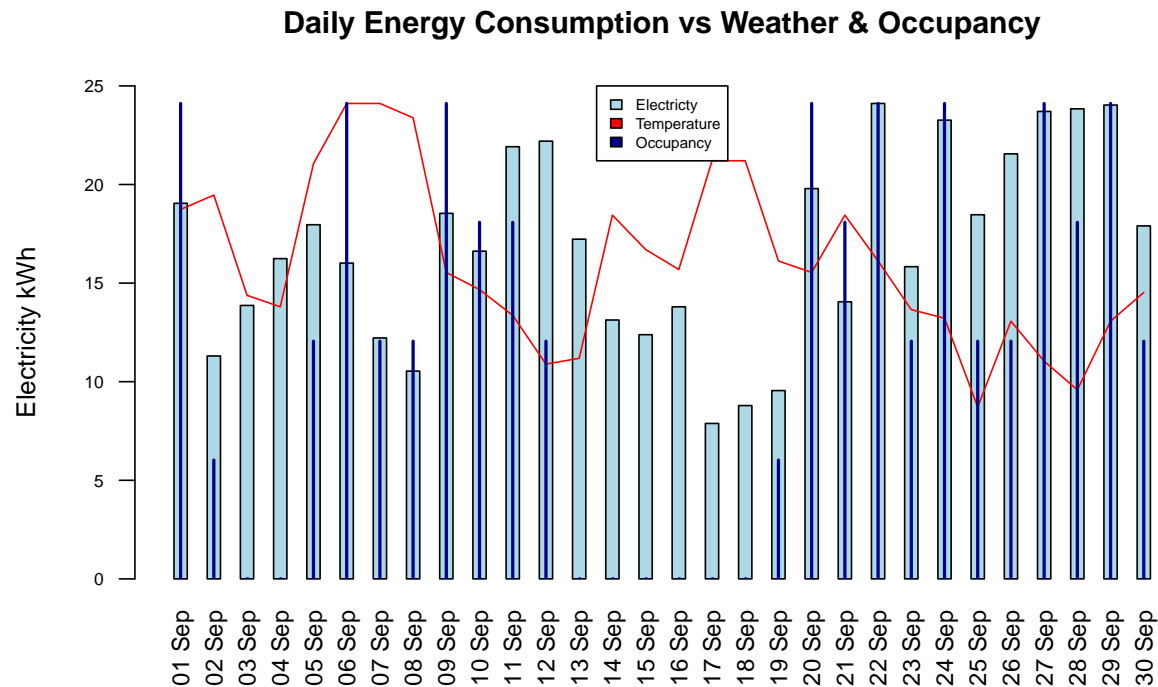
## Energy Consumption

The energy consumption (electricity kWh) of the house was compared to the weather and its occupancy. The diagram below summarises this. The temperature and occupancy values have been scaled to only show the relative size and so no values are displayed.

```
# Base barplot of energy consumption
barplot_result <- barplot(energy_df$Home_Total/1000, names.arg = format(energy_df$Date, "%d %b"),
  cex.name=0.9, las=2, cex.axis=0.7, main="Daily Energy Consumption vs Weather & Occupancy",
  ylab="Electricity kWh", ylim=c(0,25), col = 'lightblue', space = 1.5)

# Add scaled lines for temperature and occupancy
scaled_temp <- energy_df$Temp * max(energy_df$Home_Total/1000) / max(energy_df$Temp)
lines(barplot_result, scaled_temp, type = 'l', col = "red")
scaled_occ <- energy_df$Occupied * max(energy_df$Home_Total/1000) / max(energy_df$Occupied)
lines(barplot_result, scaled_occ, type = 'h', col = "darkblue", lwd = 2)

legend('top', legend=c("Electricity", "Temperature", "Occupancy"),
  fill=c("lightblue", "red", "darkblue"), cex = 0.6)
```



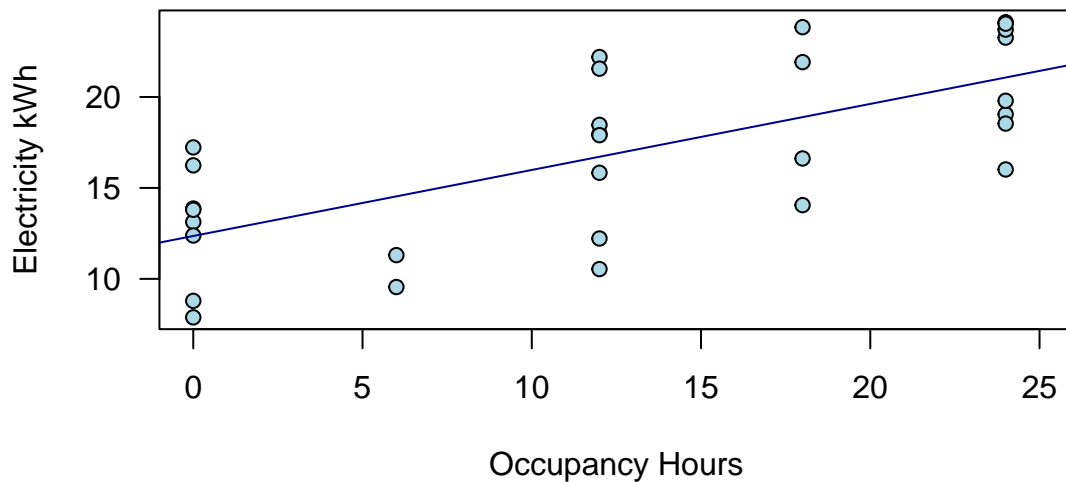
It was expected that the energy demand from the house should be related to the occupancy and the temperature, however visually there does not appear to be a strong link with temperature but potentially there is a link with occupancy. So three linear regressions were carried out to better identify any relationships.

First looking at the relationship between energy demand and occupancy.

```
# Plots and linear regression for energy and occupancy
regression_model <- lm(energy_df$Home_Total/1000 ~ energy_df$Occupied)
regression_summary <- summary((regression_model))
alpha <- regression_summary$coefficients["(Intercept)", "Estimate"]
beta <- regression_summary$coefficients["energy_df$Occupied", "Estimate"]
p_value <- regression_summary$coefficients["energy_df$Occupied", "Pr(>|t|)"]
adj_r_squared <- regression_summary$adj.r.squared

plot(energy_df$Home_Total/1000 ~ energy_df$Occupied, main="Daily Energy Demand vs Occupancy",
     xlab="Occupancy Hours", ylab="Electricity kWh", las=1, xlim=c(0,25),
     pch=21, bg="lightblue")
abline(regression_model, col="darkblue")
```

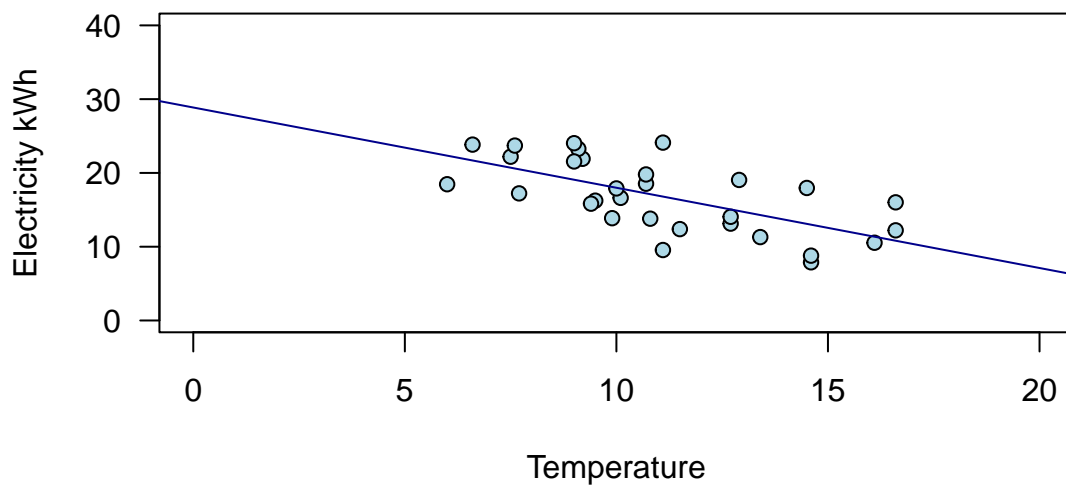
### Daily Energy Demand vs Occupancy



The regression line determined was  $\text{energy} = 12.36 (\alpha) + 0.3631 (\beta) * \text{occupancy}$ . This fit had a p-value of  $2.24e-05$  and adjusted R-squared of 0.46. For a linear regression the null hypothesis is that  $\beta$  is zero, ie  $H_0 : \beta = 0$ . Here the p-value is less than the 5% critical value and so we can reject  $H_0$  and infer that energy and occupancy are correlated, however the relative low value of adjusted R-squared suggests that the correlation is not very strong.

Then looking at the relationship between energy demand and temperature (NB: same R Code so not printed to save space).

### Daily Energy Demand vs Temperature



The regression line determined was  $\text{energy} = 28.87 (\alpha) + -1.088 (\beta) * \text{temperature}$ . This fit had a p-value

of 0.00011 and adjusted R-squared of 0.4. Here the p-value is less than the 5% critical value and so we can reject  $H_0$  and infer that energy and temperature are inversely correlated, however the relative low value of adjusted R-squared suggests that the correlation is not very strong.

Then a multi-linear regression was performed looking at the relationship between energy demand and temperature plus occupancy (NB: A three dimensional scatter plot was not attempted).

```
# Multiple linear regression for energy and occupancy + temperature
regression_model <- lm(energy_df$Home_Total/1000 ~ energy_df$Temp + energy_df$Occupied)
regression_summary <- summary((regression_model))
alpha <- regression_summary$coefficients["(Intercept)", "Estimate"]
beta0 <- regression_summary$coefficients["energy_df$Occupied", "Estimate"]
p_value0 <- regression_summary$coefficients["energy_df$Occupied", "Pr(>|t|)"]
beta1 <- regression_summary$coefficients["energy_df$Temp", "Estimate"]
p_value1 <- regression_summary$coefficients["energy_df$Temp", "Pr(>|t|)"]
adj_r_squared <- regression_summary$adj.r.squared

#print(sprintf("Alpha: %.3f, Beta Occupancy: %.3f, Beta Temp: %.3f", alpha, beta0, beta1))
#print(sprintf("p-value Occp: %.4f, p-value Temp: %.4f, Adj R-Squared: %.3f", p_value0, p_value1, adj_r_squared))
```

The regression line determined was  $\text{energy} = 23.45 (\alpha) + 0.3282 (\beta) * \text{occupancy} + -0.9654 (\beta) * \text{temperature}$ . This fit had a p-value of 7.59e-08 for occupancy and 3.54e-07 for temperature and adjusted R-squared of 0.79. Here both p-values are less than the 5% critical value and so we can reject  $H_0$  and infer that energy and occupancy plus temperature (inversely) are correlated, the high value of adjusted R-squared suggests that the correlation is quite strong.

## Solar Generation Effectiveness

The amount of energy generated by solar panels is a function of their size and the level of sunshine received, measured by irradiance. There is also a loss factor that reflects several things including the efficiency of the solar panels and the inverter. Additionally, irradiance observations are taken from local weather stations which may not experience the same shading from the sun as that experienced at the site of the solar panels.

$$\text{Power(kWh)} = \text{Area(m}^2\text{)} * \text{Irradiance(kWh/m}^2\text{)} * \text{LossFactor}$$

```
# The installation assumptions
solar_panels_count <- 36
solar_panels_area <- solar_panels_count * 2
solar_panels_max <- solar_panels_count * 275
solar_estimated_pa <- 7920
loss_factor <- solar_estimated_pa / solar_panels_max
irradiance_assumed <- solar_panels_max / solar_panels_area
```

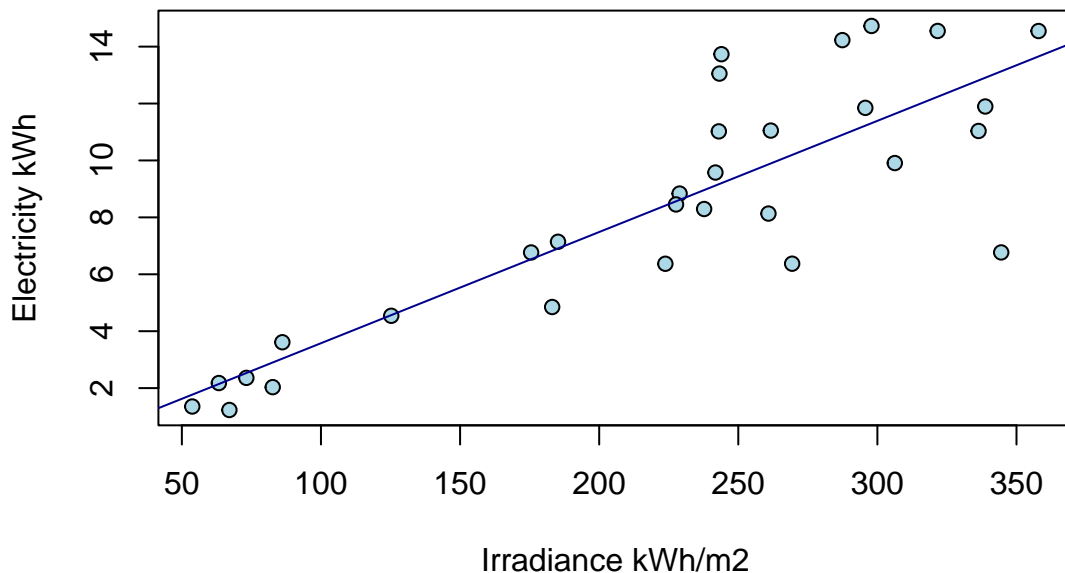
At installation, the annual generation power for this solar array was calculated as 7,920 kWh pa using an area of 72  $\text{m}^2$ , irradiance 137.5  $\text{kWh/m}^2$  pa and a loss factor of 0.8. Using the September observations of irradiance and generated solar energy, the effectiveness of the installation can be compared to the estimates (sales promises) made originally. It is very likely that the sales estimates were optimistic.

*Link Between Energy Generation & Irradiance*

```
# For solar generation, how well is this predicted by irradiance?
regression_model <- lm(energy_df$Solar_Total/1000 ~ energy_df$Irrdnce)
regression_summary <- summary(regression_model)
```

```
plot(energy_df$Solar_Total/1000 ~ energy_df$Irrdnce,
     main="Fit for Daily Solar Energy vs Irradiance", xlab="Irradiance kWh/m2", ylab="Electricity kWh",
     pch=21, bg="lightblue")
abline(regression_model, col="darkblue")
```

### Fit for Daily Solar Energy vs Irradiance



```
alpha <- regression_summary$coefficients["(Intercept)", "Estimate"]
beta0 <- regression_summary$coefficients["energy_df$Irrdnce", "Estimate"]
p_value1 <- regression_summary$coefficients["energy_df$Irrdnce", "Pr(>|t|)"]
adj_r_squared <- regression_summary$adj.r.squared

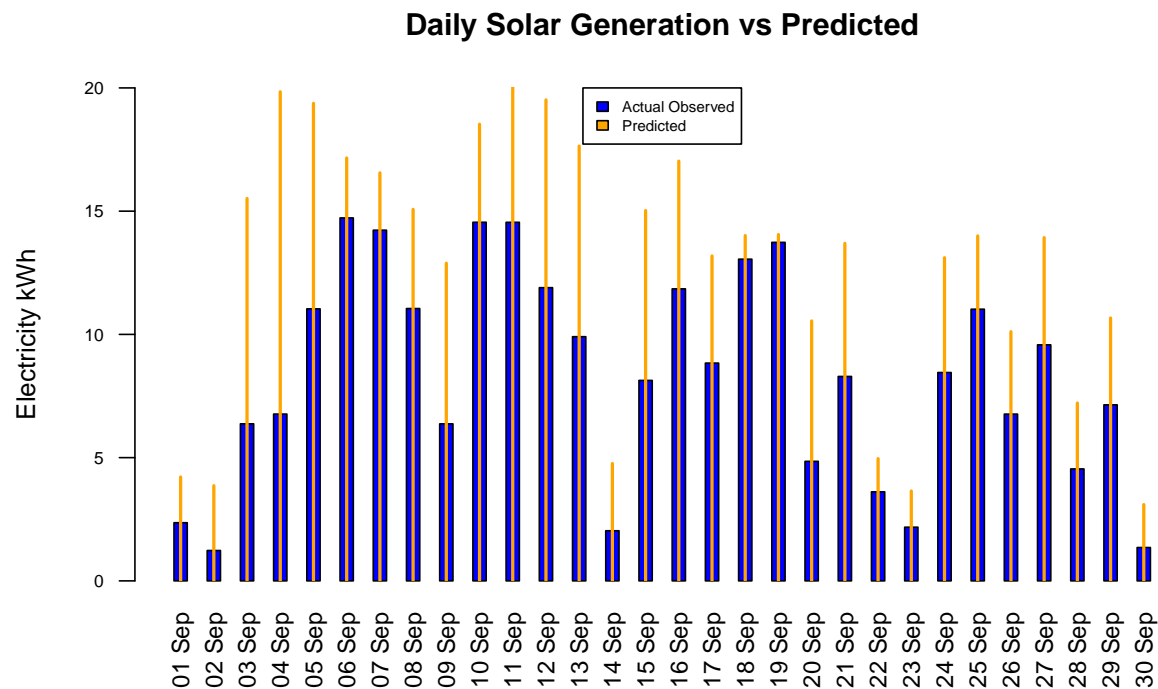
#print(sprintf("Alpha: %.3f, Beta Irradiation: %.3f", alpha, beta0))
#print(sprintf("p-value Irradiation: %.4f, Adj R-Squared: %.3f", p_value0, adj_r_squared))
# Good fit between solar energy and irradiation with p-value < 5% and a high r-squared
```

It was expected that the generated solar energy is related to the irradiance and a quick scatter plot and linear regression confirmed this, with a p-value < 0.005 and a high adjusted r-squared 0.71.

#### Evaluate Effectiveness - T-Test

The sales prediction for the daily solar energy was calculated using the formulae above and with the irradiance observations during September. The means for the month are 8.35 kWh and 12.8 kWh respectively. The daily totals are compared in the plot below.



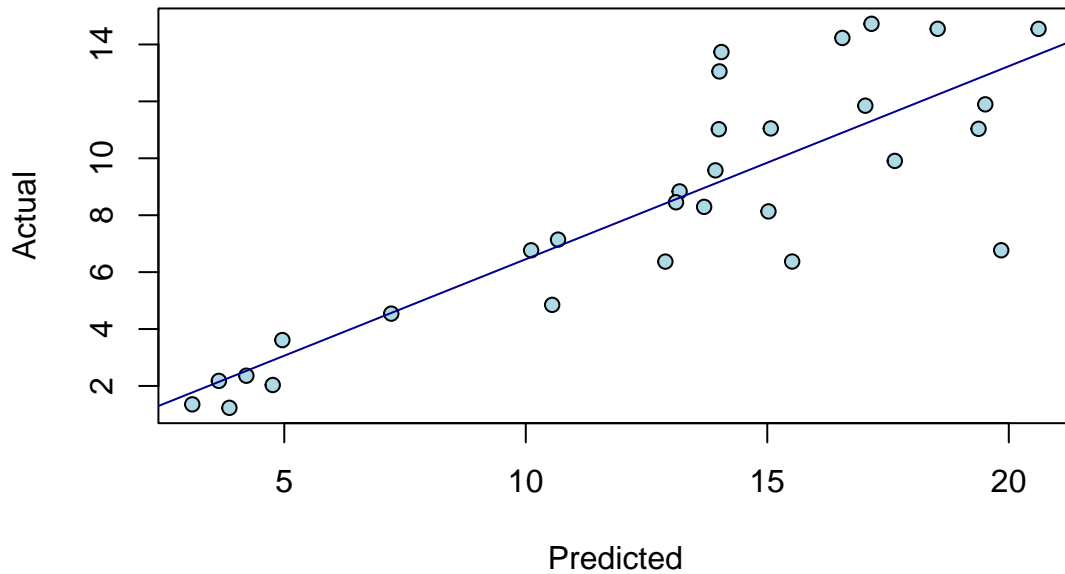


The figure clearly shows that the predicted generation is higher than the actual, although the differences do vary. Further examination was completed to better understand this. A scatter plot below shows the relationship between the predicted and generated energy.

```
# Scatter plot and regression for solar generation & solar predicted
regression_model <- lm(energy_df$Solar_Total/1000 ~ solar_predicted)
regression_summary <- summary(regression_model)

plot(energy_df$Solar_Total/1000 ~ solar_predicted,
     main="Solar Energy Actual vs Predicted (kWh)", xlab="Predicted", ylab="Actual",
     pch=21, bg="lightblue")
abline(regression_model, col="darkblue")
```

## Solar Energy Actual vs Predicted (kWh)



```
alpha <- regression_summary$coefficients["(Intercept)", "Estimate"]
beta0 <- regression_summary$coefficients["solar_predicted", "Estimate"]
p_value1 <- regression_summary$coefficients["solar_predicted", "Pr(>|t|)"]
adj_r_squared <- regression_summary$adj.r.squared

#print(sprintf("Alpha: %.3f, Beta Irradiation: %.3f", alpha, beta0))
#print(sprintf("p-value Irradiation: %.4f, Adj R-Squared: %.3f", p_value0, adj_r_squared))
```

As expected the actual generated solar energy vs predicted do appear to fit well, with a p-value < 0.005 and a high adjusted r-squared 0.71.

To try to better understand the relationship between the actual and predicted energy, a T-Test was completed. The null hypothesis is that the installation performs as well as the sales promises; specifically that the average daily energy production in September  $\mu_{act}$  is the same as that promised (expected)  $\mu_{exp}$ .

$$H_0 : \mu_{act} = \mu_{exp}$$

the alternative hypothesis is that the installation does not perform as promised

$$H_1 : \mu_{act} < \mu_{exp}$$

A T-Test was performed to evaluate  $H_0$  and to calculate the p-value which is the probability, if  $H_0$  is true, of obtaining the observation, or an observation more extreme. In this case more extreme is that the actual mean observed is less than the expected mean because we suspect that the installation does not perform as well as the sales promises; so a one-tailed paired test was used. Log values were used to try to make both samples closer to a normal distribution.

```
# Evaluate the T-Test
#t_test_result <- t.test(energy_df$Solar_Total/1000, solar_predicted, alternative = "less", paired = TR
t_test_result <- t.test(log10(energy_df$Solar_Total/1000), log10(solar_predicted), alternative = "less"
```

The p-value is  $< 0.005$  which is statistically significant and the null hypothesis can be rejected at the 5% level. From this we can infer that the installation is not generating solar energy as effectively as originally promised.

#### *Evaluate Effectiveness - Binomial Probability*

```
# How often are the actual values less then the predicted?
occurrences = sum(table(energy_df$Solar_Total/1000 < solar_predicted))
```

To confirm the above, further analysis was performed, to look at the frequency of actual daily generation being less than the predicted generation ( $energy_{act} < energy_{exp}$ ). In September this happened on 30 days out of 30. The null hypothesis is that we should always achieve the predicted generation, so

$$H_0 : p(\text{energy}_{act} < \text{energy}_{exp}) = 0$$

```
# Determine the p-value for act < predicted on the given number of occurrences
# In a 30 day month and H0 saying the probability is zero
p_value <- dbinom(occurrences, 30, 0)
```

The p-value is  $< 0.005$  which is statistically significant and the null hypothesis can be rejected at the 5% level. From this we can infer that the installation is not generating solar energy as effectively as originally promised.

XXXXXXXXXXXXXXXXXXXX

## Solar Energy Sufficiency

**!! work on this ....**

Next, the energy demands of the house were examined to see how well solar generated power matched this and how much imported electricity is needed to fill the gaps.

?? plot ?? in line figures rather than printed...

The first bar plot below shows the total energy consumed per day by the house and a breakdown of where this energy is sourced from. It appears that the energy generated by the solar panels only meets a small proportion of the total daily consumption, on average 26%.

```
average_solar_percent <- mean(c(energy_df$From_Solar / energy_df$Home_Total)) * 100
print(paste("Straight percentage:", average_solar_percent))
```

```
## [1] "Straight percentage: 26.0395444324628"
```

However, this is misleading as generated solar energy is often first stored in the battery for later use or exported to the grid if the battery becomes full. The second bar plot shows this more clearly. With solar energy distributed to the home, the battery or exported to the grid on average: 62%, 30% and 8% respectively.

```
solar_home_percent <- mean(c(energy_df$To_Home / energy_df$Solar_Total)) * 100
solar_powerwall_percent <- mean(c(energy_df$To_Pwall / energy_df$Solar_Total)) * 100
solar_grid_percent <- mean(c(energy_df$To_Grid / energy_df$Solar_Total)) * 100
print(paste("To Home:", solar_home_percent, "To Powerwall:", solar_powerwall_percent,
            "To Grid:", solar_grid_percent))
```

```
## [1] "To Home: 62.0971977996547 To Powerwall: 29.9891405679381 To Grid: 7.91530834160956"
```

A good measure of how well the solar generation meets the needs of the house's energy consumption is to simply look at how much usage is catered for without any import from the grid, and this is 80% on average.

```
adjusted_solar_percent <- mean(c((energy_df$Home_Total - energy_df$From_Grid) /
                                energy_df$Home_Total)) * 100
print(paste("Without Grid Imported:", adjusted_solar_percent))
```

```
## [1] "Without Grid Imported: 79.9652308577059"
```

## Conclusions - 20 marks

### Energy Consumption

- Analysis showed that energy consumption is strongly linked to combined temperature and occupancy
- However the link with temperature only was surprisingly weak
  - The temperature in September was only in the range 6 to 17 with a mean of 11 and this is not very extreme.
  - Good insulation with the above may have reduced the impact, colder weather will have a bigger impact, the calculated regression line suggests this with an energy demand of nearly 30 kWh at freezing point

### Solar Generation Effectiveness

- Analysis showed that the solar power generated is consistently less than that predicted during the month of September
- ?? probably throughout the year but only have monthly
- ?? but is this caused by inefficiency in the generation equipment or is it caused because the irradiance experienced is less than that predicted? don't have annual irradiance and generation figures to determine this

##?? Discussion

*TO DO: Structure into two sections*

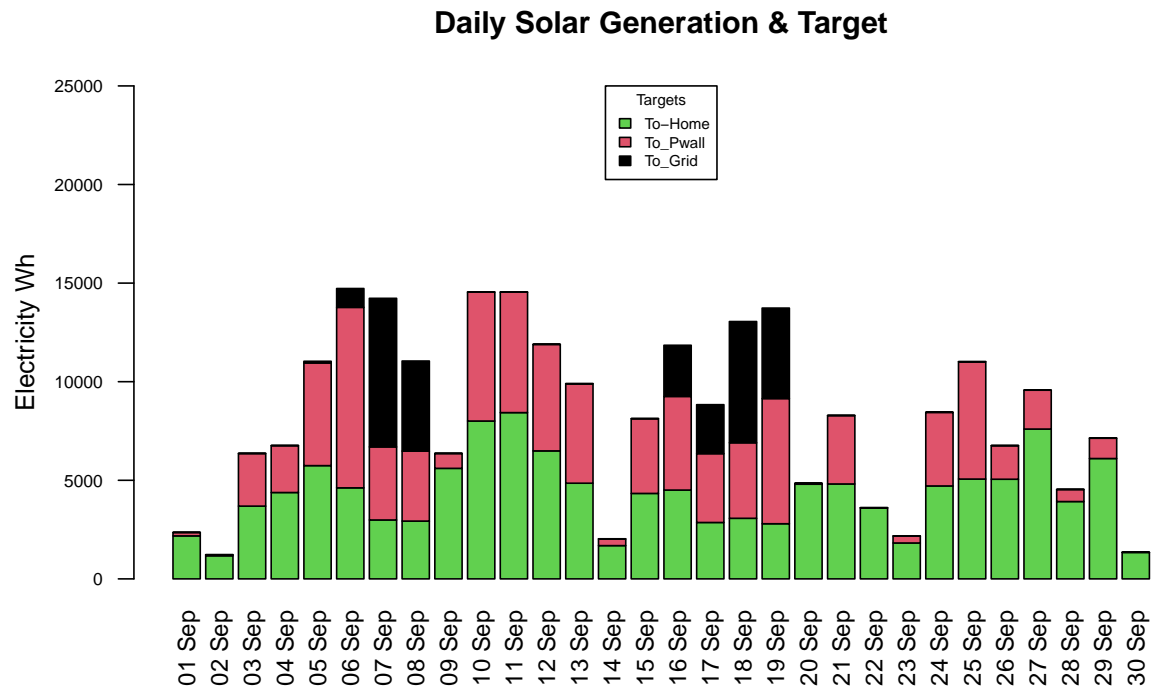
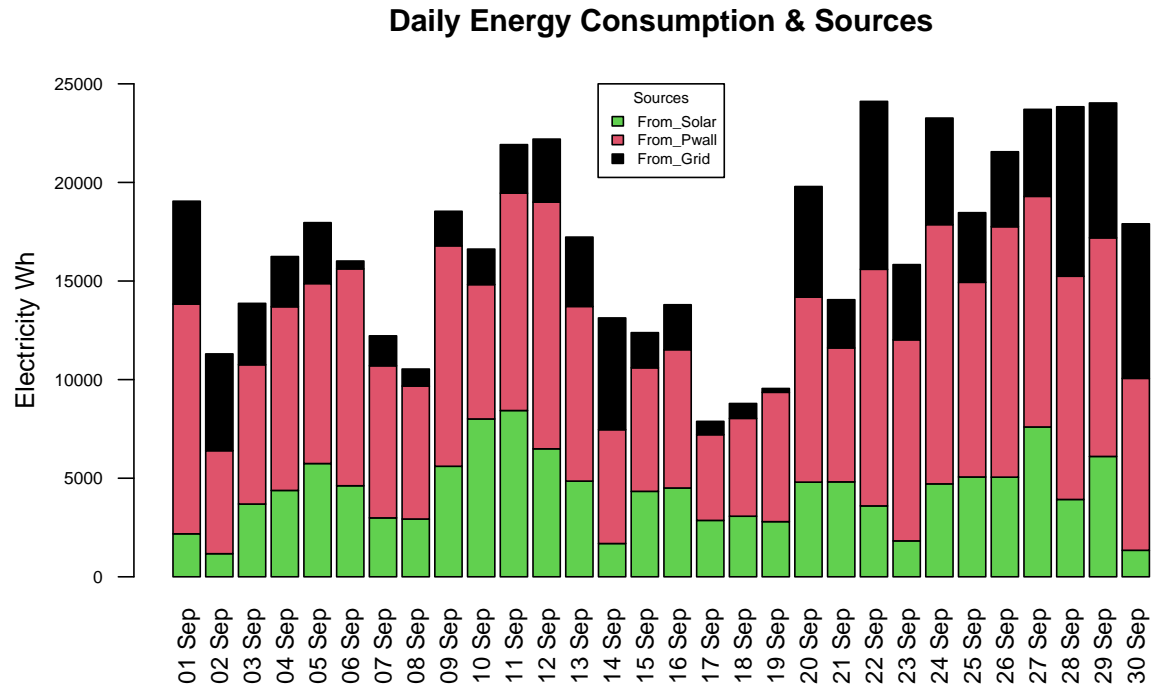
##?? Conclusion

Test citations (Crawley, 2014) and as Fraix-Burnet (2016)

Spiegel and Schiller (2012)

## Appendices

?? Energy Consumption ....



## References

Crawley, M.J. (2014) *Statistics: An introduction using R*. 2nd Edition. John Wiley & Sons.

Fraix-Burnet, D. (2016) ‘Introduction to R’, *Statistics for Astrophysics: Clustering and Classification*, Volume 77(2016), pp. 3–12. Available at: <https://doi.org/10.1051/eas/1677002>.

Spiegel, M.R. and Schiller, J. (2012) *Schaum’s outline probability and statistics*. 4th edn. McGraw Hill.