

# 기온 변화가 수도 사용량에 미치는 영향

## 목표

계절에 따른 기온 변화가 수도 사용량에 미치는 영향 분석

## 가설

여러 가지 요인이 수도 사용에 영향을 미치겠지만, 위생 관리, 수분 섭취, 여름철 온도 조절 등 여러 이유로 기온과 수도 사용량 사이에 연관관계가 존재할 것으로 예상했습니다.

## 데이터 수집

1. 한국소비자원 수도 에너지 사용량 데이터

출처 : 공공 데이터 포털

<https://www.data.go.kr/data/15100600/fileData.do>

2. 기상청 기온 분석 데이터

출처 : 기상청

<https://data.kma.go.kr/stcs/grnd/grndTaList.do>

## 데이터 확인

파이썬을 활용하여 데이터를 확인했습니다.

수도 사용량 :

	연도	월	상하수도사용량
0	2013	1	1287
1	2013	2	0
2	2013	3	1219
3	2013	4	0
4	2013	5	1434

[145 rows x 3 columns]

기온 데이터 :

	날짜	지점	평균기온	최저기온	최고기온
0	Wt2014-07-01	108	26.0	21.1	31.2
1	Wt2014-07-02	108	25.7	22.1	31.1
2	Wt2014-07-03	108	22.6	21.4	24.7
3	Wt2014-07-04	108	26.8	21.6	32.3
4	Wt2014-07-05	108	25.7	22.3	29.7

[3926 rows x 5 columns]

수도 사용량 데이터: 각 년도, 월별로 수도 사용량이 기록되어 있습니다.

기온 데이터: 날짜별로 평균, 최저, 최고 기온이 기록되어 있음을 확인할 수 있습니다.

## 결측치 확인 및 데이터 전처리

수도 사용량 결측치 :

연도 0  
월 0  
상하수도사용량 0  
dtype: int64

기온 데이터 결측치 :

날짜 0  
지점 0  
평균기온 0  
최저기온 1  
최고기온 1  
dtype: int64

수도 사용량 0인 데이터 :

	연도	월	상하수도사용량
1	2013	2	0
3	2013	4	0
5	2013	6	0
7	2013	8	0
9	2013	10	0
11	2013	12	0
13	2014	2	0
15	2014	4	0
17	2014	6	0

기온 데이터 0인 데이터 :

	날짜	지점	평균기온	최저기온	최고기온
596	Wt2016-02-17	108	0.0	-5.2	5.4
1618	Wt2018-12-05	108	0.0	-4.3	3.4
1629	Wt2018-12-16	108	0.0	-2.6	2.1
2429	Wt2021-02-23	108	0.0	-3.1	4.3
3087	Wt2022-12-13	108	0.0	-6.4	3.1

결측치 확인 결과 기온 데이터에 결측치가 존재하고

수도 사용량이 2014년 6월까지 격월로 기록되어 있습니다.

우선 결측치가 어떤 값인지 확인했습니다.

기온 데이터 결측 행 :

	날짜	지점	평균기온	최저기온	최고기온
1199	WWt2017-10-12	108	11.4	8.8	NaN
2960	WWt2022-08-08	108	26.8	NaN	28.4

확인 결과, 두 날짜의 데이터에서 최저 및 최고 기온이 누락되어 있었으며,

이 결측치는 월별 기온 평균값 산출에 미치는 영향이 미미하다고 판단되어 제거했습니다.

위의 과정으로

1. 2014년 이전 데이터 제거 기온과 수도 사용량 간의 상관관계를 분석하기 위해, 상대적으로 적은 2년 치 데이터는 분석에 큰 영향을 미치지 않는다고 판단하여 2014년 이전 데이터를 제거했습니다.
2. 25년 수도 사용량 제거 2025년의 수도 사용량 데이터는 1분기 데이터만 존재하여 비교 분석 목적에 부족하므로 제거했습니다.
3. 기온 데이터 결측행 제거 결측치가 있는 두 날짜의 데이터를 제거하였으며, 이로 인한 월별 기온 평균값 산출에는 영향이 없다고 판단했습니다.

```
# 수도 사용량 데이터 0이 포함된 연도 제거
water_df = water_df[water_df["연도"] > 2014 ]
# 25년 이후의 데이터는 측정하기엔 부족하므로 제거
water_df = water_df[water_df["연도"] != 2025]
# 기온 데이터 결측치 제거
temp_df = temp_df.dropna(subset=['평균기온', '최저기온', '최고기온'])
```

## 데이터 병합

데이터 병합을 위해, 온도 데이터의 '날짜'를 datetime 형식으로 변환한 후 연도와 월을 추출했습니다.

```
# 기온 데이터 전처리: '날짜' 열에서 공백 제거 후 datetime 변환
temp_df['날짜'] = temp_df['날짜'].str.strip()
temp_df['날짜'] = pd.to_datetime(temp_df['날짜'], format='%Y-%m-%d')

# 연도와 월 추출 및 월별 평균 온도 계산
temp_df['연도'] = temp_df['날짜'].dt.year
temp_df['월'] = temp_df['날짜'].dt.month

# 월별(연도+월)로 그룹화하여 평균 계산
temp_grouped = temp_df.groupby(['연도', '월']).agg({
    '평균기온': 'mean',
    '최저기온': 'mean',
    '최고기온': 'mean'
}).reset_index()

# 두 데이터프레임을 연도와 월을 기준으로 병합
merged_df = pd.merge(water_df, temp_grouped, on=['연도', '월'], how='left')
```

이후 병합을 진행하여 연도, 월을 기준으로 수도 사용량 데이터와 기온 데이터를 병합했습니다.

## 데이터 저장 및 확인

```
merged_df.to_csv('data/merged_data.csv', index=False)
print(f"{merged_df.head(3)}")
```

	연도	월	상하수도사용량	평균기온	최저기온	최고기온
0	2015	1	899	-0.935484	-4.806452	3.583871
1	2015	2	820	1.042857	-2.885714	5.960714
2	2015	3	878	6.290323	0.980645	12.445161

데이터를 data/ 폴더에 저장 후 head() 함수를 사용하여 데이터를 확인했습니다.

## 상관관계 확인

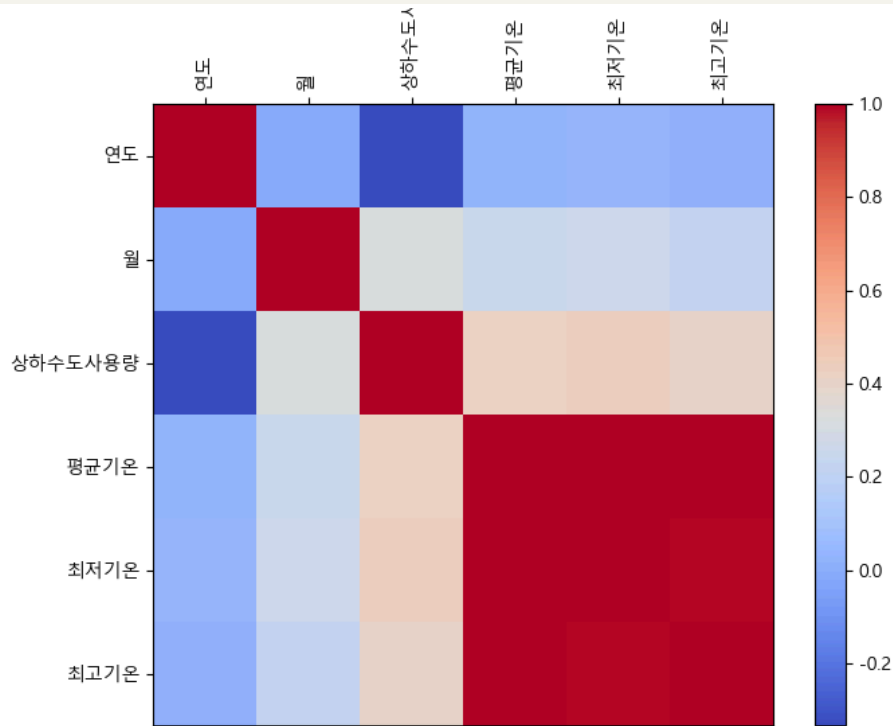
```
import pandas as pd
import matplotlib.pyplot as plt

#한글 설정
plt.rcParams['font.family'] = 'Malgun Gothic'
plt.rcParams['axes.unicode_minus'] = False

merged_df = pd.read_csv('data/merged_data.csv')

corr_matrix = merged_df.corr()

fig, ax = plt.subplots(figsize=(8, 6))
cax = ax.matshow(corr_matrix, cmap='coolwarm')
plt.xticks(range(len(corr_matrix.columns)), corr_matrix.columns, rotation=90)
plt.yticks(range(len(corr_matrix.columns)), corr_matrix.columns)
fig.colorbar(cax)
plt.title('Correlation Matrix', pad=20)
plt.show()
```



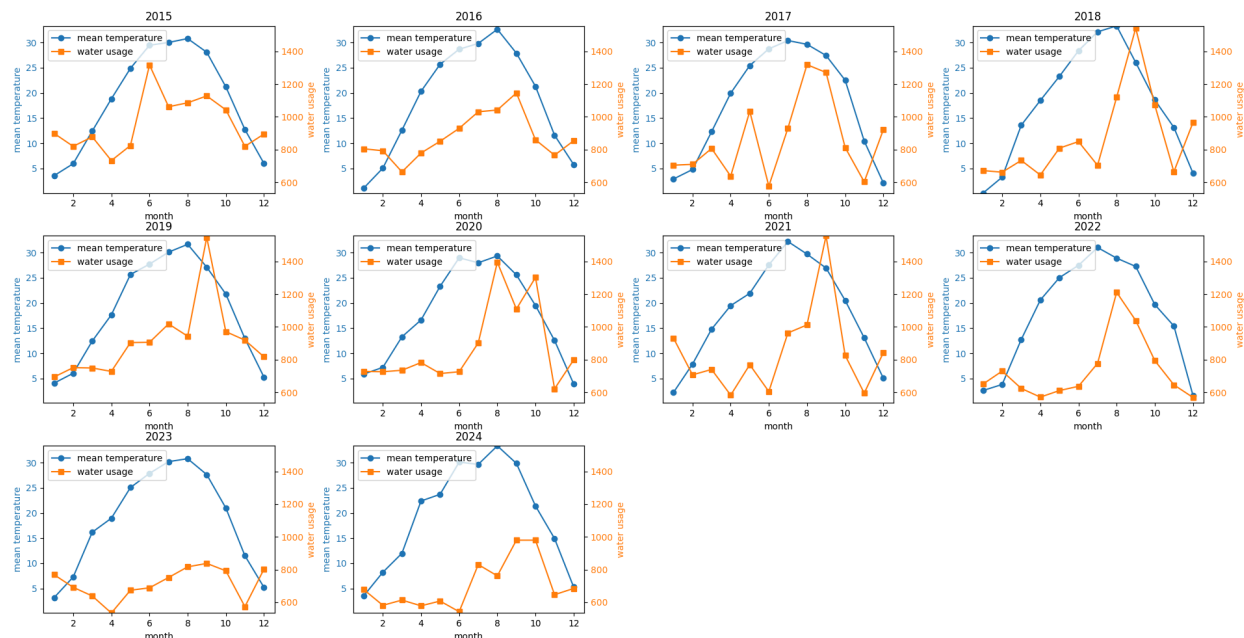
우선 파이썬을 활용해 상관관계를 분석했습니다.

전체적인 상관관계를 분석결과, 기온과 상하수도 사용량 간의 상관도는 낮은 것으로 나타났습니다.

다만, 기온에 따른 사용량의 변동이 일상적인 수도 사용량으로 인해 왜곡될 수 있다고 판단하였습니다.

## 그래프 그리기

따라서 연도별, 월별 기온에 따른 상하수도 사용량 변화 그래프를 matplotlib를 통해 그려 이를 확인하고자 했습니다.

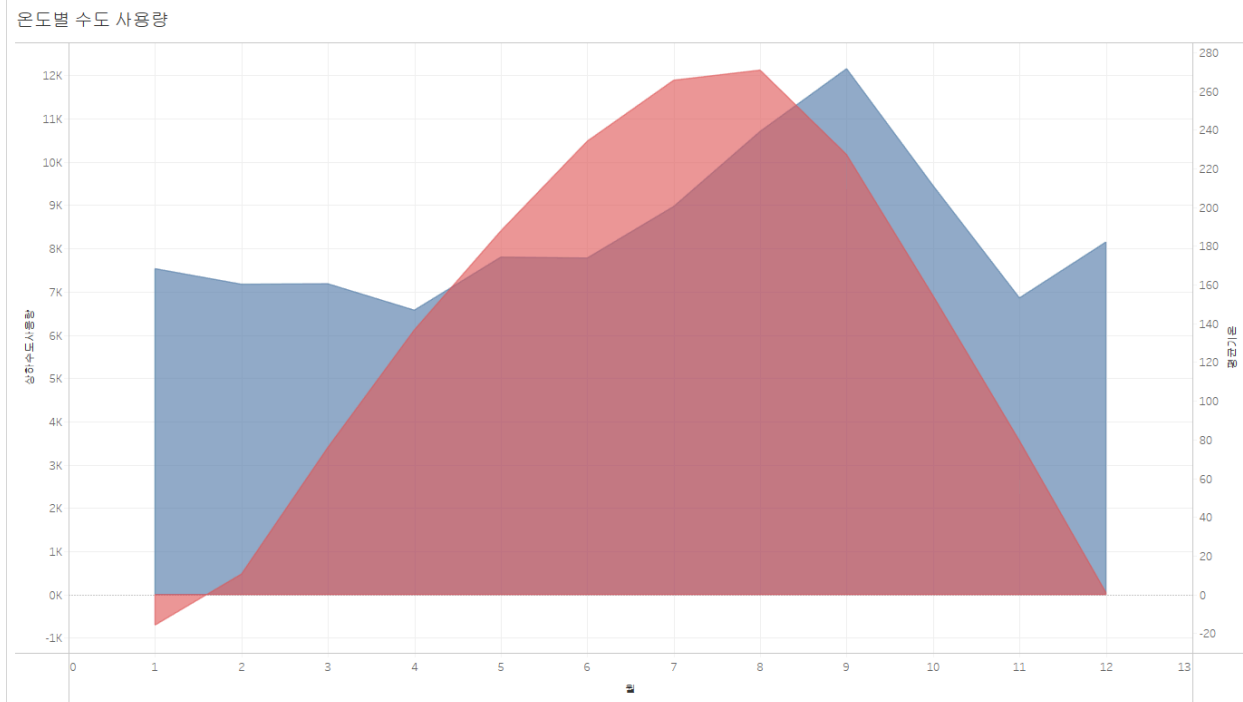


위와 같이 연도 별로 수도의 사용량 차이는 존재하지만

몇 가지 패턴을 확인할 수 있었습니다.

1. 기온이 올라가는 8, 9월에는 수도 사용량이 증가한다.
2. 기온이 떨어지는 겨울철 또한 미미하지만 대체로 수도 사용량이 증가하는 추세를 보인다.

Tableau를 통하여 전체 기온에 따른 수도 사용량을 확인해 본 결과 위와 동일한 결론을 도출했습니다.



## 결론

수도 사용량은 일정 수준 이하로 감소하지 않으며, 기온 상승에 따라 증감하는 경향을 보였습니다.

특히, 8월과 9월처럼 기온이 급격히 상승하는 시기에 수도 사용량이 크게 증가하고, 겨울철(12월과 1월)에는 소폭 증가하는 것으로 나타났습니다.

이를 통해 기온 변화가 수도 사용량에 영향을 미친다는 결론을 도출할 수 있습니다.

## 문제점

현재 데이터에서는 일부 연도의 수도 사용량이 낮게 나타난 이유에 대한 설명이 부족합니다. 이는 단순히 기온이 높아지면 수도 사용량이 증가하는 것이 아니라, 기온 상승에 따른 서서히 증가하는 경향이 존재하기 때문에 추가적인 조사가 필요합니다.