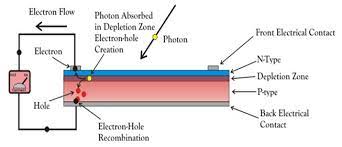
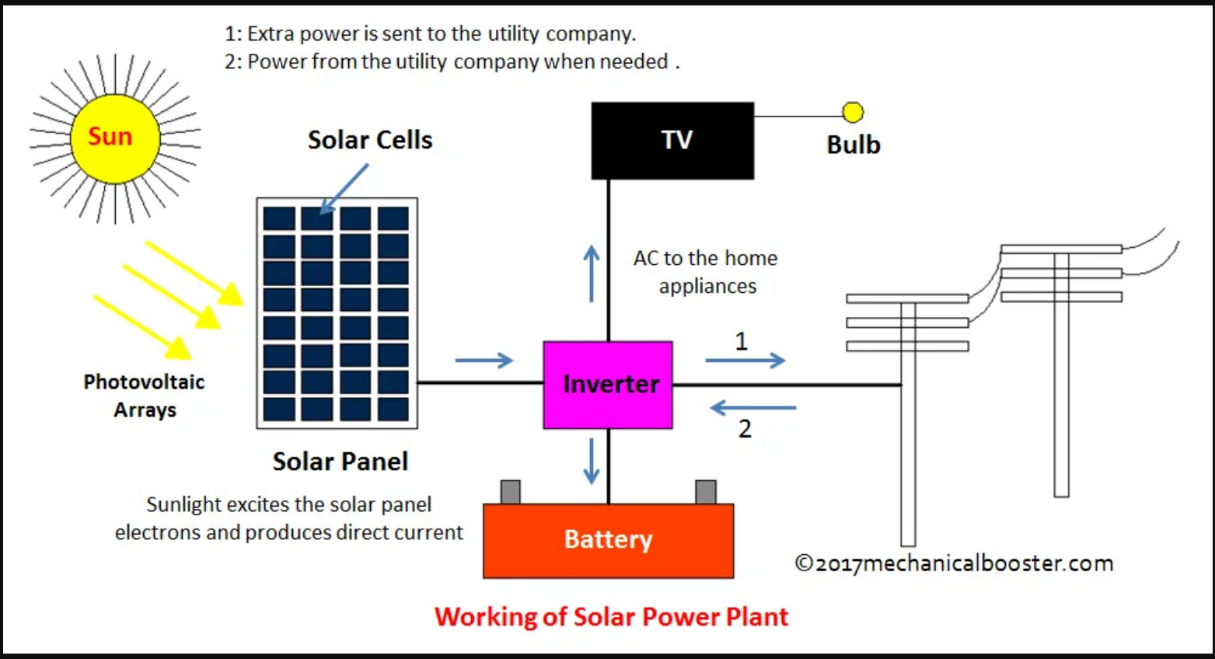
Solar Power Plants. How do they work?

* **Solar Panels are made up of different photovoltaic cells connected in series.**
* **To understand the working of the solar power generation. We first have to understand what are photovoltaic cells made up of and how do they generate Dc Power with the help of irradiation from the Sun.**
* **Photovoltaic cells are made up by combining two types of silicone materials one over another. The upper part of the cell contains N-Type semiconductors containing excess electrons and the bottom part being made up of P-Type semiconductor containing holes.**
* **When they are connected a P-N Junction is formed at boundary of the two.**

****

* **When photons hit the electrons at the N-Type junction the electron absorbs its energy and try to move to the P-Type semiconductor creating a flow of charge thus creating a potential difference between the two sides resulting in Dc Power Generation.**
* **Then the Dc Power is transferred to the Inverters which converts it into Ac Power and then steps up the current according to the requirement of the plant.**

****

### **Typically, the factors that determine the efficiency of the solar power plant are:**

1. The Irradiation falling on the solar panels.
2. Cleanliness of the solar panels.
3. Inverter efficiency
4. Ambient Temperature (For the plant to work efficiently, the ideal temperature ranges between 35 - 50 degrees C.)
5. Inverter or Panel Seniority

#### **The questions which we will be trying to answer through our analysis:**

* Can we predict the power generation based on the weather?
* can we identify faulty or sub optimally performing equipment?
* Can we identify the need for panel cleaning/maintenance?
* Can we calculate the total wasted energy in the plant ….?
* **first let’s start by loading the libraries with which we are going to be working with**

library(tidyverse)

library(dplyr)

library(skimr)

library(janitor)

library(here)

library (RColorBrewer)

library(plotly)

library(psych)

library(lubridate)

* **Now lets load the dataset**

*# Loading the dataset into Data frames.*

plant1\_gen <- read\_csv("../input/solar-power-generation-data/Plant\_1\_Generation\_Data.csv")

plant2\_gen <- read\_csv("../input/solar-power-generation-data/Plant\_2\_Generation\_Data.csv")

plant1\_we <- read\_csv("../input/solar-power-generation-data/Plant\_1\_Weather\_Sensor\_Data.csv")

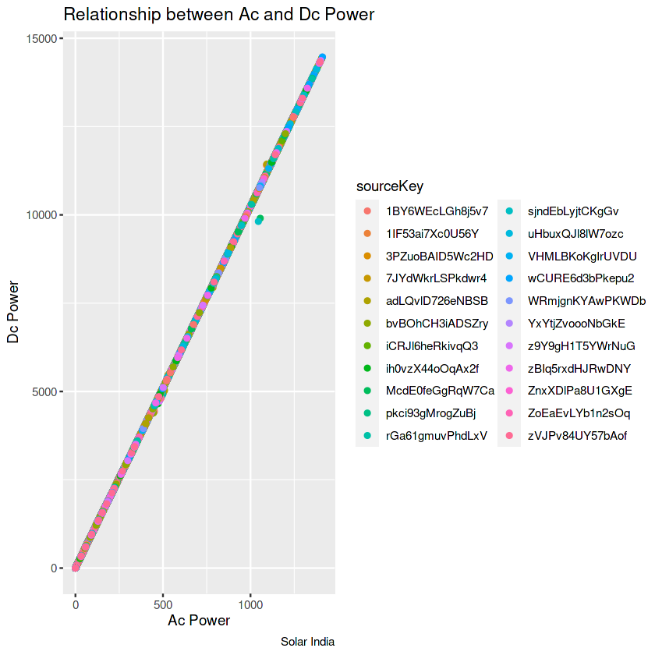
plant2\_we <- read\_csv("../input/solar-power-generation-data/Plant\_2\_Weather\_Sensor\_Data.csv")

* One important metric to determine if an inverter or solar panel is working efficiently or not will be the Electrical Efficiency So lets just calculate that

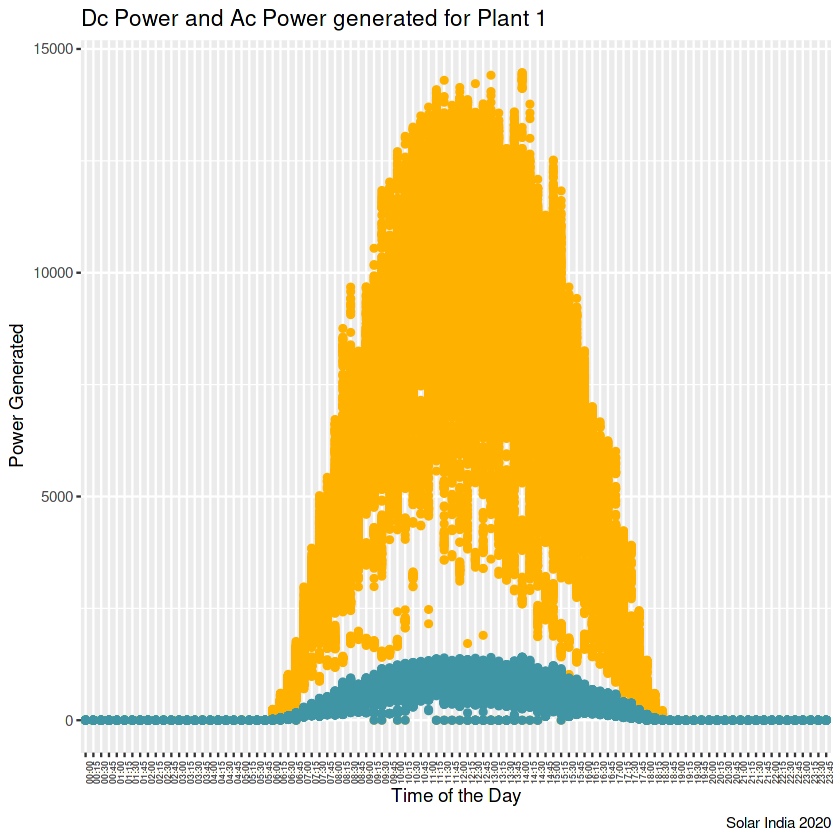
*# Calculating electrical efficiency for the plant.*

plant1\_gen <- plant1\_gen %>%

mutate(electrical\_eff = if\_else(dcPower == 0 | acPower == 0, 0, (acPower/dcPower)\*100))



* **We can see that the Ac and Dc Power have linear correlation, with absolutely no outliers**

****

* **Since, Dc Power Generated by a solar panel is Irradiation Area of the panel solar efficiency. And according to our assumption, the area and the irradiation received is the same for every panel.**
* **So higher the solar efficiency of the panel, the more Dc Power it produces or We can say that Dc Power is directly proportional to the solar efficiency of the solar panel (In this case).**
* **Now, let’s calculate the daily yield on different days by different inverters**

*# Daily yield for Plant.*

plant\_by\_yield <- plant1\_gen %>%

group by (date, source Key) %>%

summarize (yield = max (daily Yield))

*# Plotting daily yield on different days.*

plant\_by\_yield %>%

group\_by(date) %>%

summarize(total\_yield = sum(yield)) %>%

ggplot(aes(x = date, y = total\_yield)) +

geom\_line(color = "#ffb100", size = 0.6) +

geom\_point(color = "#ffb100") +

labs(title = "Power Generated by Plant 1", x = "Date", y = "Power Generated (Kw)") +

theme(axis.text.x = element\_text(face = "bold", size = 8, color = "#4095a5"))

*# Plotting daily yield on different days.*

plant\_by\_yield %>%

group\_by(date) %>%

summarize (total\_yield = sum(yield)) %>%

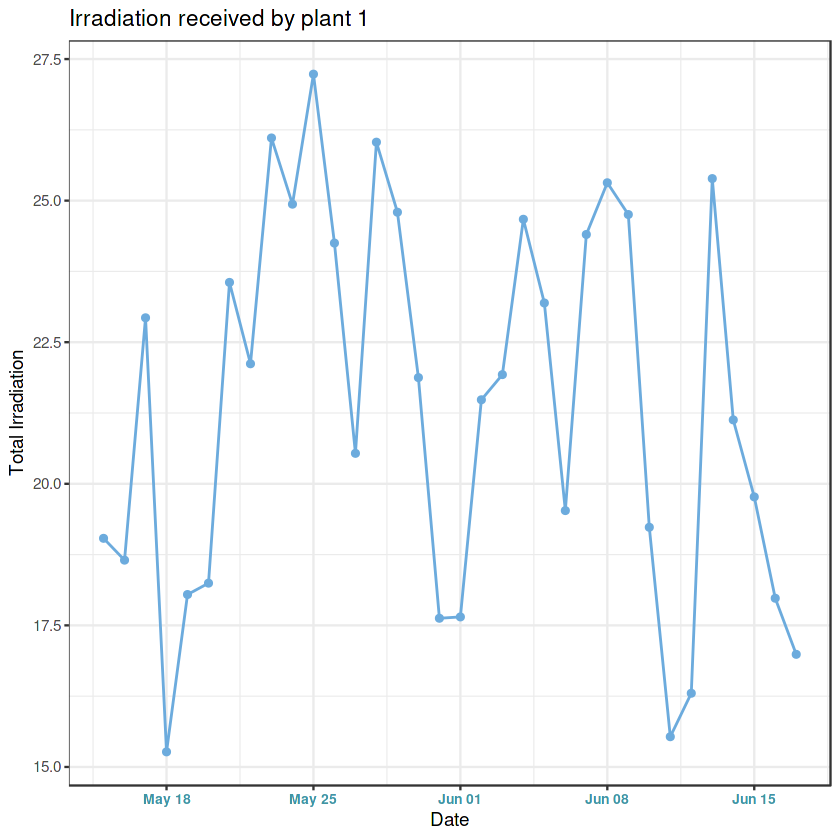
ggplot (aes(x = date, y = total\_yield)) +

geom\_line (color = "#ffb100", size = 0.6) +

geom\_point (color = "#ffb100") +

labs (title = "Power Generated by Plant 1", x = "Date", y = "Power Generated (Kw)") +

theme (axis.text.x = element\_text(face = "bold", size = 8, color = "#4095a5"))



**We can see a similar trend between the power generated and irradiation.**

* **Now let’s have a look at the Dc Power generated by different source keys at different time intervals to see if there are any underperforming inverters**

*# For the first 11 inverters*

plant1\_gen %>%

filter(sourceKey %in% unlist(first11\_inv\_1)) %>%

group\_by(time, sourceKey) %>%

summarize(dc\_power = sum(dcPower)) %>%

ggplot(aes(x = time, y = dc\_power,group = sourceKey)) +

geom\_line(aes(color = sourceKey)) +

scale\_color\_brewer(palette = "Paired") +

theme(axis.text.x = element\_text(face = "bold", angle = 90, size = 7), legend.position = c(0.9,0.6))

*# For the last 11 inverters*

plant1\_gen %>%

filter(sourceKey %in% unlist(last11\_inv\_1)) %>%

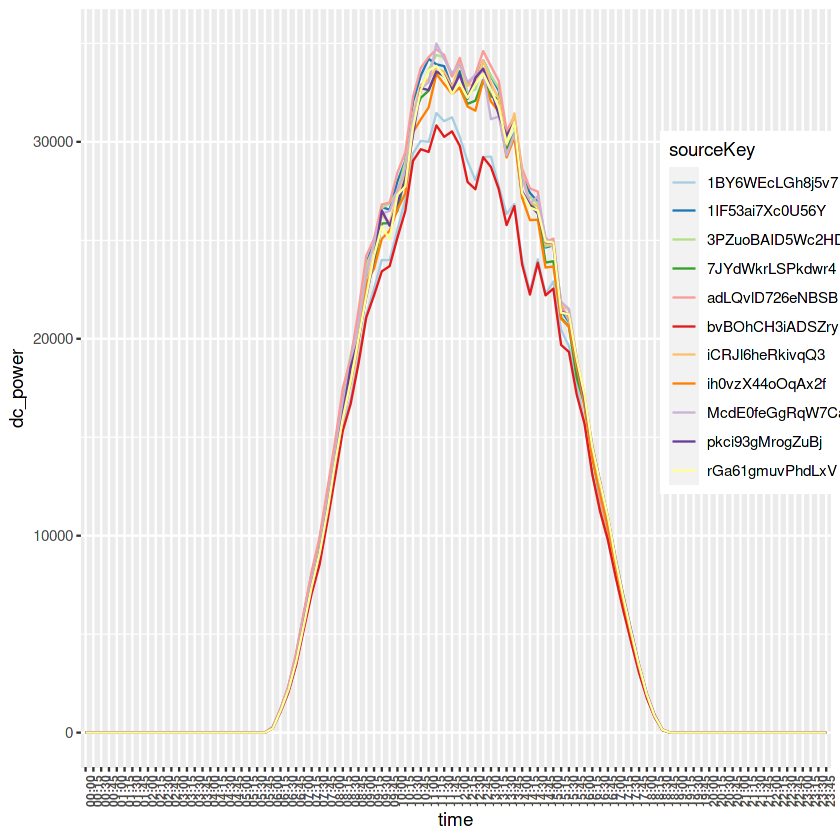
group\_by(time, sourceKey) %>%

summarize(dc\_power = sum(dcPower)) %>%

ggplot(aes(x = time, y = dc\_power,group = sourceKey)) +

geom\_line(aes(color = sourceKey)) +scale\_color\_brewer(palette = "Paired") +

theme(axis.text.x = element\_text(face = "bold", angle = 90, size = 7), legend.position = c(0.9,0.6))



We can see that the solar panels connected to the inverters **"1BY6WEcLGh8j5v7"** and **"bvBOhCH3iADSZry"** are producing less Dc Power as compared to other panels. They might need cleaning or maintenance.

**The sub optimally performing solar panels are panels connected to the inverters "1BY6WEcLGh8j5v7" and "bvBOhCH3iADSZry"\*\***

**Now let’s have a look at how at the proportion of the Dc Power generated by different inverters in the plants**

data <- plant1\_gen %>%

group\_by(sourceKey) %>%

summarize(DcPower = sum(dcPower))

plot\_ly(data, labels = ~sourceKey, values = ~DcPower, type = 'pie') %>%

layout(title = 'DC Power Received in Different Inverters',

xaxis = list(showgrid = FALSE, zeroline = FALSE, showticklabels = FALSE),

yaxis = list(showgrid = FALSE, zeroline = FALSE, showticklabels = FALSE))

**For plant 1 we can see that proportion of Dc Power generated is almost evenly distributed with inverter contributing around 4.2 to 4.7 % of the total power.**

**Now let’s try to find some relationship between the generation data and the sensor data.**

*# Joining the two tables for Plant 1.*

plant1\_comb <- inner\_join(plant1\_gen, plant1\_we, by = "dateTime")

*# Plotting the ambient and module temperature on different days.*

plant1\_we %>%

select(date, time, ambientTemperature, moduleTemperature) %>%

ggplot() +

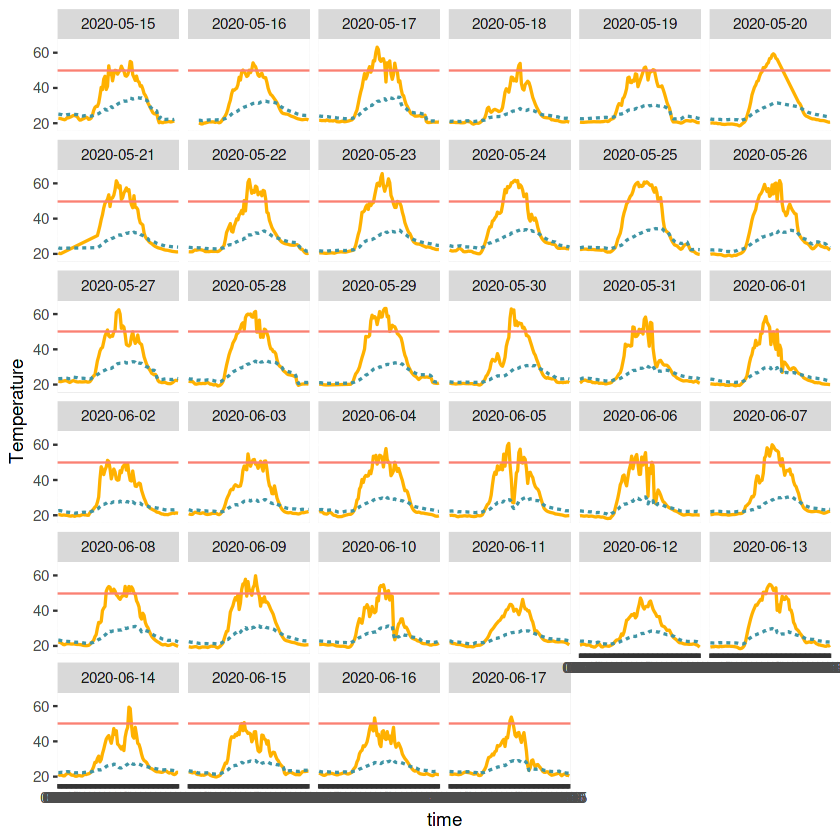
geom\_line(aes(x = time, y = moduleTemperature,group = 1), color = "#ffb100", size = 0.7, linetype = "solid") +

geom\_line(aes(x = time, y = ambientTemperature,group = 2), color = "#4095a5", size = 0.7, linetype = "dashed") +

geom\_hline(yintercept = 50, linetype = "solid", color = "salmon") +

scale\_y\_continuous(name = "Temperature", labels = c(0,20,40,60), breaks = c(0,20,40,60) ) +

facet\_wrap(~date)

****

**REALIZATION ON THE SOLAR GENARATION**

**The horizontal line in the plot is 50 degrees C which is the ideal temperature for the plant working We can see that on some of the days the ambient temperature goes above 50 degrees C which is not good for the equipment in the plant.**

**The plant needs some improvement in the cooling system. They should investigate if the cooling system is working as expected or not.**