Problem Statement:

Isolate the Rooftop Solar generation (Electricity Supply) with a load or group of loads (Electricity Demand) that will ultimately match Demand and Supply as closely as possible. The assumption is that any excess energy will be going into storage systems and any deficit will be taken from the local utility Grid. Objective is to minimize any excess going into storage and limit usage from the local utility grid

Tools used : python(colab), excel

Summary steps:

1. Loading the data:
   First load the data both (rooftops and client wise data) .in google colab (df and df_load)

2. Understanding the data:
   Performed some basic steps (shape,first five rows, checking the columns , dtypes, descriptive stat)

3. Cleaning the data for future:
   Checked null values in both df and df_load.

   Dropped null values in both df and df_load .

   There are two subsystems (1147 , 1035) for solar.
   I filtered them both in `df_solar1 and df_solar2` respectively

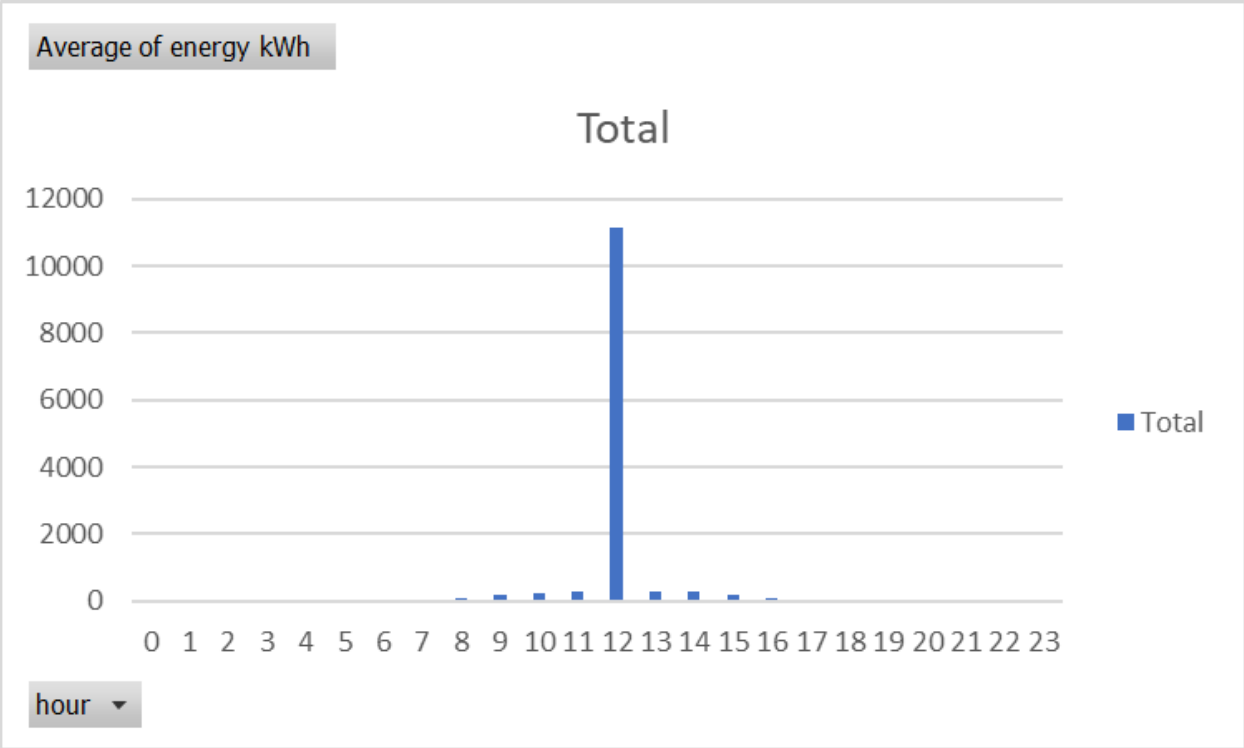
4. Making minute time series to hourly time series in both df_solar1 , df_solar2. Making two new dataframes `hourly_df_solar1`, `hourly_df_solar2` respectively.(with columns date, hour, day of week, quarter, month etc respectively)

5. Making a new column in `df_load_cleaned` namely tenant_floor_building to get the unique load combination present.There are 32 combinations of unique loads with tenant , floor and building.

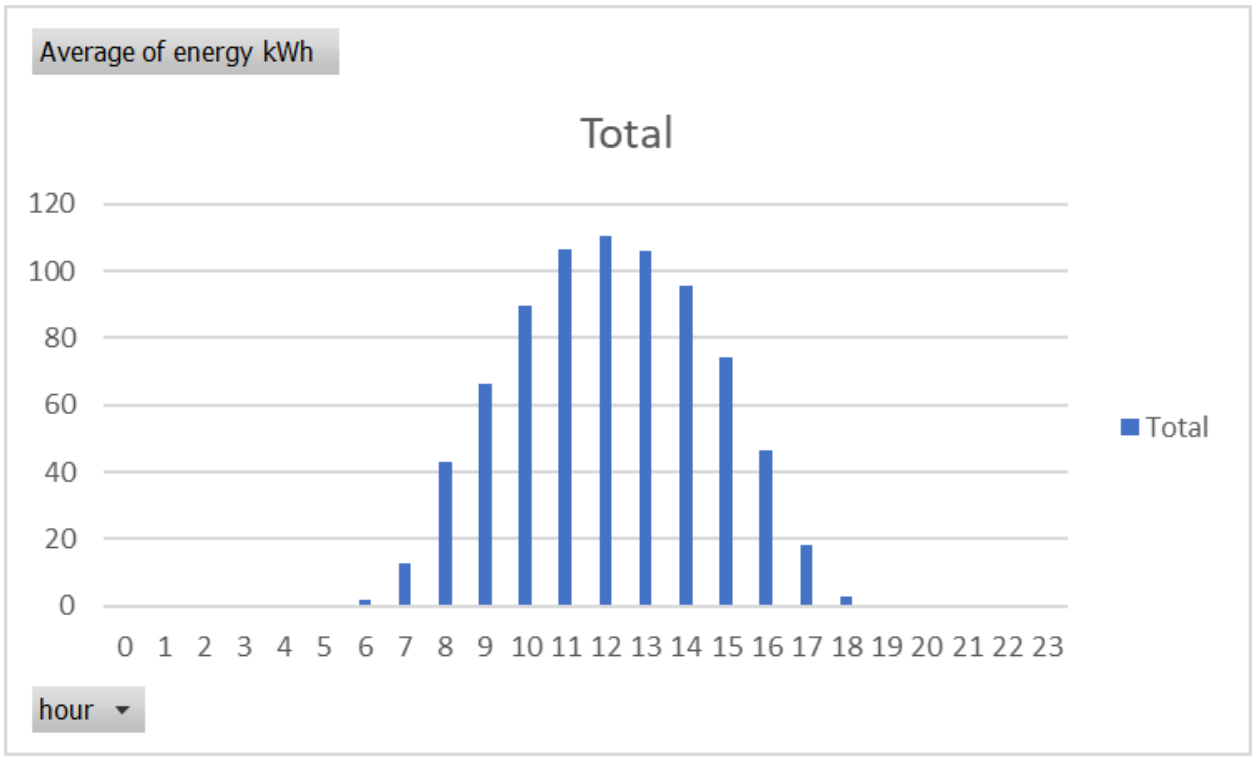
6. In solar_df_first sheet I tried to see the variation of Energy hourly .

It is clearly seen in hour 12 there is an anomaly in solar_df_1 data.

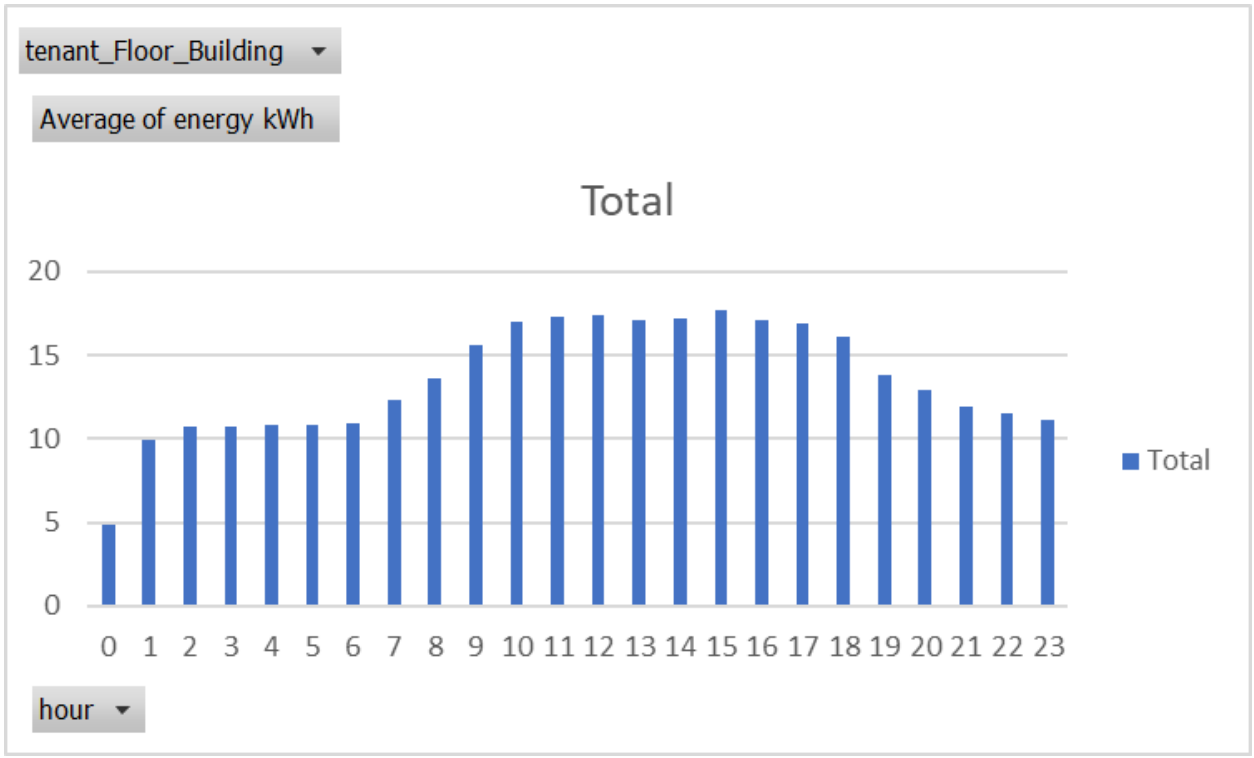
12	11164.37306



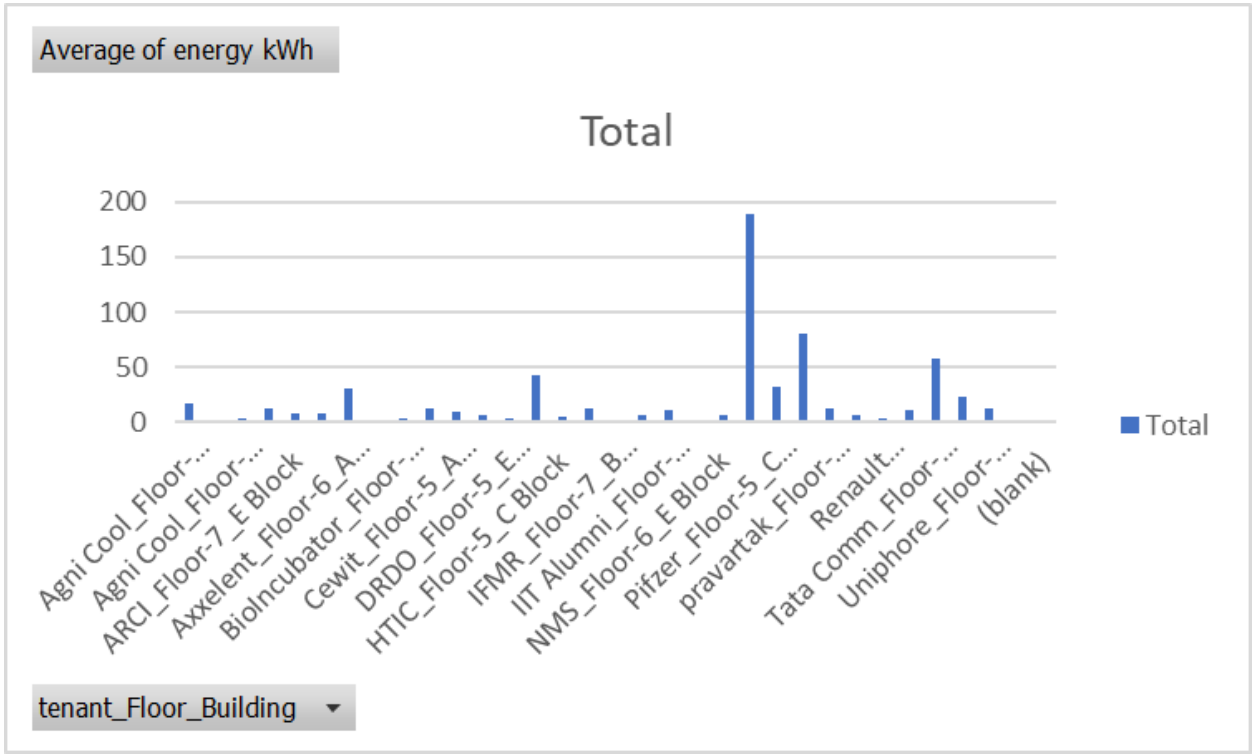
7. In solar_df_second sheet I tried to see variation of energy hourly.



8.
In the load sheet first we have tried to find the variation load consumption vs hour.



Then we have tried to see the variation of load consumption tenant floor building wise.



9.
Then I have built a sheet which has both solar1,solar2, load consumption hourwise. Please refer to the load and solar both hourwise in load excel.

In that sheet , I have tried to make a matrix (34 , 24) . In that I have used sparkline,databar features of excel to understand the distribution of data in a better way.

We can clearly see the solar generation of 0,1,2,3,4,19,20,21,22,23 is zero. For both solar 1 and solar 2

0	1	2	3	4
0	0	0	0	0
0	0	0	0	0

19	20	21	22	23
0	0	0	0	0
0	0	0	0	0

Similarly hour 5,6,7 and 17, 18 is considerably low in terms of solar generation.

10.

As it is an NP hard problem We have to reduce the search space as much as we can to go towards the optimum solution.

To do that We have to sort tenant floor block column in the hour of

[0,1,2,3,4,5,6,7,17,18,19,20,21,22,23] as solar power generation is considerably low in this period.

Load consumption is considerably high in this period for this tenant floor block combinations .They are sorted top to bottom.(consumption wise)

Pifzer_Floor-1_E Block
Pifzer_Ground Floor_MLCP
Tata Comm_Floor-9_E Block
Ginger_Floor-1_E Block
Pifzer_Floor-5_C Block
Axxelent_Floor-6_A Block
TCS_Floor-2_A Block

11. We changed the 12 hour anomaly of solar generation 1 to 303 (approximating looking at other solar generation data).

After filtering out the first three worst tenant floor block combinations we will subset the data on the basis from hour 8 to hour 16(because solar

generation from both solar 1 and solar 2 is statistically significant in this period)

Now we will go to the solar_final excel.

We will use try to find optimum solution based on Total column of subset data first sheet on solar_final excel.

For ease of analysis , we have made final.xlsx file , which we read in colab.

12.

In the first iteration of optimization , we made it a **multiple knapsack**

Problem . We used google or tools . We use linear solver module to find the optimum solution.

Suppose there are two bins(here the solar generation capacity).

We have to maximize the load combination but it cant exceed the capacity of of individual bin.`[1993, 738]`

Simultaneously the load which appears in one bin (one solar panel) can not appear in another bin.(other solar panel). Ideally the scenario is presented here.

We found the coefficient matrix of solar 1

```
[1. 0. 0. 1. 1. 1. 0. 1. 1. 0. 0. 0. 1. 1. 0. 0. 0. 1. 1. 1. 1. 0. 1. 1.
1. 0. 1. 1. 0.]
```

Total bin weight is 1993 (solar1 generation)

Packed weight by load is also 1993

Then we found the coefficient matrix of solar 2.

```
[0. 0. 1. 0. 0. 0. 1. 0. 0. 0. 1. 1. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0.
0. 0. 0. 0. 0.]
```

Total bin weight is 738 (solar 2 generation)

Packed weight by load is also 738

13.

Now we have to backtest our solution in excel to find the loss as lowering of loss is our objective.

We have calculated the dot product of coefficient matrix(both solar 1 and solar2) with each hourly data. We calculated the the difference of solar generation hourly with the result. Then we have took the sum of the absolute value of them.

In the first iteration total loss came out be 3705..

14. In the second iteration optimization we will reduce our search space again After filtering out the next three worst tenant floor block combinations we will subset the data on the basis from hour 8 to hour 16(because solar generation from both solar 1 and solar 2 is statistically significant in this period) .

Ginger_Floor-1_E Block
Pifzer_Floor-5_C Block
Axxelent_Floor-6_A Block

These three will be filtered out.

So total number combination remains (29 -3) = 26 load combination.

Like in the first iteration for the ease of analysis made final_second.xlsx file , which we read in colab.

After running the necessary codes in colab

We have found the coefficient matrix of solar1 load combination and solar 2 load combination.

For solar 1 load combo :

```
[1. 0. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 0.
0. 1.]
```

Total bin weight solar 1 (bin 0) – 1993

Total Load combination of solar 1 (bin 0) - 1951

For solar 2 load combo:

```
[0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1.
1. 0.]
```

Total bin weight solar 2 (bin 1) – 738

Total Load combination of solar 2 (bin 1) - 377

15.

Now we have to backtest our solution in excel to find the loss as lowering of loss is our objective.

We have calculated the dot product of coefficient matrix(both solar 1 and solar2) with each hourly data. We calculated the the difference of solar generation hourly with the result. Then we have took the sum of the absolute value of them.

In the second iteration total loss came out be **3251..**
Which is indeed lower than the previous one.

After further reducing the search space of worst load i.e TCS_Floor-2_A Block I have done the backtest again. But it does not reduce loss. So we will stop here in reducing our search space.

16.

Result and conclusion:

As second iteration of optimization results the lowest loss of in our back testing that will be the result.

solar_coeff_1(1147)	solar_coeff_2(1035)	Solar1
1	0	Agni Cool_Floor-1_A Block
0	0	Agni Cool_Floor-1_E Block
1	0	Agni Cool_Floor-3_E Block
1	0	ARCI_Floor-2_E Block
1	0	ARCI_Floor-7_E Block
1	0	ARCI_Floor-8_E Block
1	0	BioIncubator_Floor-3_E Block
1	0	BioIncubator_Floor-5_C Block
1	0	CaterPillar_Floor-7_E Block
1	0	Cewit_Floor-5_A Block
1	0	Cewit_Floor-6_A Block
1	0	DRDO_Floor-5_E Block

1	0	HTIC_Floor-5_C Block
1	0	ICCW_Ground Floor_B Block
1	0	IFMR_Floor-7_B Block
1	0	IFMR_Floor-8_E Block
1	0	IIT Alumni_Floor-10_C Block
1	0	NMS_Floor-4_E Block
1	0	NMS_Floor-6_E Block
1	0	pravartak_Floor-6_B Block
1	0	Renault Nissan_Floor-4_E Block
1	0	Renault Nissan_Floor-8_E Block
1	0	SGRI_Ground Floor_C Block
0	1	TCS_Floor-2_A Block
0	1	Uniphore_Floor-4_A Block
1	0	Uniphore_Floor-8_E Block

We need to think if we can dynamically schedule load based on consumption rather than static solution.