

DB 응용

20191048 박윤희

INTRO

- 2000년부터 2022년까지 날씨의 변화로 지구온난화의 영향을 찾고자 함.

ABOUT DATA

- 데이터 출처: <https://data.kma.go.kr/data/grnd/selectAsosRltmList.do?pgmNo=36&tabNo=1>
- 위치: 서울(108)/ 위도 : 37.57142 경도 : 126.9658
- 2000년부터 2022년까지 2년마다 7월의 날씨



SURFACE_
ASOS_108_
MI_2000-07
2000-07...



SURFACE_
ASOS_108_
MI_2002-07
2002-07...



SURFACE_
ASOS_108_
MI_2004-07
2004-07...



SURFACE_
ASOS_108_
MI_2006-07
2006-07...



SURFACE_
ASOS_108_
MI_2008-07
2008-07...



SURFACE_
ASOS_108_
MI_2010-07
2010-07...



SURFACE_
ASOS_108_
MI_2012-07
2012-07...



SURFACE_
ASOS_108_
MI_2014-07
2014-07...



SURFACE_
ASOS_108_
MI_2016-07
2016-07...



SURFACE_
ASOS_108_
MI_2018-07
2018-07...



SURFACE_
ASOS_108_
MI_2020-07
2020-07...



SURFACE_
ASOS_108_
MI_2022-07
2022-07...

RAW DATA

B2 2022-07-01 12:01:00 AM												
	A	B	C	D	E	F	G	H	I	J	K	L
1	지점	일시	기온(℃)	누적강수량	풍향(deg)	풍속(m/s)	현지기압(h)	해면기압(h)	습도(%)	일사(MJ/m	일조(Sec)	
2	108	#####	21.9	0	46.8	0.6	1003	1012.9	98.8	0	0	
3	108	#####	21.9	0	45.8	0.6	1003	1012.9	99.1	0	0	
4	108	#####	22	0	41.3	0.9	1003	1012.9	99.3	0	0	
5	108	#####	22	0	53.1	0.9	1003	1012.9	99.2	0	0	
6	108	#####	22	0	42.1	0.6	1003	1012.9	99.1	0	0	
7	108	#####	22	0	45.2	0.6	1003.1	1013	98.8	0	0	
8	108	#####	22	0	45.2	0	1003.1	1013	98.7	0	0	
9	108	#####	22	0	53.3	0.6	1003.1	1013	98.4	0	0	
10	108	#####	22	0	49.7	0.7	1003.1	1013	98.1	0	0	
11	108	#####	22	0	49	0.6	1003.1	1013	98.4	0	0	
12	108	#####	22	0	36	1.3	1003.1	1013	98.8	0	0	
13	108	#####	22	0	22.5	1.5	1003.1	1013	99.5	0	0	
14	108	#####	22	0	56.1	0.2	1003.2	1013.1	98.7	0	0	
15	108	#####	22	0	73.3	0.2	1003.2	1013.1	98.7	0	0	
16	108	#####	22	0	80	0.1	1003.2	1013.1	98.9	0	0	
17	108	#####	22	0	36.7	0.8	1003.2	1013.1	99.1	0	0	
18	108	#####	22	0	35.7	1.6	1003.2	1013.1	99	0	0	
19	108	#####	22	0	34.1	0.4	1003.3	1013.2	98.7	0	0	
20	108	#####	22	0	165.8	0.7	1003.3	1013.2	98.7	0	0	
21	108	#####	22.1	0	125.9	0.1	1003.3	1013.2	99.7	0	0	
22	108	#####	22.1	0	27.5	0.5	1003.3	1013.2	98.8	0	0	
23	108	#####	22.1	0	41	1.6	1003.3	1013.2	98.9	0	0	
24	108	#####	22.1	0	37.1	1.7	1003.3	1013.2	99.5	0	0	
25	108	#####	22.1	0	36.7	1.3	1003.3	1013.2	98.8	0	0	
26	108	#####	22.1	0	47.6	0.7	1003.3	1013.2	98.7	0	0	
27	108	#####	22.1	0	49.3	1.2	1003.3	1013.2	99.3	0	0	
28	108	#####	22.1	0	41.6	1.5	1003.3	1013.2	98.8	0	0	
29	108	#####	22.1	0	48	1.1	1003.3	1013.2	98.5	0	0	
30	108	#####	22.1	0	46.3	0.8	1003.3	1013.2	98.7	0	0	

- rows = 527327, columns = 11
- 일시가 년-월-일-시-분-초까지 있음.
- columns =
[지점, 일시, 기온, 누적강수량, 풍향, 풍속, 현지기압, 해면기압, 습도, 일사, 일조]
- yyyy-08-01 12:00:00 AM 데이터가 있음.

USE DATA

- rows = 372, columns = 8
- 일시를 년-월-일까지 자름
- columns =
[일시, 기온_평균, 기온_최대, 기온_최소, 누적강수량_평균, 습도_평균, 일사_평균, 일조_평균]
- 소수점 둘째자리까지 나타냄 (반올림)
- 8월 데이터 삭제

MySQL

MyUnknownColumn	일시	기온(°C)	누적강수량(mm)	습도(%)	일사(MJ/m^2)	일조(Sec)	max_degree	min_degree
0	2000-07-01 00:00:00	25.09	0	70.21			30.8	0
1	2000-07-02 00:00:00	26.76	0	71.52			33.4	0
2	2000-07-03 00:00:00	27.53	0	70.58			33.2	23.3
3	2000-07-04 00:00:00	27.96	0	66.66			33.8	23.5
4	2000-07-05 00:00:00	27.81	0	67.12			33.9	0
5	2000-07-06 00:00:00	28.53	0	67.22			34.2	23.8
6	2000-07-07 00:00:00	27.74	0	61.47			33.4	24.3
7	2000-07-08 00:00:00	26.72	0	51.93			33	22.1
8	2000-07-09 00:00:00	26.33	0	65.59			30.5	22.7
9	2000-07-10 00:00:00	28.23	0	59.31			34	22.8
10	2000-07-11 00:00:00	23.56	12.38	86.85			27.6	20.1
11	2000-07-12 00:00:00	24.78	4.2	80.17			28.7	20.1
12	2000-07-13 00:00:00	25.36	0	78.96			31	23.1
13	2000-07-14 00:00:00	26.22	0	71.25			30.5	0
14	2000-07-15 00:00:00	26.35	0	71.97			30.7	24
15	2000-07-16 00:00:00	25.33	0	79.17			28.6	23.5
16	2000-07-17 00:00:00	26.78	0	75.96			30.5	25.1
17	2000-07-18 00:00:00	27.36	0.59	75.91			32	25.1
18	2000-07-19 00:00:00	26.85	5.5	82.94			28.9	24.7
19	2000-07-20 00:00:00	28.66	2	71.26			31.8	26.6
20	2000-07-21 00:00:00	27.22	0	78.67			30.6	0

Jupyter notebook

MyUnknownColumn	일시	기온(* C)	누적강수량(mm)	습도(%)	일사(MJ/m^2)	일조(Sec)	max_degree	min_degree
0	2000-07-01	25.09	0.00	70.21				
1	2000-07-02	26.76	0.00	71.52				
2	2000-07-03	27.53	0.00	70.58				
3	2000-07-04	27.96	0.00	66.66				
4	2000-07-05	27.81	0.00	67.12				
...
367	2022-07-27	29.34	0.00	67.06			8.87	
368	2022-07-28	29.56	0.00	63.38			12.18	
369	2022-07-29	30.63	0.00	67.89			10	
370	2022-07-30	31.08	0.38	66.78			11.59	
371	2022-07-31	27.43	8.77	84.59			2.5	

	일조(Sec)	max_degree	min_degree
0		30.8	0.0
1		33.4	0.0
2		33.2	23.3
3		33.8	23.5
4		33.9	0.0
...
367	7233.62	33.4	26.4
368	18812.61	34.4	25.1
369	13989.38	34.3	27.3
370	18104.5	36.1	27.3
371	23.29	29.0	25.4

[372 rows x 9 columns]

PYMYSQL

▶ SELECT * FROM weatherTBL

```
{'MyUnknownColumn': 0, '일시': datetime.datetime(2000, 7, 1, 0, 0), '기온(' C)': 25.09, '누적강수량(mm)': 0.0, '습도(%)': 70.21, '일사(MJ/m^2)': '', '일조(Sec)': '', 'max_degree': 30.8, 'min_degree': 0.0}
{'MyUnknownColumn': 1, '일시': datetime.datetime(2000, 7, 2, 0, 0), '기온(' C)': 26.76, '누적강수량(mm)': 0.0, '습도(%)': 71.52, '일사(MJ/m^2)': '', '일조(Sec)': '', 'max_degree': 33.4, 'min_degree': 0.0}
{'MyUnknownColumn': 2, '일시': datetime.datetime(2000, 7, 3, 0, 0), '기온(' C)': 27.53, '누적강수량(mm)': 0.0, '습도(%)': 70.58, '일사(MJ/m^2)': '', '일조(Sec)': '', 'max_degree': 33.2, 'min_degree': 23.3}
{'MyUnknownColumn': 3, '일시': datetime.datetime(2000, 7, 4, 0, 0), '기온(' C)': 27.96, '누적강수량(mm)': 0.0, '습도(%)': 66.66, '일사(MJ/m^2)': '', '일조(Sec)': '', 'max_degree': 33.8, 'min_degree': 23.5}
{'MyUnknownColumn': 4, '일시': datetime.datetime(2000, 7, 5, 0, 0), '기온(' C)': 27.81, '누적강수량(mm)': 0.0, '습도(%)': 67.12, '일사(MJ/m^2)': '', '일조(Sec)': '', 'max_degree': 33.9, 'min_degree': 0.0}
{'MyUnknownColumn': 5, '일시': datetime.datetime(2000, 7, 6, 0, 0), '기온(' C)': 28.53, '누적강수량(mm)': 0.0, '습도(%)': 67.22, '일사(MJ/m^2)': '', '일조(Sec)': '', 'max_degree': 34.2, 'min_degree': 23.8}
{'MyUnknownColumn': 6, '일시': datetime.datetime(2000, 7, 7, 0, 0), '기온(' C)': 27.74, '누적강수량(mm)': 0.0, '습도(%)': 61.47, '일사(MJ/m^2)': '', '일조(Sec)': '', 'max_degree': 33.4, 'min_degree': 24.3}
{'MyUnknownColumn': 7, '일시': datetime.datetime(2000, 7, 8, 0, 0), '기온(' C)': 26.72, '누적강수량(mm)': 0.0, '습도(%)': 51.93, '일사(MJ/m^2)': '', '일조(Sec)': '', 'max_degree': 33.0, 'min_degree': 22.1}
{'MyUnknownColumn': 8, '일시': datetime.datetime(2000, 7, 9, 0, 0), '기온(' C)': 26.33, '누적강수량(mm)': 0.0, '습도(%)': 65.59, '일사(MJ/m^2)': '', '일조(Sec)': '', 'max_degree': 30.5, 'min_degree': 22.7}
{'MyUnknownColumn': 9, '일시': datetime.datetime(2000, 7, 10, 0, 0), '기온(' C)': 28.23, '누적강수량(mm)': 0.0, '습도(%)': 59.31, '일사(MJ/m^2)': '', '일조(Sec)': '', 'max_degree': 34.0, 'min_degree': 22.8}
{'MyUnknownColumn': 10, '일시': datetime.datetime(2000, 7, 11, 0, 0), '기온(' C)': 23.56, '누적강수량(mm)': 12.38, '습도(%)': 86.85, '일사(MJ/m^2)': '', '일조(Sec)': '', 'max_degree': 27.6, 'min_degree': 20.1}
{'MyUnknownColumn': 11, '일시': datetime.datetime(2000, 7, 12, 0, 0), '기온(' C)': 24.78, '누적강수량(mm)': 4.2, '습도(%)': 80.17, '일사(MJ/m^2)': '', '일조(Sec)': '', 'max_degree': 28.7, 'min_degree': 20.1}
{'MyUnknownColumn': 12, '일시': datetime.datetime(2000, 7, 13, 0, 0), '기온(' C)': 25.36, '누적강수량(mm)': 0.0, '습도(%)': 78.96, '일사(MJ/m^2)': '', '일조(Sec)': '', 'max_degree': 31.0, 'min_degree': 23.1}
{'MyUnknownColumn': 13, '일시': datetime.datetime(2000, 7, 14, 0, 0), '기온(' C)': 26.22, '누적강수량(mm)': 0.0, '습도(%)': 71.25, '일사(MJ/m^2)': '', '일조(Sec)': '', 'max_degree': 30.5, 'min_degree': 0.0}
{'MyUnknownColumn': 14, '일시': datetime.datetime(2000, 7, 15, 0, 0), '기온(' C)': 26.35, '누적강수량(mm)': 0.0, '습도(%)': 71.97, '일사(MJ/m^2)': '', '일조(Sec)': '', 'max_degree': 30.7, 'min_degree': 24.0}
{'MyUnknownColumn': 15, '일시': datetime.datetime(2000, 7, 16, 0, 0), '기온(' C)': 25.33, '누적강수량(mm)': 0.0, '습도(%)': 79.17, '일사(MJ/m^2)': '', '일조(Sec)': '', 'max_degree': 28.6, 'min_degree': 23.5}
{'MyUnknownColumn': 16, '일시': datetime.datetime(2000, 7, 17, 0, 0), '기온(' C)': 26.78, '누적강수량(mm)': 0.0, '습도(%)': 75.96,
```

▶ SELECT * FROM weatherTBL WHERE YEAR(일시) = %s ▷ 2022년 데이터만 추출

```
{'MyUnknownColumn': 352, '일시': datetime.date(2022, 7, 1), '기온(' C)': 26.47, '누적강수량(mm)': 0.0, '습도(%)': 83.41, '일사(MJ/m^2)': '9.86', '일조(Sec)': '10604.36', 'max_degree': 30.9, 'min_degree': 21.9}
{'MyUnknownColumn': 353, '일시': datetime.date(2022, 7, 2), '기온(' C)': 28.65, '누적강수량(mm)': 0.0, '습도(%)': 73.8, '일사(MJ/m^2)': '12.23', '일조(Sec)': '17339.54', 'max_degree': 33.8, 'min_degree': 24.0}
{'MyUnknownColumn': 354, '일시': datetime.date(2022, 7, 3), '기온(' C)': 29.24, '누적강수량(mm)': 0.0, '습도(%)': 68.19, '일사(MJ/m^2)': '12.46', '일조(Sec)': '18500.42', 'max_degree': 34.2, 'min_degree': 24.4}
{'MyUnknownColumn': 355, '일시': datetime.date(2022, 7, 4), '기온(' C)': 29.21, '누적강수량(mm)': 0.0, '습도(%)': 71.88, '일사(MJ/m^2)': '6.88', '일조(Sec)': '5356.92', 'max_degree': 32.6, 'min_degree': 26.4}
{'MyUnknownColumn': 356, '일시': datetime.date(2022, 7, 5), '기온(' C)': 29.21, '누적강수량(mm)': 0.0, '습도(%)': 73.65, '일사(MJ/m^2)': '7.38', '일조(Sec)': '3929.31', 'max_degree': 33.1, 'min_degree': 26.7}
{'MyUnknownColumn': 357, '일시': datetime.date(2022, 7, 6), '기온(' C)': 29.38, '누적강수량(mm)': 1.17, '습도(%)': 75.62, '일사(MJ/m^2)': '6.66', '일조(Sec)': '1582.08', 'max_degree': 33.3, 'min_degree': 26.2}
{'MyUnknownColumn': 358, '일시': datetime.date(2022, 7, 7), '기온(' C)': 28.34, '누적강수량(mm)': 0.01, '습도(%)': 82.74, '일사(MJ/m^2)': '4.06', '일조(Sec)': '687.79', 'max_degree': 31.0, 'min_degree': 26.4}
{'MyUnknownColumn': 359, '일시': datetime.date(2022, 7, 8), '기온(' C)': 26.05, '누적강수량(mm)': 1.59, '습도(%)': 84.03, '일사(MJ/m^2)': '3.12', '일조(Sec)': '1.25', 'max_degree': 27.9, 'min_degree': 24.0}
{'MyUnknownColumn': 360, '일시': datetime.date(2022, 7, 9), '기온(' C)': 26.79, '누적강수량(mm)': 0.1, '습도(%)': 82.2, '일사(MJ/m^2)': '6.13', '일조(Sec)': '2206.67', 'max_degree': 30.9, 'min_degree': 24.0}
{'MyUnknownColumn': 361, '일시': datetime.date(2022, 7, 10), '기온(' C)': 29.05, '누적강수량(mm)': 0.0, '습도(%)': 69.32, '일사(MJ/m^2)': '11.09', '일조(Sec)': '14952.72', 'max_degree': 35.1, 'min_degree': 24.6}
{'MyUnknownColumn': 362, '일시': datetime.date(2022, 7, 11), '기온(' C)': 26.97, '누적강수량(mm)': 3.63, '습도(%)': 74.44, '일사(MJ/m^2)': '6.97', '일조(Sec)': '5393.08', 'max_degree': 30.6, 'min_degree': 24.7}
{'MyUnknownColumn': 363, '일시': datetime.date(2022, 7, 12), '기온(' C)': 26.72, '누적강수량(mm)': 0.01, '습도(%)': 72.91, '일사(MJ/m^2)': '6.02', '일조(Sec)': '2047.92', 'max_degree': 29.4, 'min_degree': 23.2}
{'MyUnknownColumn': 364, '일시': datetime.date(2022, 7, 13), '기온(' C)': 25.12, '누적강수량(mm)': 44.93, '습도(%)': 92.8, '일사(MJ/m^2)': '0.63', '일조(Sec)': '2.04', 'max_degree': 27.1, 'min_degree': 24.0}
{'MyUnknownColumn': 365, '일시': datetime.date(2022, 7, 14), '기온(' C)': 25.73, '누적강수량(mm)': 0.83, '습도(%)': 85.02, '일사(MJ/m^2)': '7.61', '일조(Sec)': '4659.75', 'max_degree': 29.9, 'min_degree': 23.3}
{'MyUnknownColumn': 366, '일시': datetime.date(2022, 7, 15), '기온(' C)': 26.09, '누적강수량(mm)': 0.0, '습도(%)': 76.67, '일사(MJ/m^2)': '11.39', '일조(Sec)': '13844.17', 'max_degree': 30.3, 'min_degree': 23.0}
{'MyUnknownColumn': 367, '일시': datetime.date(2022, 7, 16), '기온(' C)': 24.98, '누적강수량(mm)': 9.77, '습도(%)': 82.96, '일사(MJ/m^2)': '5.42', '일조(Sec)': '2496.92', 'max_degree': 28.2, 'min_degree': 23.2}
```

PYMYSQL

▷ 데이터 타입 확인

```
{'Field': 'MyUnknownColumn', 'Type': 'int', 'Null': 'YES', 'Key': '', 'Default': None, 'Extra': ''}  
{'Field': '일시', 'Type': 'datetime', 'Null': 'YES', 'Key': '', 'Default': None, 'Extra': ''}  
{'Field': '기온(* C)', 'Type': 'double', 'Null': 'YES', 'Key': '', 'Default': None, 'Extra': ''}  
{'Field': '누적강수량(mm)', 'Type': 'double', 'Null': 'YES', 'Key': '', 'Default': None, 'Extra': ''}  
{'Field': '습도(%)', 'Type': 'double', 'Null': 'YES', 'Key': '', 'Default': None, 'Extra': ''}  
{'Field': '일사(MJ/m^2)', 'Type': 'text', 'Null': 'YES', 'Key': '', 'Default': None, 'Extra': ''}  
{'Field': '일조(Sec)', 'Type': 'text', 'Null': 'YES', 'Key': '', 'Default': None, 'Extra': ''}  
{'Field': 'max_degree', 'Type': 'double', 'Null': 'YES', 'Key': '', 'Default': None, 'Extra': ''}  
{'Field': 'min_degree', 'Type': 'double', 'Null': 'YES', 'Key': '', 'Default': None, 'Extra': ''}
```

▷ '일시' datetime -> date로 변경 후 데이터 타입 재확인

```
{'Field': 'MyUnknownColumn', 'Type': 'int', 'Null': 'YES', 'Key': '', 'Default': None, 'Extra': ''}  
{'Field': '일시', 'Type': 'date', 'Null': 'YES', 'Key': '', 'Default': None, 'Extra': ''}  
{'Field': '기온(* C)', 'Type': 'double', 'Null': 'YES', 'Key': '', 'Default': None, 'Extra': ''}  
{'Field': '누적강수량(mm)', 'Type': 'double', 'Null': 'YES', 'Key': '', 'Default': None, 'Extra': ''}  
{'Field': '습도(%)', 'Type': 'double', 'Null': 'YES', 'Key': '', 'Default': None, 'Extra': ''}  
{'Field': '일사(MJ/m^2)', 'Type': 'text', 'Null': 'YES', 'Key': '', 'Default': None, 'Extra': ''}  
{'Field': '일조(Sec)', 'Type': 'text', 'Null': 'YES', 'Key': '', 'Default': None, 'Extra': ''}  
{'Field': 'max_degree', 'Type': 'double', 'Null': 'YES', 'Key': '', 'Default': None, 'Extra': ''}  
{'Field': 'min_degree', 'Type': 'double', 'Null': 'YES', 'Key': '', 'Default': None, 'Extra': ''}
```

PYMYSQL

▶ INSERT into weatherTBL VALUES (500, '2022-11-30', -7, 0, 56, '10', '100', -3, -8)

▷ INSERT 확인

▶ SELECT * FROM weatherTBL WHERE MyUnknownColumn=%s

```
{'MyUnknownColumn': 500, '일시': datetime.date(2022, 11, 30), '기온(* C)': -7.0, '누적강수량(mm)': 0.0, '습도(%)': 56.0, '일사(MJ/m^2)': '10', '일조(Sec)': '100', 'max_degree': -3.0, 'min_degree': -8.0}
```

▶ UPDATE weatherTBL SET 일시 = %s WHERE MyUnknownColumn=%s

▷ UPDATE 확인

▶ SELECT * FROM weatherTBL WHERE MyUnknownColumn=%s

```
{'MyUnknownColumn': 500, '일시': datetime.date(2021, 11, 30), '기온(* C)': -7.0, '누적강수량(mm)': 0.0, '습도(%)': 56.0, '일사(MJ/m^2)': '10', '일조(Sec)': '100', 'max_degree': -3.0, 'min_degree': -8.0}
```

▶ DELETE FROM weatherTBL WHERE MyUnknownColumn=%s

DATA PREPROCESSING

▷ 열 'MyUnknownColumn' 삭제 및 열 'year' 만들기

결측치 처리

○ 2000~2006년의 일사, 일조 데이터가 결측.

▷ **use_df1** => 2000~2006년 행 제외

	일시	기온(°C)	누적강수량(mm)	습도(%)	일사(MJ/m^2)	일조(Sec)	max_degree	min_degree	year
124	2008-07-01	24.60	0.00	65.68	7.64	5549.71	29.3	20.2	2008
125	2008-07-02	22.55	4.68	80.27	0.62	9.06	24.5	21.1	2008
126	2008-07-03	21.96	7.62	83.57	3.77	2144.73	25.1	20.5	2008
127	2008-07-04	24.20	0.01	80.42	1.89	103.13	27.2	21.8	2008
128	2008-07-05	24.44	1.63	83.26	3.49	588.58	28.0	22.0	2008
...
367	2022-07-27	29.34	0.00	67.06	8.87	7233.62	33.4	26.4	2022
368	2022-07-28	29.56	0.00	63.38	12.18	18812.61	34.4	25.1	2022
369	2022-07-29	30.63	0.00	67.89	10.00	13989.38	34.3	27.3	2022
370	2022-07-30	31.08	0.38	66.78	11.59	18104.50	36.1	27.3	2022
371	2022-07-31	27.43	8.77	84.59	2.50	23.29	29.0	25.4	2022

248 rows × 9 columns

▷ **use_df2** => 일사, 일조 열 제외

	일시	기온(°C)	누적강수량(mm)	습도(%)	max_degree	min_degree	year
0	2000-07-01	25.09	0.00	70.21	30.8	0.0	2000
1	2000-07-02	26.76	0.00	71.52	33.4	0.0	2000
2	2000-07-03	27.53	0.00	70.58	33.2	23.3	2000
3	2000-07-04	27.96	0.00	66.66	33.8	23.5	2000
4	2000-07-05	27.81	0.00	67.12	33.9	0.0	2000
...
367	2022-07-27	29.34	0.00	67.06	33.4	26.4	2022
368	2022-07-28	29.56	0.00	63.38	34.4	25.1	2022
369	2022-07-29	30.63	0.00	67.89	34.3	27.3	2022
370	2022-07-30	31.08	0.38	66.78	36.1	27.3	2022
371	2022-07-31	27.43	8.77	84.59	29.0	25.4	2022

372 rows × 7 columns

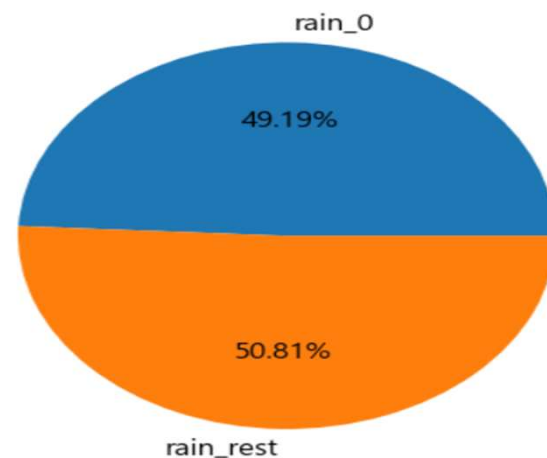
DATA PREPROCESSING

데이터 범주화

```
def rainfunc1(x):  
    if x == 0:  
        return 'rain_0'  
    elif x > 0 and x <= 25:  
        return 'rain_0_25'  
    else:  
        return 'rain_25_'  
  
def rainfunc2(x):  
    if x == 0:  
        return 'rain_0'  
    elif x > 0 and x <= 10:  
        return 'rain_0_10'  
    elif x > 10 and x <= 20:  
        return 'rain_10_20'  
    else:  
        return 'rain_20_'
```

	누적강수량(mm)	rain1	rain2
169	0.00	rain_0	rain_0
206	0.00	rain_0	rain_0
285	0.00	rain_0	rain_0
246	0.00	rain_0	rain_0
176	0.05	rain_0_25	rain_0_10
292	0.00	rain_0	rain_0
45	0.51	rain_0_25	rain_0_10
209	0.00	rain_0	rain_0
193	0.00	rain_0	rain_0
325	0.00	rain_0	rain_0
121	17.68	rain_0_25	rain_10_20
327	0.00	rain_0	rain_0
54	0.03	rain_0_25	rain_0_10
139	4.54	rain_0_25	rain_0_10
183	0.41	rain_0_25	rain_0_10
267	0.00	rain_0	rain_0
161	0.00	rain_0	rain_0
46	0.00	rain_0	rain_0
65	43.29	rain_25_	rain_20_
269	0.00	rain_0	rain_0

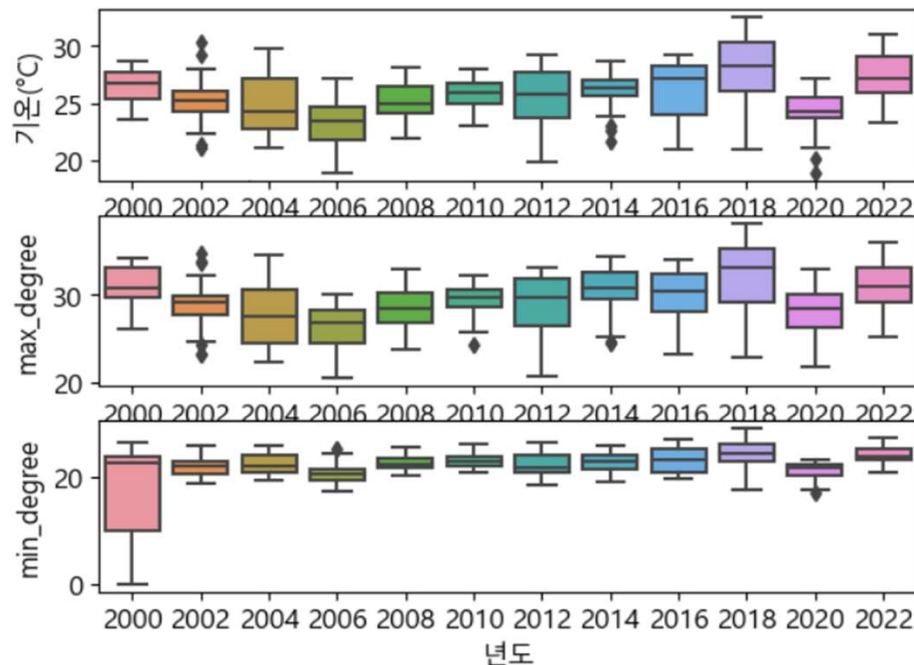
▷ 파이그래프를 위한 누적 강수량 범주화
▷ 누적 강수량 = 0 인 데이터가 49.91%로,
범주화 할 때 x == 0 은 무조건 한 범주로
됨.



VISUALIZATION

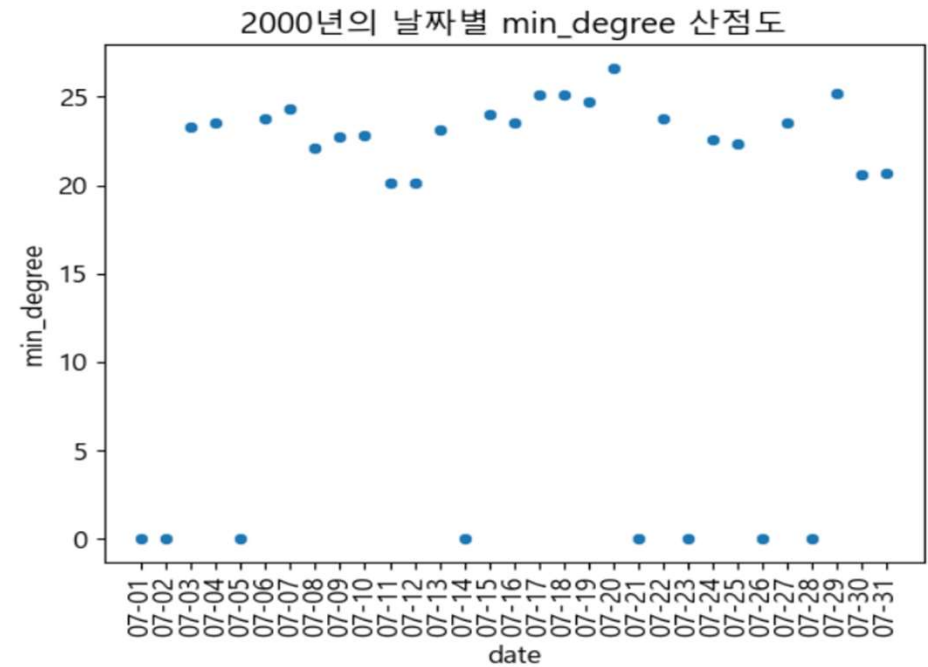
boxplot

▷ 이상치를 찾기 위한 boxplot



☞ 2000년 min_degree에 이상치가 있음을 발견

▷ 2000년 min_degree 산점도로 데이터 확인

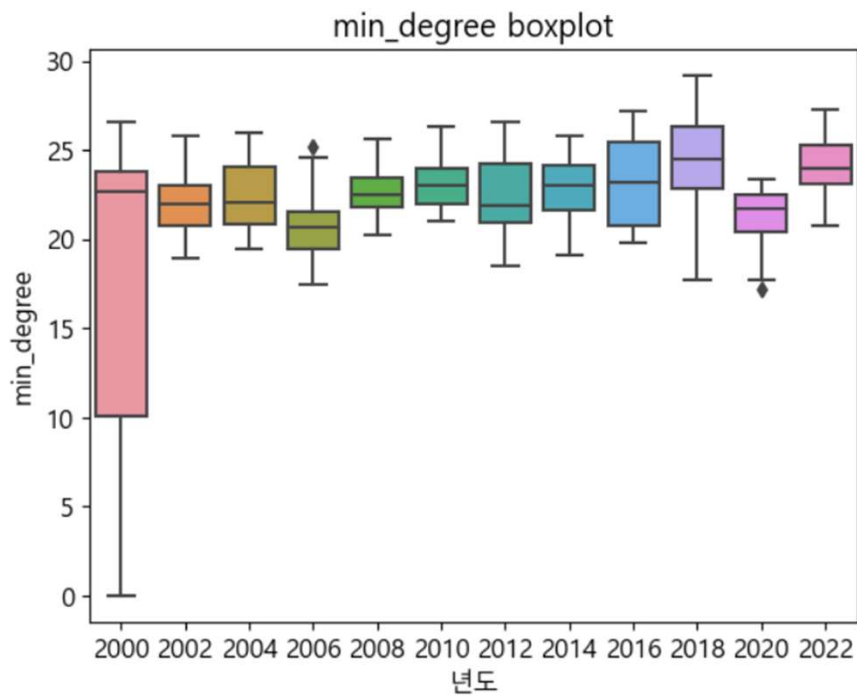


☞ min_degree = 0인 데이터가 이상치임을 확인

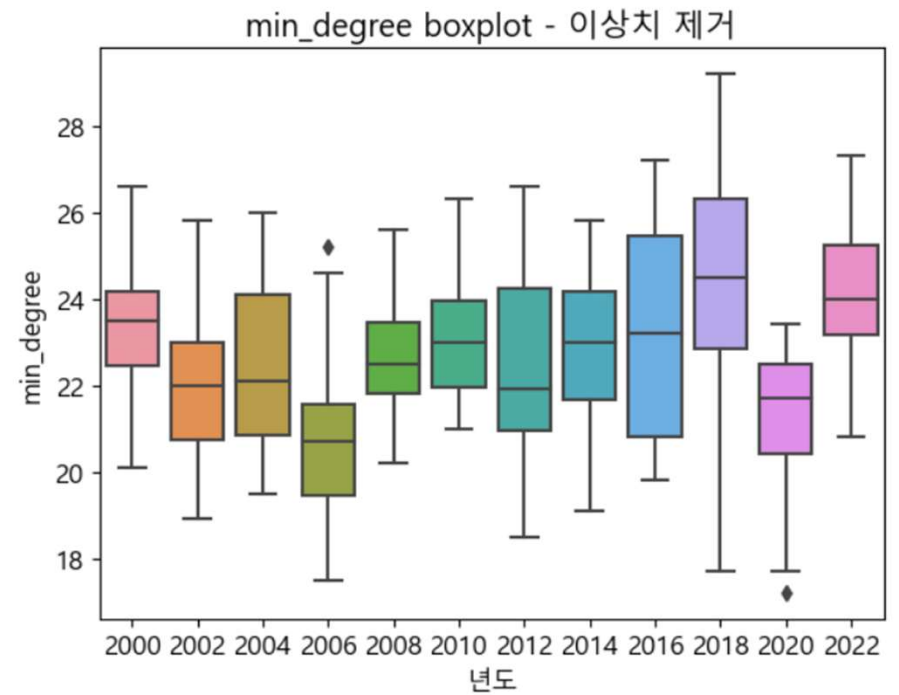
VISUALIZATION

boxplot

▷ 이상치 제거 전 min_degree boxplot

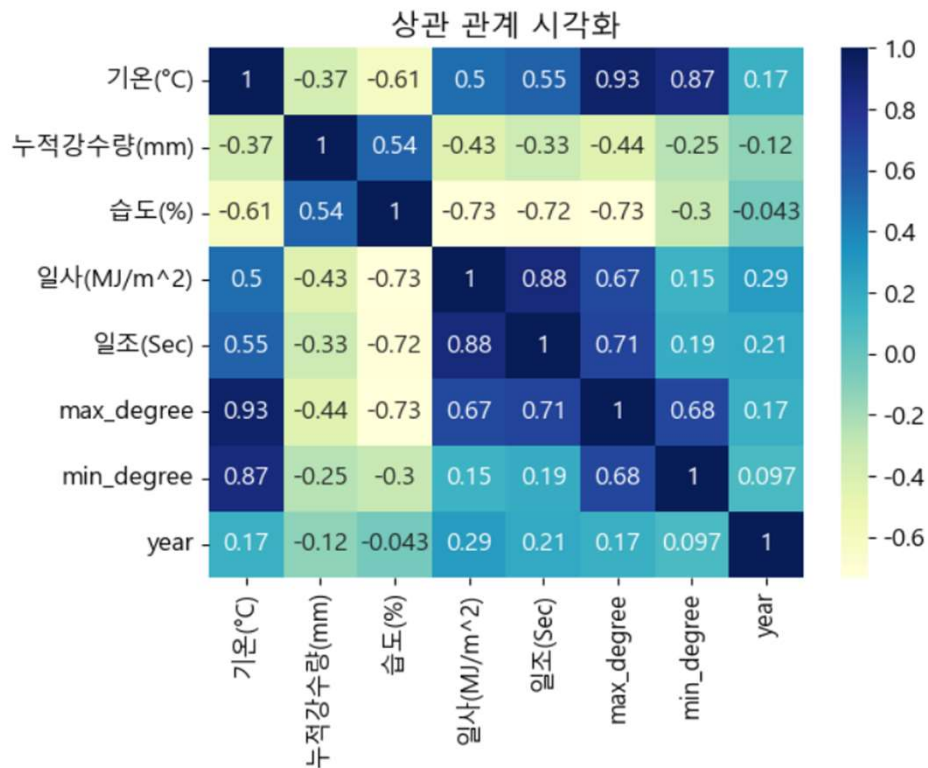


▷ 이상치 제거 후 min_degree boxplot



VISUALIZATION

heatmap - corr

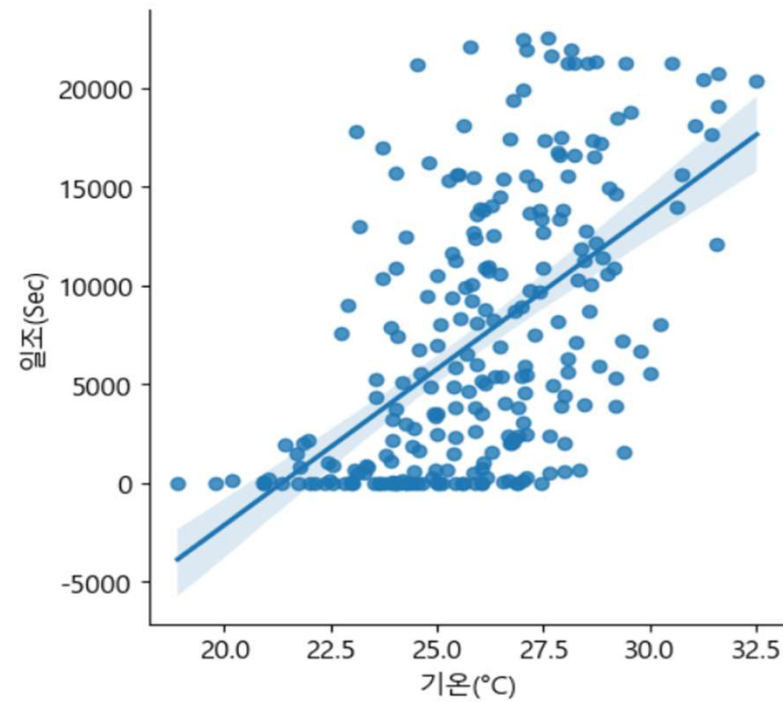


- ☞ 기온은 일별 평균이고 max_degree와 min_degree는 일별 최대 최소이기에 이 셋의 상관성이 깊음.
- ☞ 일사와 일조 또한 비슷한 의미이기에 상관성이 깊음.
- ☞ 습도와 강수량도 상관성이 깊음.

VISUALIZATION

Implot

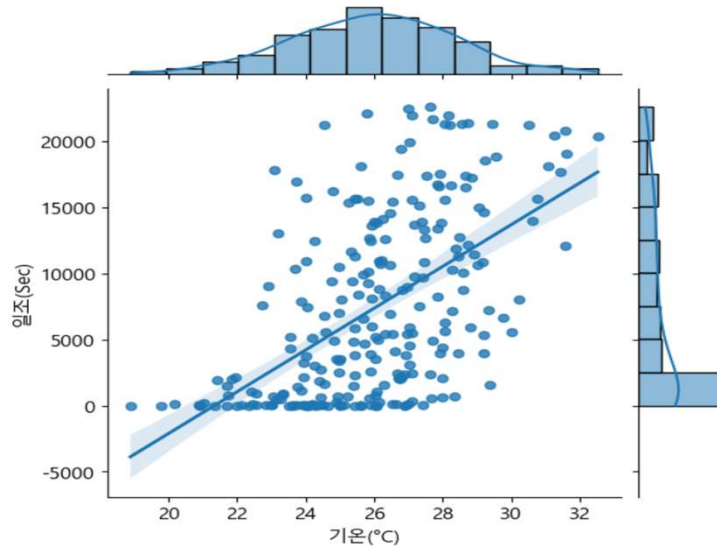
▷ 일사와 기온/ 일조와 기온의 상관성이 0.5이상이기때 Implot을 그려봄.



VISUALIZATION

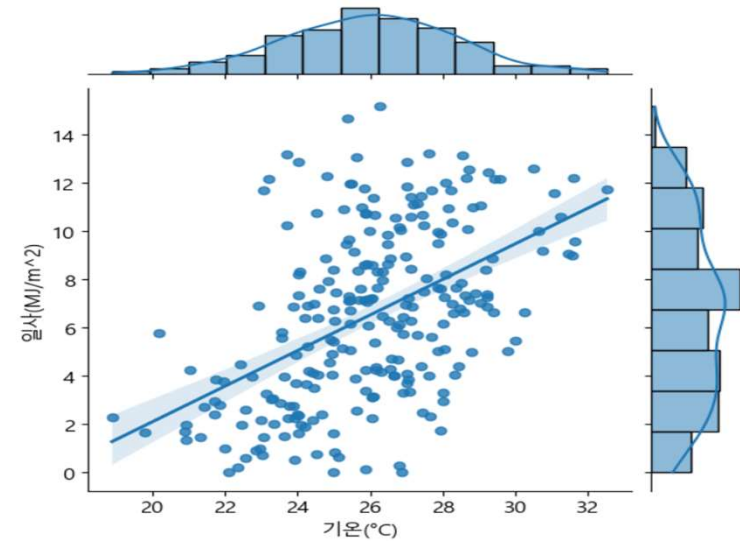
jointplot

▷ 기온과 일조.



☞ 일조는 0에 몰려 있고, 기온은 대략 평균 26 °C에 적절히 분포되어 있음. 또한 기온이 높다는 것은 태양이 내리쬘고 있다는 것이기 때문에 기온이 높아지면, 일조가 길어짐.

▷ 기온과 일사.

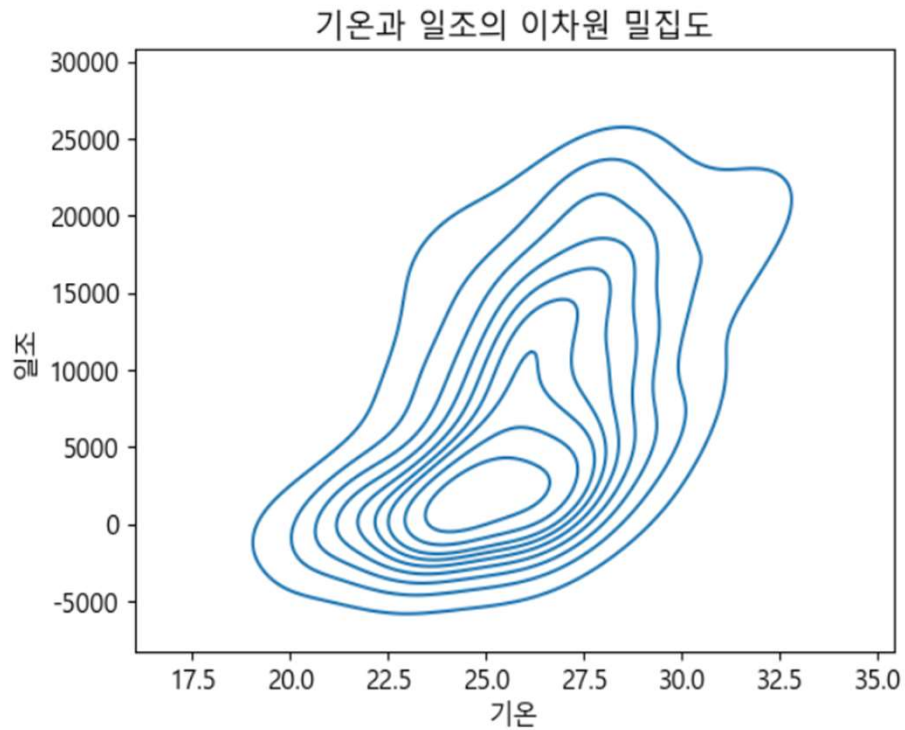


☞ 일조보다는 일사의 선형회귀가 더 일리있어 보임. (둘다 맞지는 않음.)

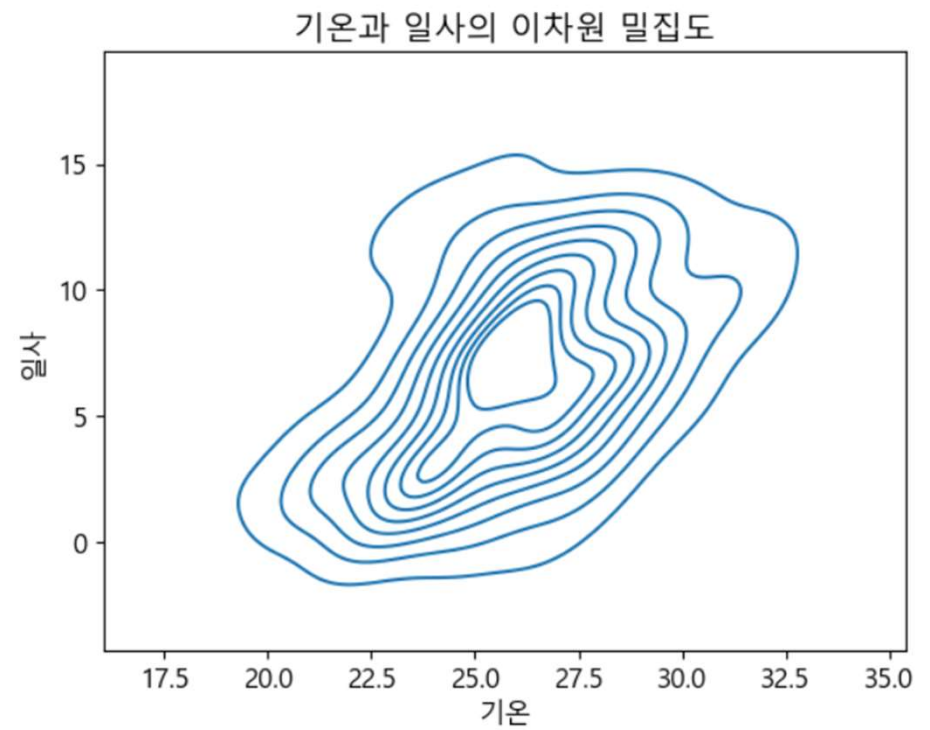
VISUALIZATION

kdeplot

▷ 기온과 일조.



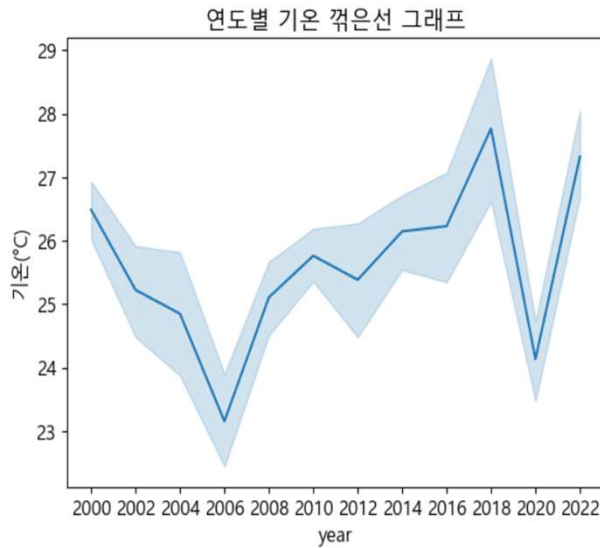
▷ 기온과 일사.



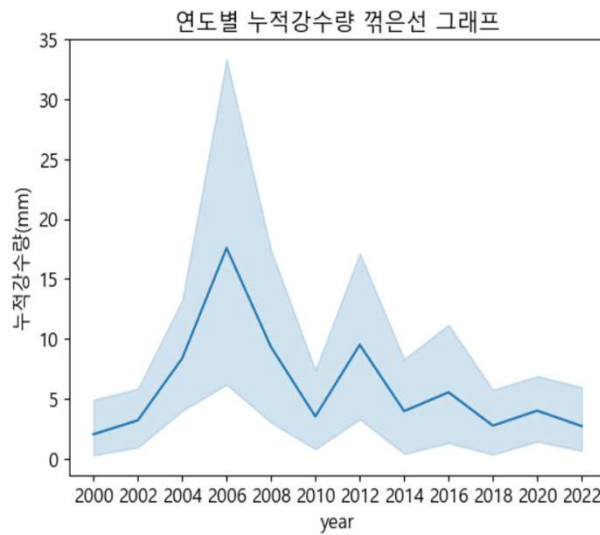
VISUALIZATION

lineplot

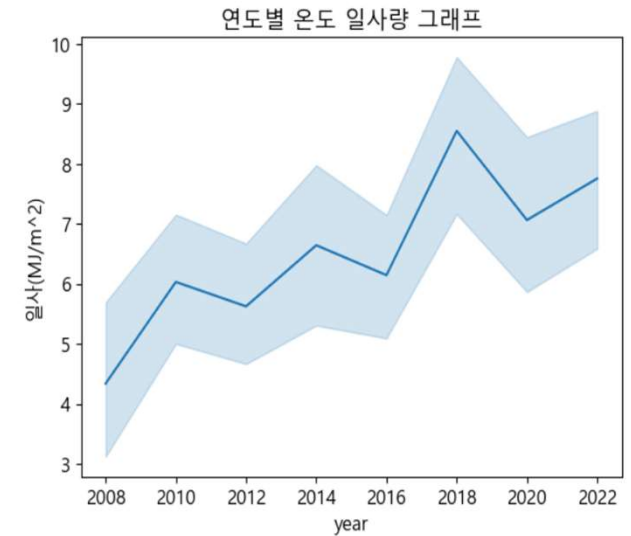
▷ 연도별 기온 추이



▷ 연도별 누적 감수량 추이



▷ 연도별 일사량 추이

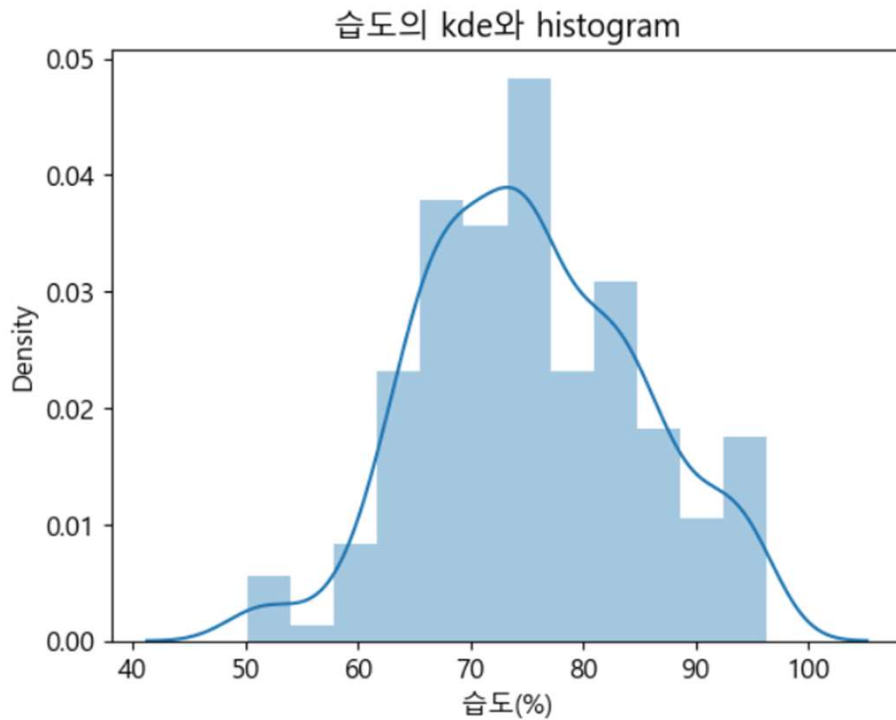


☞ 2006년은 다른 해와 비교했을 때 기온과 감수량 면에서 큰 차이를 보임.

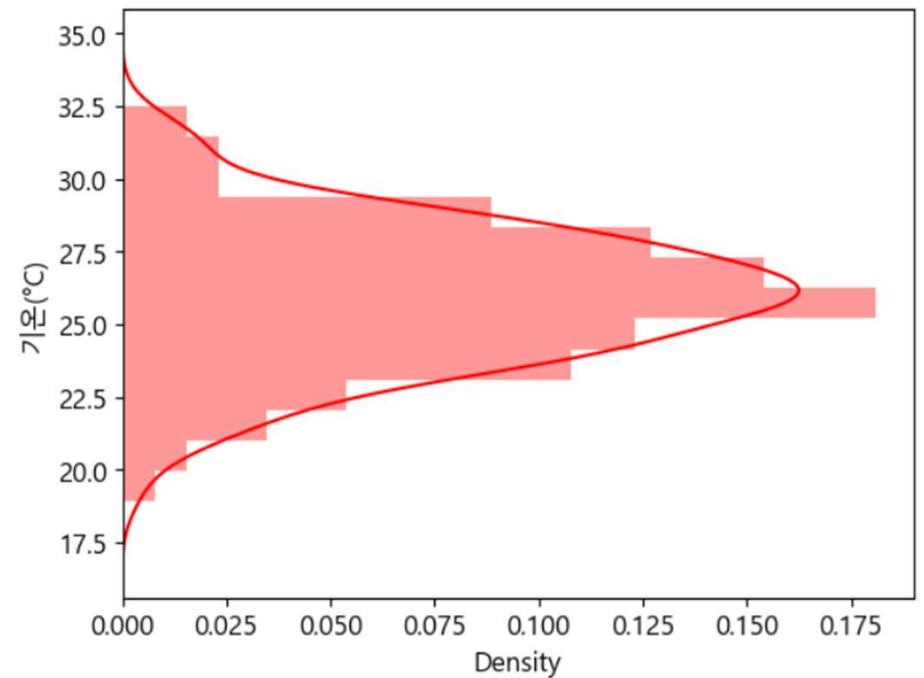
VISUALIZATION

histogram

▷ 습도의 kde와 histogram



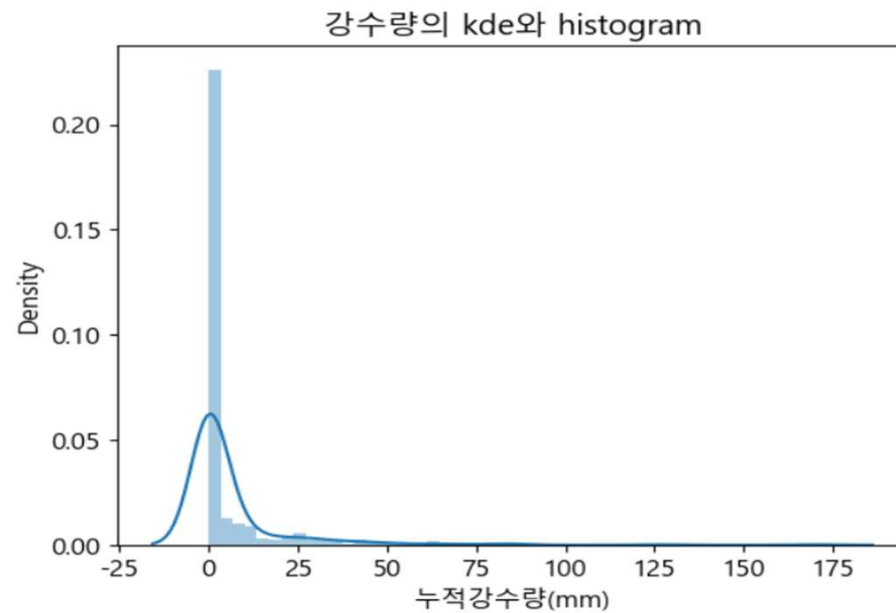
▷ 전치 및 색 변환



VISUALIZATION

histogram

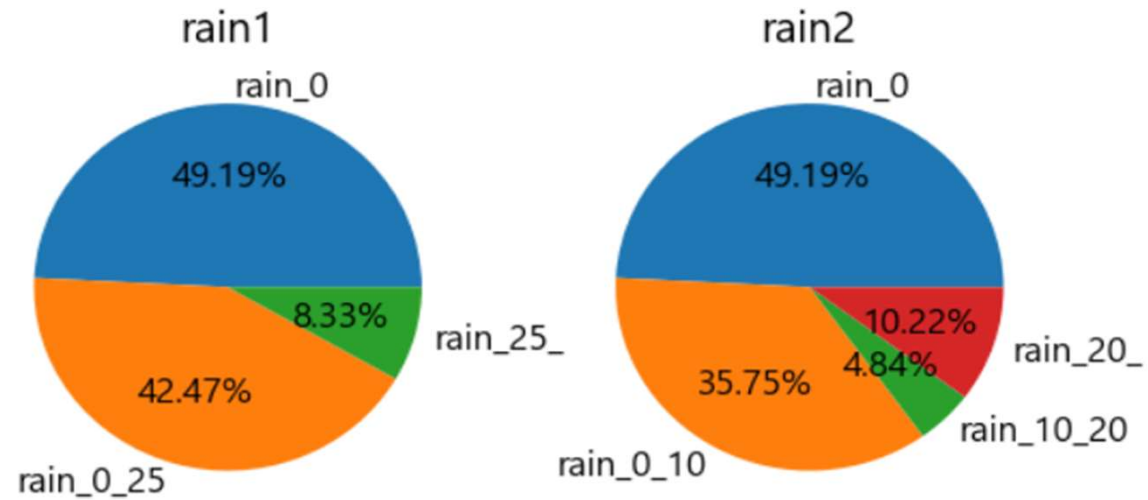
▷ 누적강수량의 kde와 histogram



☞ 누적 강수량은 0에 밀집되어 있음.

VISUALIZATION

piegraph



OUTRO

● 예상했던 것과 다름.

- ◇ 지구온난화로 기온이 전체적으로 올랐을 것을 예상함.
- ◆ 연도별 기온 꺾은 선 그래프를 확인해본 결과, 2008년부터 2018년까지 올랐지만 (2006년은 특이케이스이기에 제외) 2000년부터 비교한다면, 전체적인 기온이 오르지 않음.
- ◇ 강수량 또한 줄어들었을 것이라고 예상함.
- ◆ 강수량도 줄어드는 추세를 보여주진 않음.
- ◆ 연도별 일사량은 오르는 추이를 보여줌.
- ☞ 오존층의 파괴로 일사량이 올랐나 추측해볼 수 있음.

OUTRO

아쉬운 점

- (1) yyyy-07-01 12:01:00 AM부터 시작이라는 것을 알았음. yyyy-07-02부터는 12:00:00 AM이 존재함. 이를 채우기 위해 6월 데이터를 사용하던지, 틀린값이라 판단하였던 yyyy-08-01 12:00:00 AM데이터를 삭제하지 말고 yyyy-07-dd 12:01:00 AM ~ yyyy-07-(dd+1) 12:00:00 AM의 데이터를 사용하였어야 함.
- (2) Input을 이용하여 INSERT문을 실행하는 것을 결국 못해서, VALUES를 일일이 다씀.
- (3) 누적 강수량을 범주화한 rain1과 rain2의 정확한 기준이 없이 범주화 함. (0제외)

※ (2)에 대한 오류 – 구글링 결과, 또는 “를 쓰지 않았을 가능성이 크다 하였지만 여러 번 확인하여 보여도 오타를 찾지 못함. 또한, 그대로 복붙하여 values () 하였을 경우 제대로 작동함.

```
ProgrammingError: (1064, "You have an error in your SQL syntax; check the manual that corresponds to your MySQL server version for the right syntax to use near ', '2022-11-30', -7, 0, 56, '10', '100', -3, -8)' at line 1")
```