

# 기상 데이터 전처리

- 기상청에서 제공하는 제주도 읍면동별 기상 데이터
- 일별 '풍속', '기온', '습도', '강수' 데이터 평균으로 되어있다.

```
In [1]: # 필요 라이브러리 로드
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

import warnings
warnings.filterwarnings(action='ignore')
```

```
In [2]: # 파일 불러오기
df = pd.read_csv('2018_2021_날씨.csv', encoding='cp949')
df.head()
```

```
Out[2]:
```

	y_m	location	풍속	기온	습도	강수
0	2018-01-01	건입동	1.616667	5.254167	54.041667	0.0
1	2018-01-01	구좌읍	3.404167	4.575000	63.333333	0.0
2	2018-01-01	남원읍	3.070833	7.212500	51.375000	0.0
3	2018-01-01	노형동	1.641667	5.379167	54.083333	0.0
4	2018-01-01	대륜동	1.108333	8.287500	53.625000	0.0

```
In [3]: # 'y_m'(년_월) 열 데이터 변환
df["y_m"] = df.apply(lambda x: x.y_m[:7], axis = 1)
df.head()
```

```
Out[3]:
```

	y_m	location	풍속	기온	습도	강수
0	2018-01	건입동	1.616667	5.254167	54.041667	0.0
1	2018-01	구좌읍	3.404167	4.575000	63.333333	0.0
2	2018-01	남원읍	3.070833	7.212500	51.375000	0.0
3	2018-01	노형동	1.641667	5.379167	54.083333	0.0
4	2018-01	대륜동	1.108333	8.287500	53.625000	0.0

```
In [4]: # 월별 행정동별 그룹화 진행 (mean)
df_mean = df.groupby(['y_m', 'location']).mean().reset_index()
df_mean
```

```
Out[4]:
```

	y_m	location	풍속	기온	습도	강수
0	2018-01	건입동	2.593011	4.800134	66.350806	0.090323
1	2018-01	구좌읍	3.899194	4.218952	74.362903	0.092608
2	2018-01	남원읍	2.619220	6.256586	62.920699	0.061022

	y_m	location	풍속	기온	습도	강수
3	2018-01	노형동	2.991129	4.887231	68.990591	0.057796
4	2018-01	대륜동	1.831317	8.004301	66.793011	0.095430
...	...	...	...	...	...	...
1717	2021-06	표선면	2.315000	22.187778	89.306944	0.307917
1718	2021-06	한경면	4.363194	22.407500	93.737500	0.242361
1719	2021-06	한림읍	2.888889	22.537500	92.995833	0.225278
1720	2021-06	화북동	2.432778	23.299861	75.793056	0.168611
1721	2021-06	효돈동	1.550833	22.751111	86.093056	0.277500

1722 rows × 6 columns

```
In [5]: # 월별 행정동별 그룹화 진행 (sum)
df_sum = df.groupby(['y_m', 'location']).sum().reset_index()
df_sum
```

	y_m	location	풍속	기온	습도	강수
0	2018-01	건입동	80.383333	148.804167	2056.875000	2.800000
1	2018-01	구좌읍	120.875000	130.787500	2305.250000	2.870833
2	2018-01	남원읍	81.195833	193.954167	1950.541667	1.891667
3	2018-01	노형동	92.725000	151.504167	2138.708333	1.791667
4	2018-01	대륜동	56.770833	248.133333	2070.583333	2.958333
...	...	...	...	...	...	...
1717	2021-06	표선면	69.450000	665.633333	2679.208333	9.237500
1718	2021-06	한경면	130.895833	672.225000	2812.125000	7.270833
1719	2021-06	한림읍	86.666667	676.125001	2789.875000	6.758333
1720	2021-06	화북동	72.983333	698.995835	2273.791667	5.058333
1721	2021-06	효돈동	46.525000	682.533334	2582.791667	8.325000

1722 rows × 6 columns

```
In [6]: # 풍속, 기온, 습도는 월별 평균값으로 // 강수는 월별 총 강수량으로 추출
new_df = df_mean.loc[:, ["y_m", "location", "풍속", "기온", "습도"]]
new_df["강수"] = df_sum.loc[:, "강수"]
new_df
```

	y_m	location	풍속	기온	습도	강수
0	2018-01	건입동	2.593011	4.800134	66.350806	2.800000
1	2018-01	구좌읍	3.899194	4.218952	74.362903	2.870833
2	2018-01	남원읍	2.619220	6.256586	62.920699	1.891667

	y_m	location	풍속	기온	습도	강수
<b>3</b>	2018-01	노형동	2.991129	4.887231	68.990591	1.791667
<b>4</b>	2018-01	대륜동	1.831317	8.004301	66.793011	2.958333
...	...	...	...	...	...	...
<b>1717</b>	2021-06	표선면	2.315000	22.187778	89.306944	9.237500
<b>1718</b>	2021-06	한경면	4.363194	22.407500	93.737500	7.270833
<b>1719</b>	2021-06	한림읍	2.888889	22.537500	92.995833	6.758333
<b>1720</b>	2021-06	화북동	2.432778	23.299861	75.793056	5.058333
<b>1721</b>	2021-06	효돈동	1.550833	22.751111	86.093056	8.325000

1722 rows × 6 columns

```
In [7]: new_df.to_csv("기상데이터_전처리.csv", encoding = "cp949", index = False)
```