

오픈소스개발프로젝트

(기말 프로젝트)



이름	박정현
학번	2019037027
담당교수	조 겨 리 교수님
담당조교	유 지 원 선생님

Calculation of Solar Power Generation with Climatic Conditions

<https://github.com/BiteSnail/DataAnalysis>

Data Collection

Climate Information: <https://data.kma.go.kr/cmmn/main.do>

Generation of Solar Power: <https://www.data.go.kr/>

Data Analysis

Table 1 Median of Generation Amount with Date

year	1	2	3	4	5	6	7	8	9	10	11	12
2019									464.7	512.4	446.8	408.4
2020	436.6	538	665	786.5	584.4	612	414	276	530	563.4	463	398.6
2021	320	556.9	618.8	697	600	567.1	459.6	294.8	357	187.2	411.6	370.4
2022	479	378	372.8									

According to this table, April and September are the months that generate most power. Interesting point is that power generation in spring and fall are more than summer.

Figure 1 bar graph with power generation
Monthly solar power median

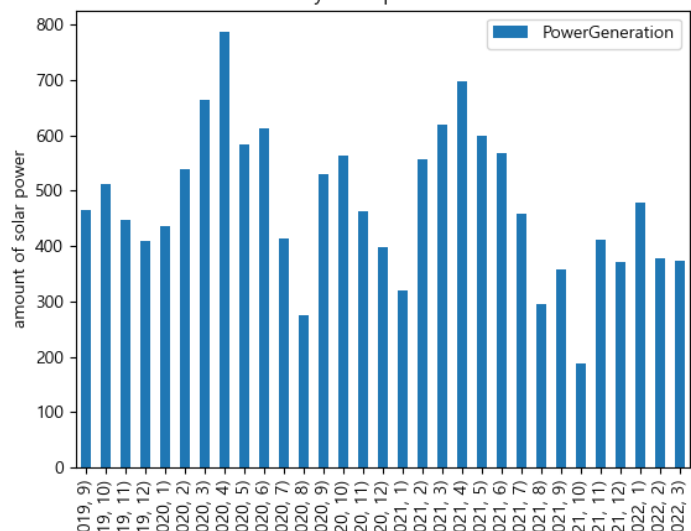
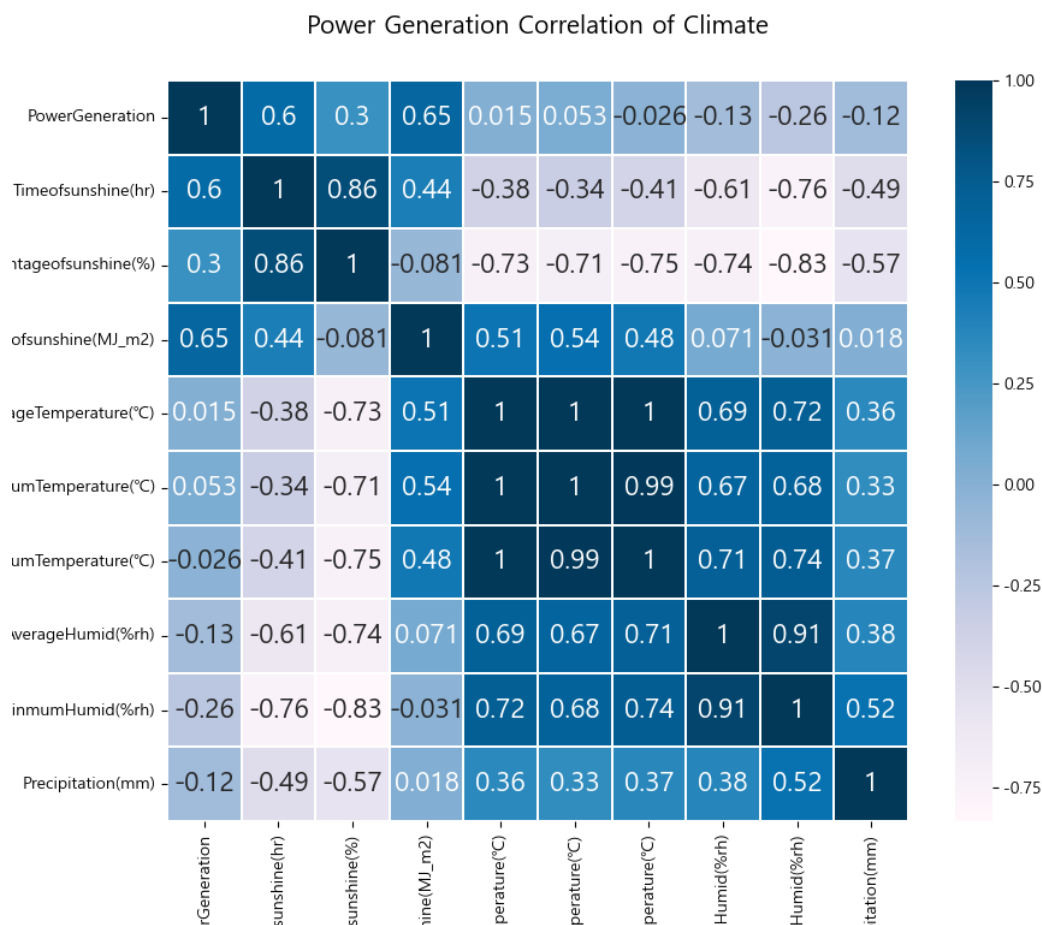
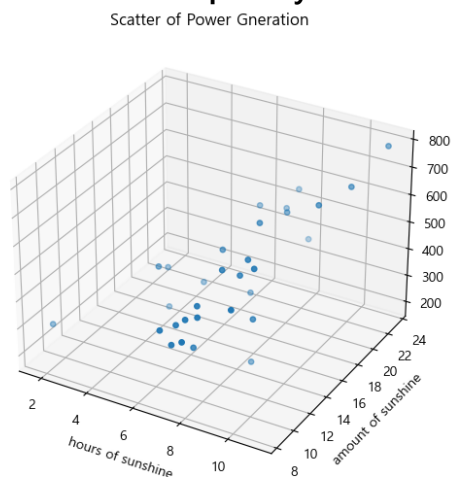


Figure 2 Correlation between solar power generation and climatic conditions



The above figure is a heatmap graph showing the correlation between solar power generation and various climatic conditions. According to this graph, time of sunshine, amount of sunshine, and percentage of sunshine have more correlation than other conditions. And it suggests that in solar power generation, high temperature is not primary condition.

Figure 3 scatter about primary conditions



Left scatter graph is about power generation with two primary conditions that hours of sunshine and amount of sunshine. It seems like that power generation in direct proportion to other two conditions.

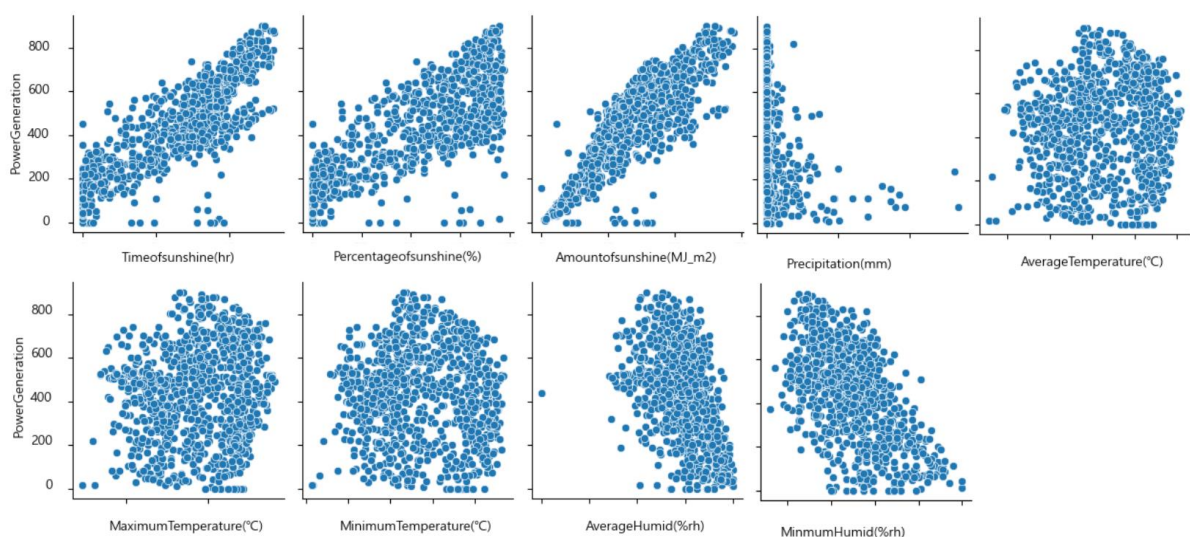
Table 2 Average climatic conditions with power generation in 100 unit

PowerGeneration	PowerGeneration	Timeofsunshine(hr)	Percentageofsunshine(%)	Amountofsunshine(MJ_m2)	AverageTemperature(°C)	MaximumTemperature(°C)	MinimumTemperature(°C)	AverageHumid(%rh)	MinumHumid(%rh)	Precipitation(mm)
0	51.05	1.23	11.38	4.01	11.02	14.47	7.77	86.28	67.08	15.39
1	153.32	1.11	9.78	5.61	13.61	16.92	10.45	83.76	64.71	11.68
2	250.80	2.74	25.14	8.28	12.97	17.25	9.00	78.53	58.37	6.15
3	354.78	5.95	53.57	11.57	11.23	16.38	6.69	73.82	48.59	1.65
4	457.90	7.74	70.17	13.32	10.26	15.66	5.41	70.04	45.27	0.92
5	547.26	8.39	73.78	15.06	10.34	15.95	5.40	68.00	41.45	0.57
6	649.44	9.60	78.49	18.35	13.52	19.45	8.29	69.94	42.85	0.22
7	739.73	10.92	84.17	22.11	15.03	21.62	9.00	66.38	37.45	0.09
8	836.69	11.88	88.60	25.28	12.34	18.95	6.13	65.60	34.22	0.66
9	900.90	12.30	94.65	25.26	8.35	13.50	3.35	59.10	27.00	0.00
51	5153.60	5.20	41.30	15.56	10.50	15.20	6.90	74.50	51.00	0.00
54	5451.00	7.00	61.90	11.81	5.70	11.50	-0.90	56.50	25.00	0.00

I found that high Percentage of sunshine, Amount of sunshine, and time of sunshine lead to increase amount of solar power generation. This insight can find in the table. However, the humidity makes the power generation decrease.

Looking at the statistics, we can know that last two samples at the bottom have some error. See about other data, their power generations are expected 500~600 kW. So, when we use these data, discard two wrong sample. Then we will get more precise result.

Figure 4 scatter about power generation and other conditions except wrong samples



Machine Learning Model and calculation of evaluations

In data preprocessing, first translate columns Korean name into English. Next step is discarding wrong samples. And then chose independent variables. To expect power generation with other variables, x consists of below variables. The last step is that converting x to quadratic polynomial.

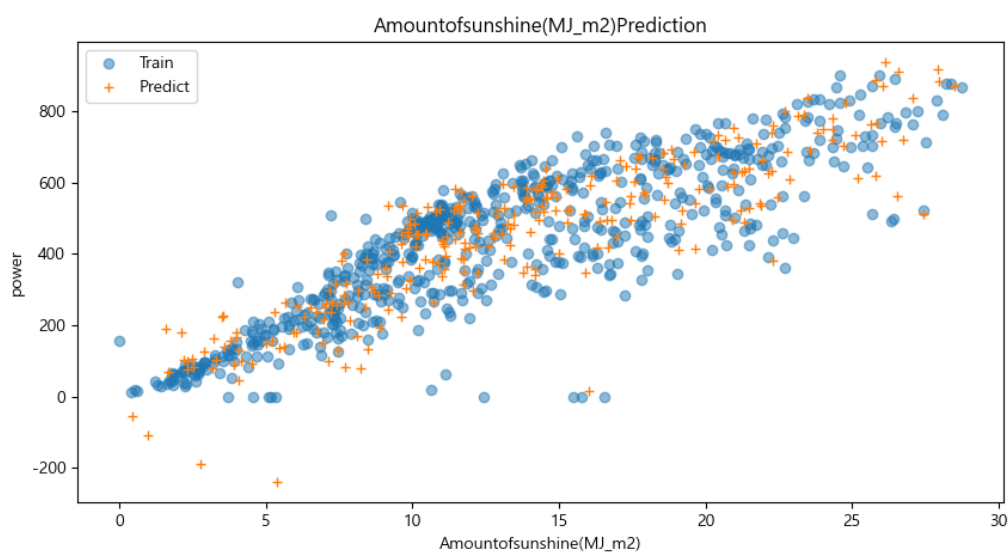
Table 3 Information of independent variables

Data columns (total 9 columns):

#	Column	Count	Non-Null	Dtype
0	Timeofsunshine(hr)	941	non-null	float64
1	Percentageofsunshine(%)	941	non-null	float64
2	Amountofsunshine(MJ_m2)	941	non-null	float64
3	AverageTemperature(°C)	941	non-null	float64
4	MaximumTemperature(°C)	941	non-null	float64
5	MinimumTemperature(°C)	941	non-null	float64
6	AverageHumid(%rh)	941	non-null	float64
7	MinmumHumid(%rh)	941	non-null	int64
8	Precipitation(mm)	941	non-null	float64

Use 'sklearn' to make prediction model. The train and test rate is 7:3. After linear regression fitting, here is scatter graphs between power generation and some conditions. Graphs about other variables can find github result directory.

Figure 5 scatter graph with predict and train



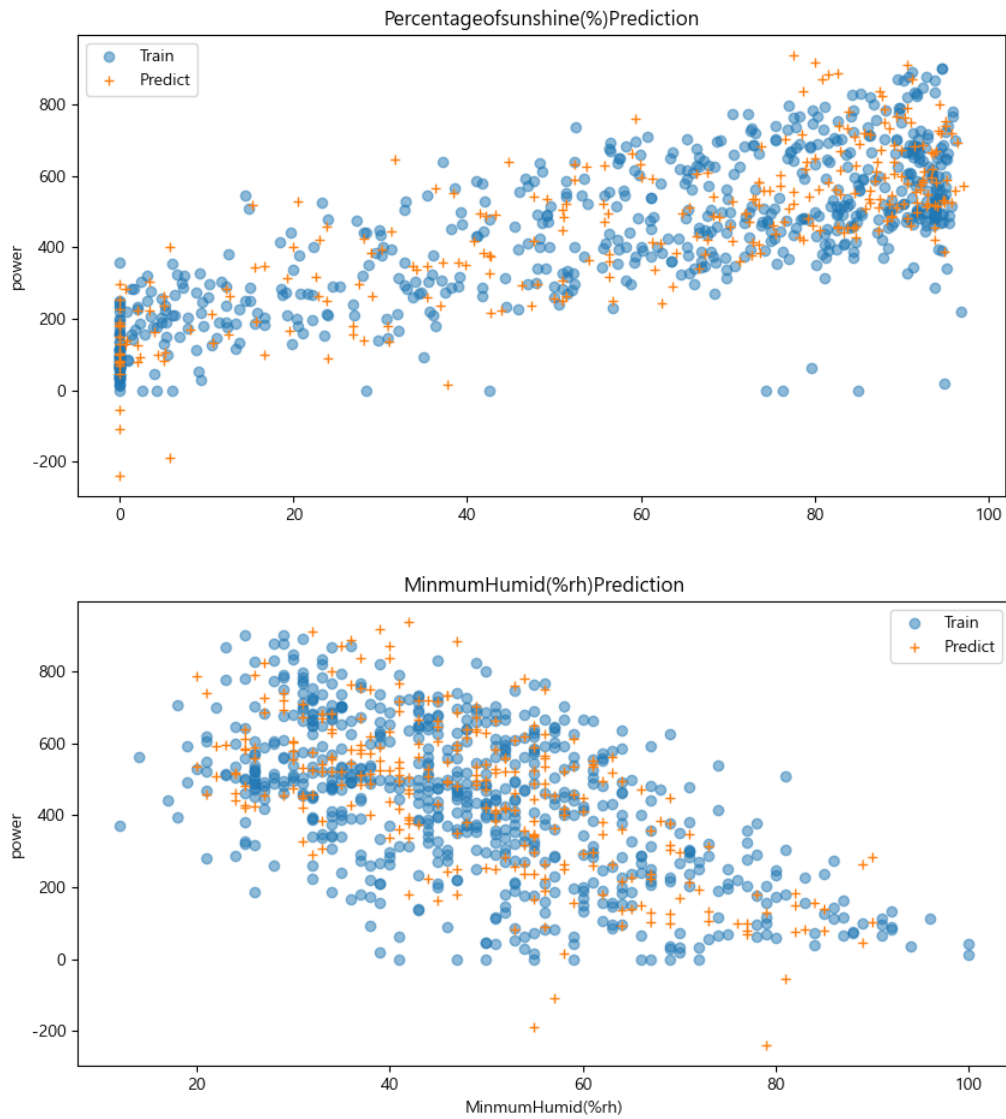
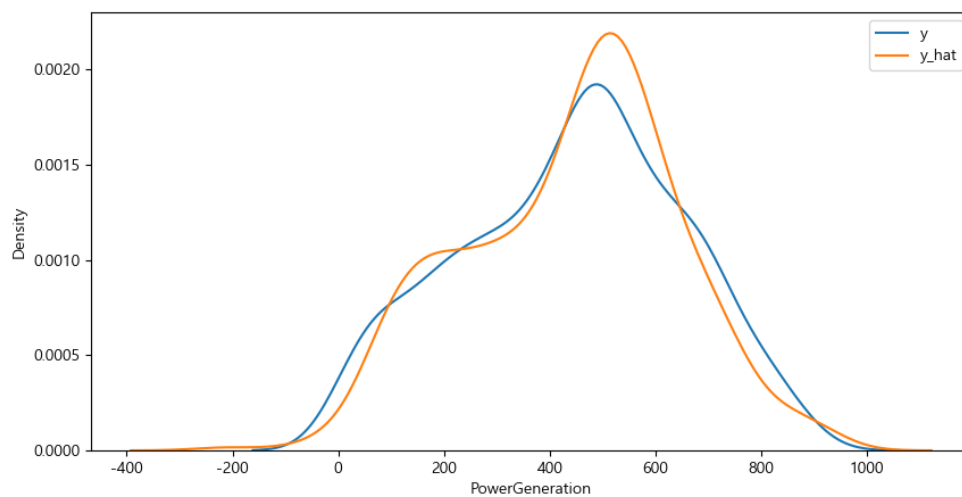


Figure 6 Density about Power generation prediction and answer data



The above graph shows that density about power generation prediction and real data. Compare to two data, it shows that model fitted with train data has about 80% accuracy.

Accuracy: 0.7943065062180832