

합천다목적댐 운영정보 분석 및 유입량 예측

Data Analysis Presentation

23-1학기 데이터사이언스캡스톤프로젝트 밤발조

2019311621 송재현

2019312564 노최유하

2019310847 송예진

2019313321 신다은

목차

1. Data Observation
2. Imputation
3. Feature Transformation
4. Exploratory Data Analysis
5. Outlier Handling

Data Observation

1. 합천다목적댐 운영정보
2. 합천군 기상정보
3. 태양고도정보

합천다목적댐 운영정보



- **설명:** K-Water에서 관리하는 합천다목적댐의 운영정보 raw data
- **수집기간:** 2000.01.01. - 2023.04.03.
- **데이터 구성:** 원본 총 23개의 Feature
 - '발전량(실적)', '발전량(계획)', '발전량(계획대비)', '전일유입량', '저수위(전년)', '저수량(전년)', '저수위(현재)', '저수량(현재)', '현재저수율', '전일방류량(본댐)', '시간' 활용
- **주 분석:** K-Water와의 논의를 통해
 - 기존 유입량 예측값인 '유입량'
 - 댐 운영과 밀접한 연관이 있는 '방류량', '유입량', '발전량'

Data Description

합천군 기상정보



기상청

Korea Meteorological
Administration

- **설명:** 기상청에서 제공하는 합천군의 종관기상관측자료
- **수집기간:** 2000.01.01. - 2023.04.03.
- **데이터 구성:** 원본 총 36개의 Feature
 - '일시', '기온(°C)', '강수량(mm)', '습도(%)' 활용
- **주 분석:** K-Water와의 논의를 통해
 - 유입량 예측에 제일 필수적인 요소인 '강수량'
 - 부가적인 날씨요소로 활용될 수 있는 '기온'과 '습도'



- **설명:** 한국천문연구원에서 제공하는 대구광역시 태양고도정보
- **수집기간:** 2016.01.01. - 2022.12.31.
- **데이터 구성:** 원본 총 10개의 Feature
 - '일자', '남중고도' 활용
- **주 분석:** K-Water와의 논의를 통해
 - 계절적인 요소와 유입량의 연관성 확인
 - '남중고도' 활용

1. 합천다목적댐 운영정보
2. 합천군 기상정보
3. 태양고도정보

Imputation

합천다목적댐 운영정보

```
[20] df2 = df.resample('H').asfreq()
df2[['전일유입량', '저수량(전년)', '저수량(예년)', '전년누계강우량', '예년누계강우량', '강우량전일', '전일방류량(본댐)', '전일방류량(조정지)']] = df2[['전일유입량', '저수량(전년)', '저수량(예년)', '전년누계강우량', '예년누계강우량', '강우량전일', '전일방류량(본댐)', '전일방류량(조정지)']]
df2[['저수량(현재)', '금년누계강우량', '현재저수율', '강우량금일']] = df2[['저수량(현재)', '금년누계강우량', '현재저수율', '강우량금일']].interpolate()
df2 = df2.round(1)
df2
```

	전일유입량	저수량(전년)	저수량(현재)	저수량(예년)	현재저수율	전일방류량(본댐)	전일방류량(조정지)	전년누계강우량	금년누계강우량	예년누계강우량	강우량전일	강우량금일
시간												
1997-01-29 23:00:00	1.7	246.6	250.7	277.6	31.7	13.0	13.0	23.0	9.8	39.6	0.0	0.0
1997-01-30 00:00:00	1.7	246.6	250.7	277.6	31.7	13.0	13.0	23.0	9.8	39.6	0.0	0.0
1997-01-30 01:00:00	1.7	246.6	250.7	277.6	31.7	13.0	13.0	23.0	9.8	39.6	0.0	0.0
1997-01-30 02:00:00	1.7	246.6	250.7	277.6	31.7	13.0	13.0	23.0	9.8	39.6	0.0	0.0
1997-01-30 03:00:00	1.7	246.6	250.7	277.6	31.7	13.0	13.0	23.0	9.8	39.6	0.0	0.0
...
2023-04-05 19:00:00	3.3	316.1	212.4	323.7	26.9	5.0	5.3	71.2	106.5	127.0	6.9	24.2
2023-04-05 20:00:00	3.3	316.1	212.5	323.7	26.9	5.0	5.3	71.2	106.5	127.0	6.9	24.2
2023-04-05 21:00:00	3.3	316.1	212.5	323.7	26.9	5.0	5.3	71.2	106.5	127.0	6.9	24.2
2023-04-05 22:00:00	3.3	316.1	212.6	323.7	26.9	5.0	5.3	71.2	106.5	127.0	6.9	24.2
2023-04-05 23:00:00	3.3	316.1	212.6	323.7	26.9	5.0	5.3	71.2	106.5	127.0	6.9	24.2

229489 rows × 12 columns

Imputation

합천군 기상정보

연속형(수치형) 데이터: 중간값

```
data_col_list = list(data_n.columns)
data_col_list.remove("전문량(10분위)")
data_col_list.remove("중하층운량(10분위)")
data_col_list
```

```
['지점',
 '지점명',
 '일시',
 '기온(° C)',
 '강수량(mm)',

 '최저운고(100m )',
 '시정(10m)',
 '지면온도(° C)']
```

```
for c in data_col_list:
    data_n[c].interpolate(inplace=True)
```

범주형 데이터: 앞 행의 값

```
data_n['전문량(10분위)'] = data_n['전문량(10분위)'].fillna(method='ffill')
data_n['중하층운량(10분위)'] = data_n['중하층운량(10분위)'].fillna(method="ffill")
data_n["풍향(16방위)"] = data_n["풍향(16방위)"].fillna(method="ffill")
```

Feature Transformation

1. Datetime Indexing
2. Daily Data

Feature Transformation

Datetime Indexing

	전일유입량	저수량(전년)	저수량(현재)	저수량(예년)	현재저수율	전일방류량(본댐)	전일방류량(조정지)	전년누계강우량	금년누계강우량	예년누계강우량	강우량전일	강우량금일
시간												
1997-01-29 23:00:00	1.7	246.6	250.7	277.6	31.7	13.0	13.0	23.0	9.8	39.6	0.0	0.0
1997-01-30 00:00:00	1.7	246.6	250.7	277.6	31.7	13.0	13.0	23.0	9.8	39.6	0.0	0.0
1997-01-30 01:00:00	1.7	246.6	250.7	277.6	31.7	13.0	13.0	23.0	9.8	39.6	0.0	0.0
1997-01-30 02:00:00	1.7	246.6	250.7	277.6	31.7	13.0	13.0	23.0	9.8	39.6	0.0	0.0
1997-01-30 03:00:00	1.7	246.6	250.7	277.6	31.7	13.0	13.0	23.0	9.8	39.6	0.0	0.0
...
2023-04-05 19:00:00	3.3	316.1	212.4	323.7	26.9	5.0	5.3	71.2	106.5	127.0	6.9	24.2
2023-04-05 20:00:00	3.3	316.1	212.5	323.7	26.9	5.0	5.3	71.2	106.5	127.0	6.9	24.2
2023-04-05 21:00:00	3.3	316.1	212.5	323.7	26.9	5.0	5.3	71.2	106.5	127.0	6.9	24.2
2023-04-05 22:00:00	3.3	316.1	212.6	323.7	26.9	5.0	5.3	71.2	106.5	127.0	6.9	24.2
2023-04-05 23:00:00	3.3	316.1	212.6	323.7	26.9	5.0	5.3	71.2	106.5	127.0	6.9	24.2

229489 rows × 12 columns

데이터 분석 및 모델에의 적용의 편의성을 위해 시간열을 인덱스로 사용

Feature Transformation

Daily Data

시간	전일유입량	저수량(전년)	저수량(현재)	저수량(예년)	현재저수율	전일방류량(본댐)	전일방류량(조정지)	전년누계강우량	금년누계강우량	예년누계강우량	강우량전일	강우량금일
2000-01-01	7.4	508.5	513.4	340.0	65.0	28.9	26.1	0.0	1491.6	0.6	0.0	0.0
2000-01-02	4.4	507.1	512.1	339.1	64.8	23.7	27.4	0.0	0.1	0.6	0.0	0.1
2000-01-03	4.7	505.8	509.8	338.0	64.6	26.1	23.8	0.0	0.2	1.0	0.2	0.0
2000-01-04	3.8	504.3	508.4	337.0	64.4	27.3	26.8	0.0	0.2	1.8	0.0	0.0
2000-01-05	6.1	503.2	506.5	336.1	64.1	25.2	26.1	0.0	1.4	3.5	0.0	1.2
...
2023-03-30	2.8	318.5	212.9	326.9	26.9	8.5	9.2	71.2	75.4	116.0	0.0	0.0
2023-03-31	2.6	318.3	212.7	326.5	26.9	8.7	7.9	71.2	75.4	117.3	0.0	0.0
2023-04-01	2.7	318.0	212.6	326.0	26.9	1.3	5.2	71.2	75.4	120.1	0.0	0.0
2023-04-02	2.3	317.6	212.4	325.5	26.9	5.0	5.3	71.2	75.4	121.7	0.0	0.0
2023-04-03	2.6	317.2	212.2	324.9	26.9	5.0	5.1	71.2	75.4	123.3	0.0	0.0

8494 rows × 12 columns

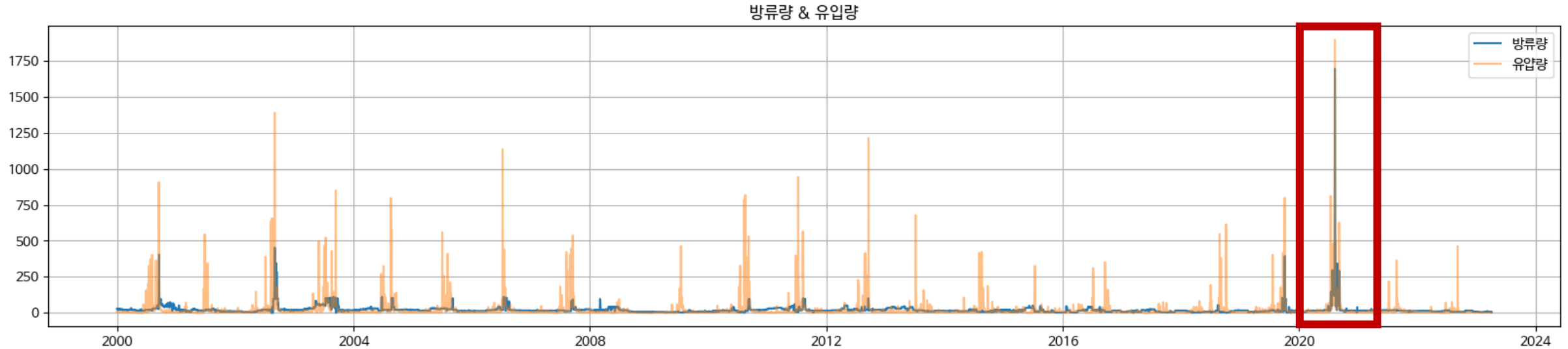
시간별 데이터를 사용해 모델을 학습시키기는 어려울 것이라는 한국수자원공사의 조언에 따라
1일 간격의 평균값 데이터로 변환

Exploratory Data Analysis

1. 합천다목적댐 운영정보
2. 합천군 기상정보
3. 태양고도정보
4. 특이사항

Exploratory Data Analysis

합천다목적댐 운영정보 - 유입량과 방류량



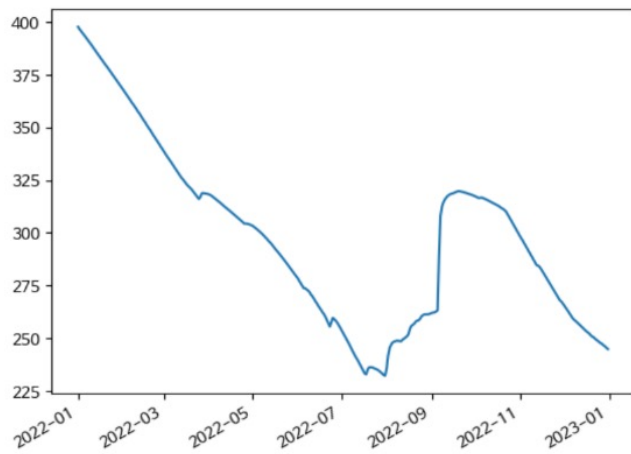
일반적으로 홍수기(6/21~9/20)에 댐 유입량이 증가함에 따라 방류량 증가

2020년 유입량의 증가에 따라 방류량도 같은 수준으로 증가
→ 홍수위 초과가 예상되던 시점으로 급히 방류

합천다목적댐 운영정보 - 저수량

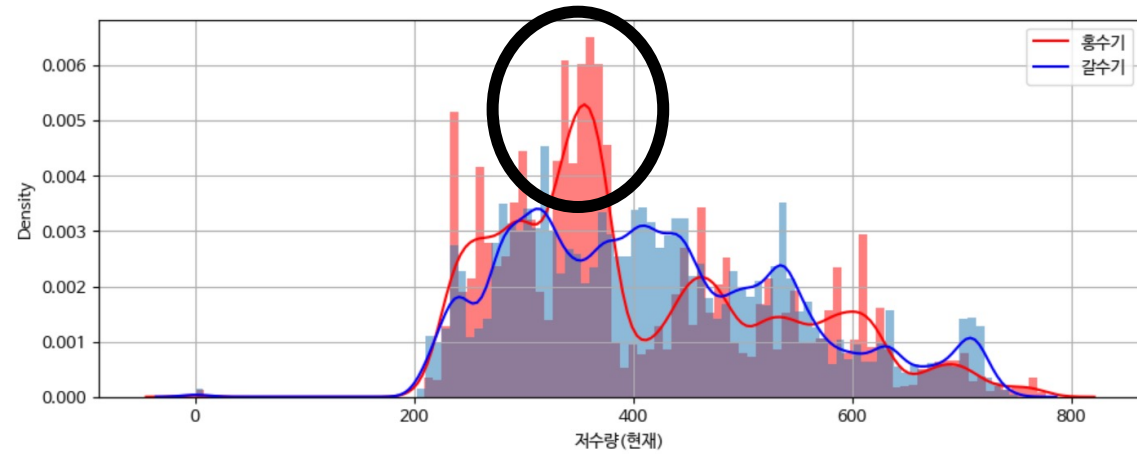
가정) 홍수기엔 여름철 호우에 대비하여 저수량이 상대적으로 적었다가,
갈수기에 홍수기 영향 및 용수 확보를 위해 저수량이 증가할 것이다.

1. 월별 저수량



홍수기에 저수량이 적었다가 갈수기에 점차 증가

2. 홍수기와 갈수기의 저수량 비교

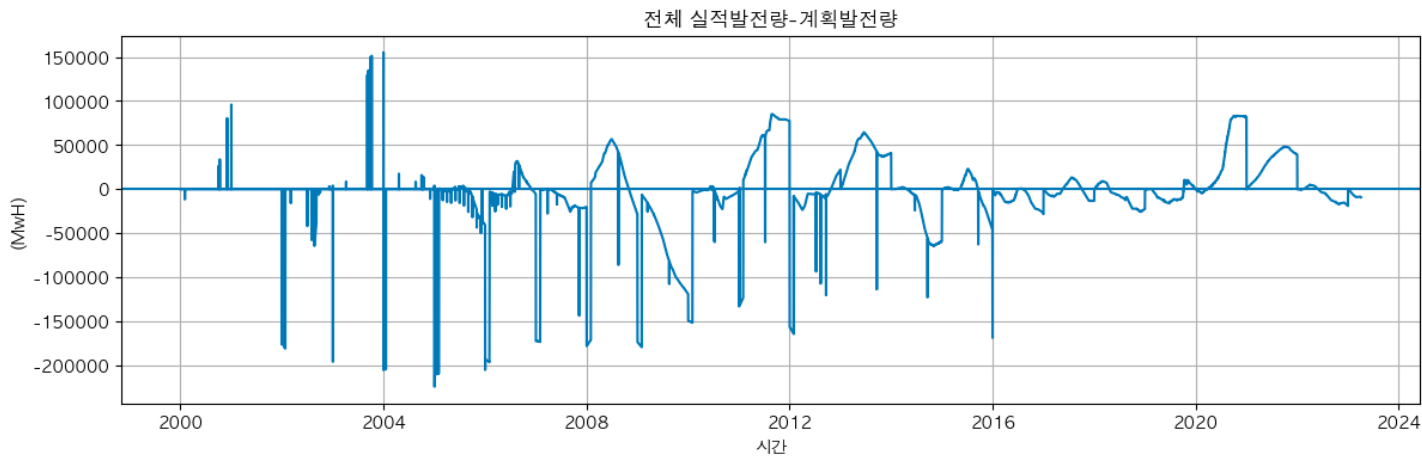
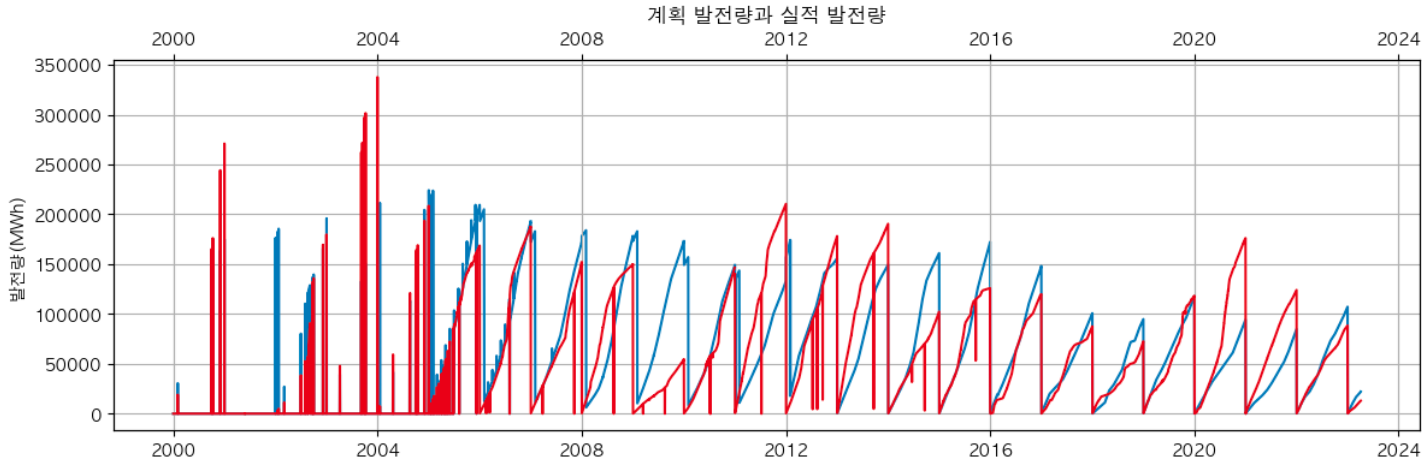


홍수기 저수량이 가장 높은 분포를 보이는 구간은
갈수기 저수량의 주요 분포 구간보다 낮음.

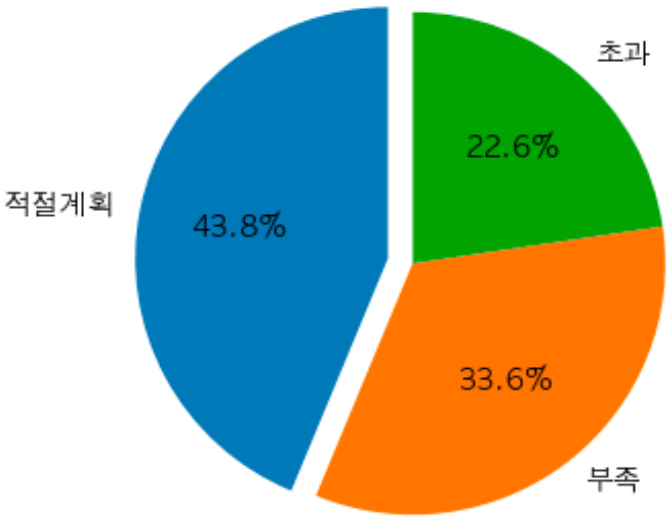
=> 가정에 부합하는 분석 결과

합천다목적댐 운영정보 - 발전량

댐 방류 시 발전기를 돌리며 생산할 계획발전량과 실적발전량 비교

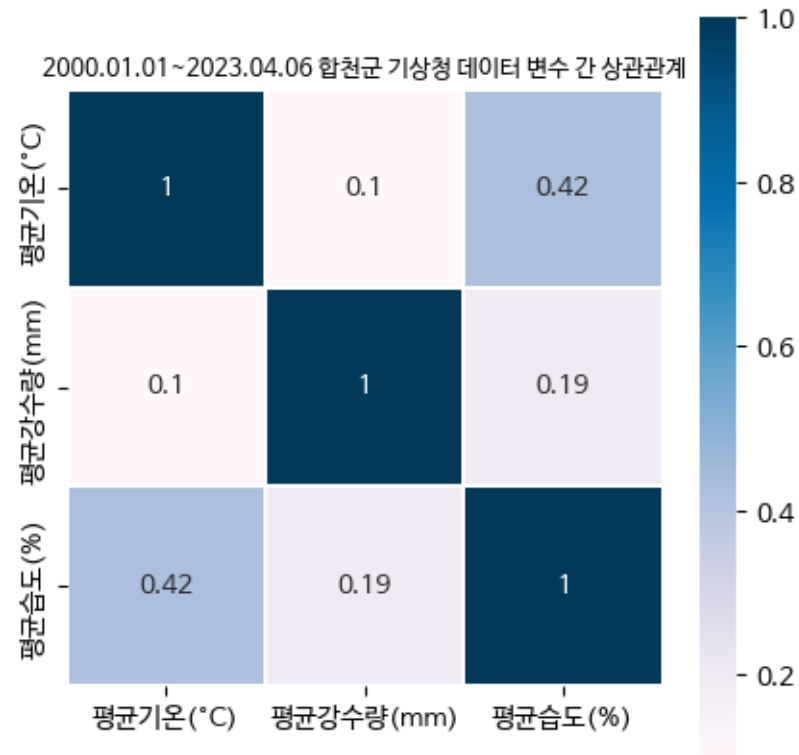


계획발전량과 실적발전량 비교



- 비슷한 추세를 예측하나, 정확한 예측이 적음.
- 절반 이상 계획이 적절히 이루어지지 않음.
- 유입량이 잘 예측된다면, 방류량을 잘 조절해 발전량이 부족/초과되는 상황이 적어질 것

합천군 기상정보

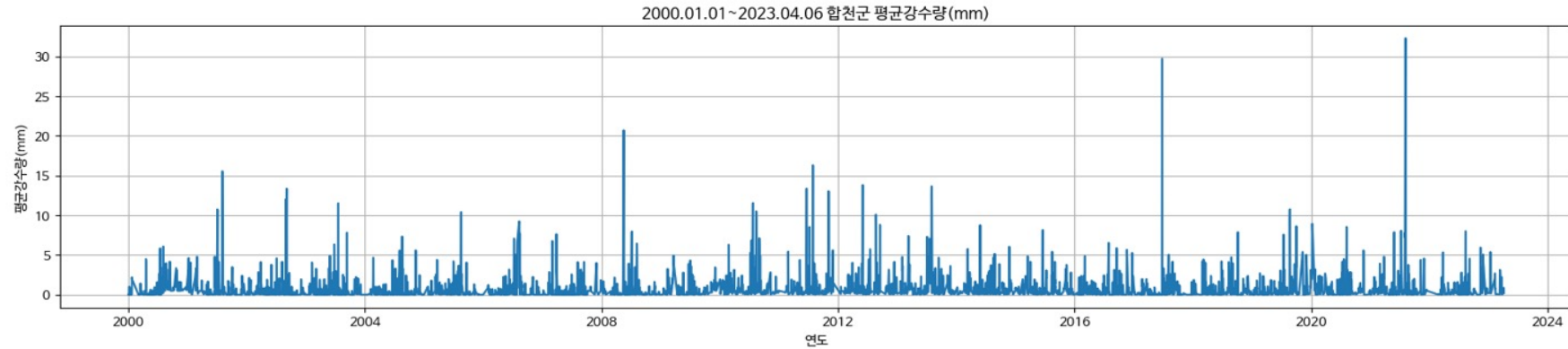


- 일일 평균기온, 평균 강수량, 평균습도 세 변수의 상관관계 분석 및 다중공선성 문제 발생 가능성 판단

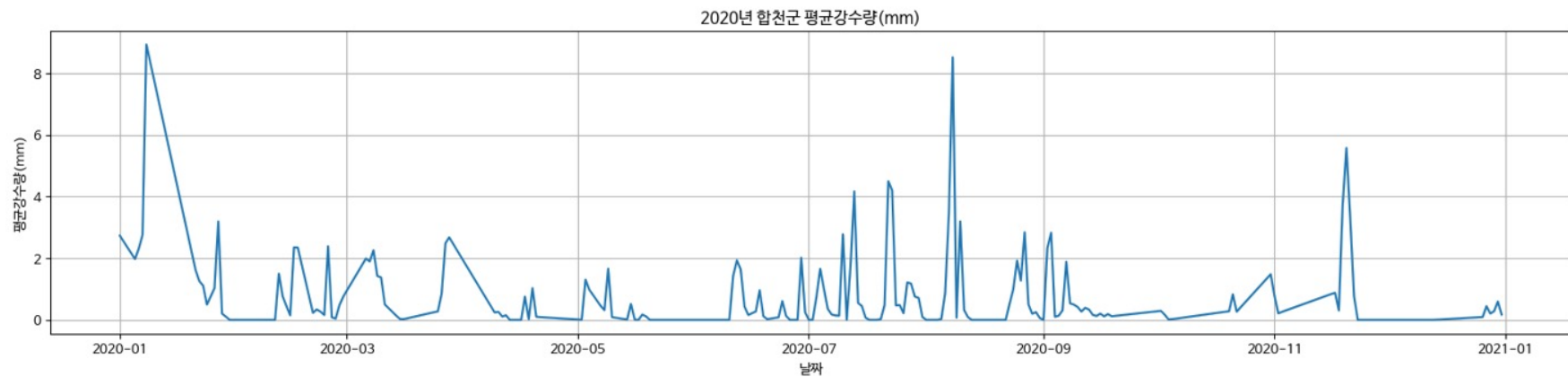
Exploratory Data Analysis

합천군 기상정보 - 강수량

2000년 1월 1일부터 2023년 4월 6일까지의 합천군 일일 평균 강수량 그래프



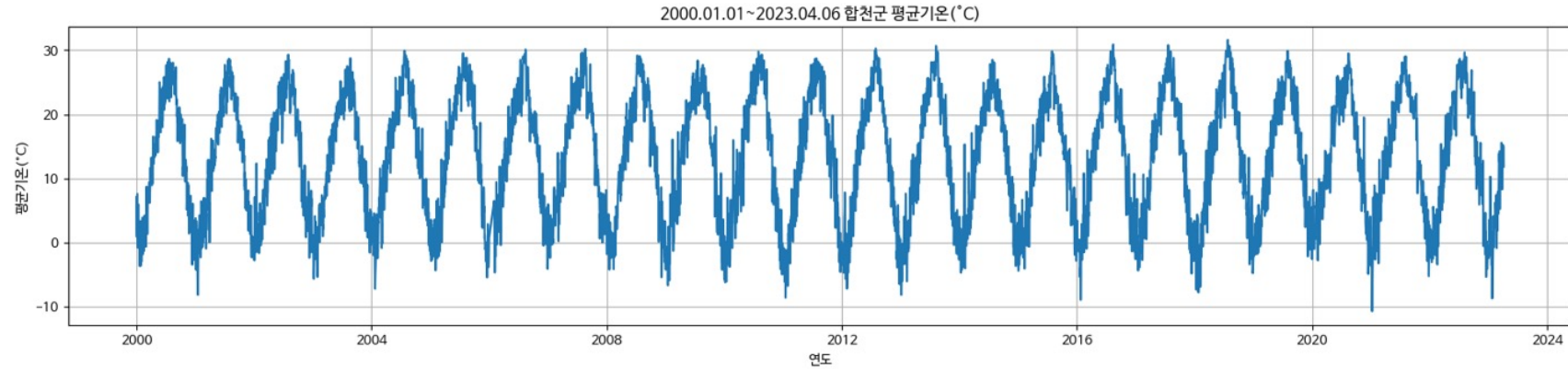
2020년의 이상치 원인을 파악하기 위한 2020년 합천군 일일 평균 강수량 그래프



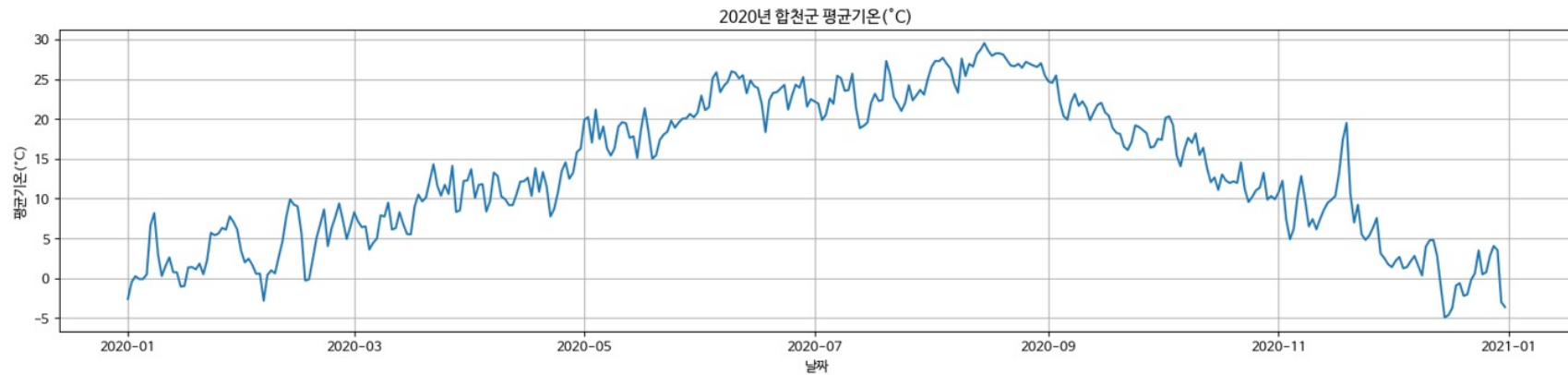
Exploratory Data Analysis

합천군 기상정보 - 기온

2000년 1월 1일부터 2023년 4월 6일까지의 합천군 일일 평균기온 그래프



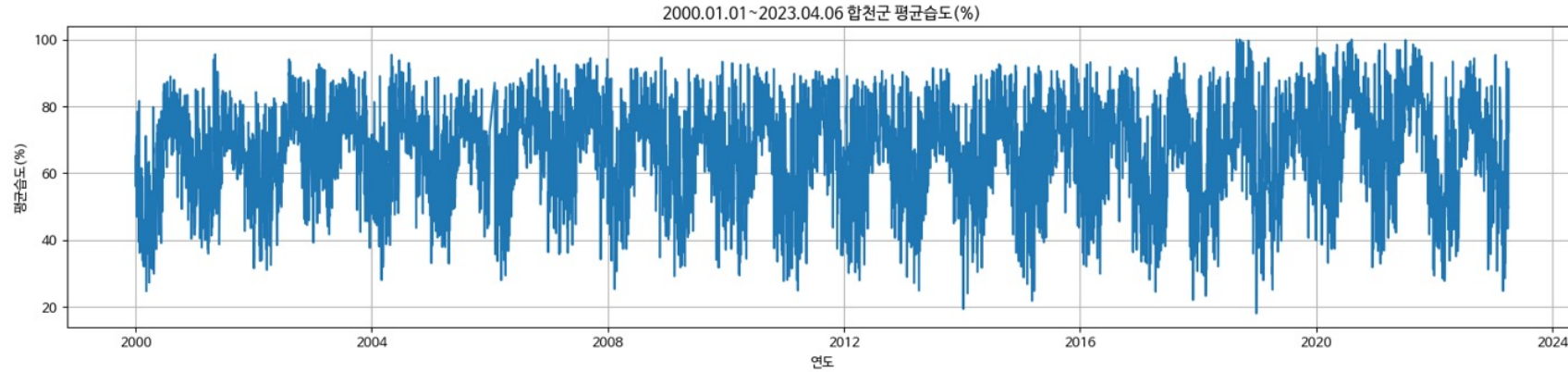
2020년의 이상치 원인을 파악하기 위한 2020년 합천군 일일 평균기온 그래프



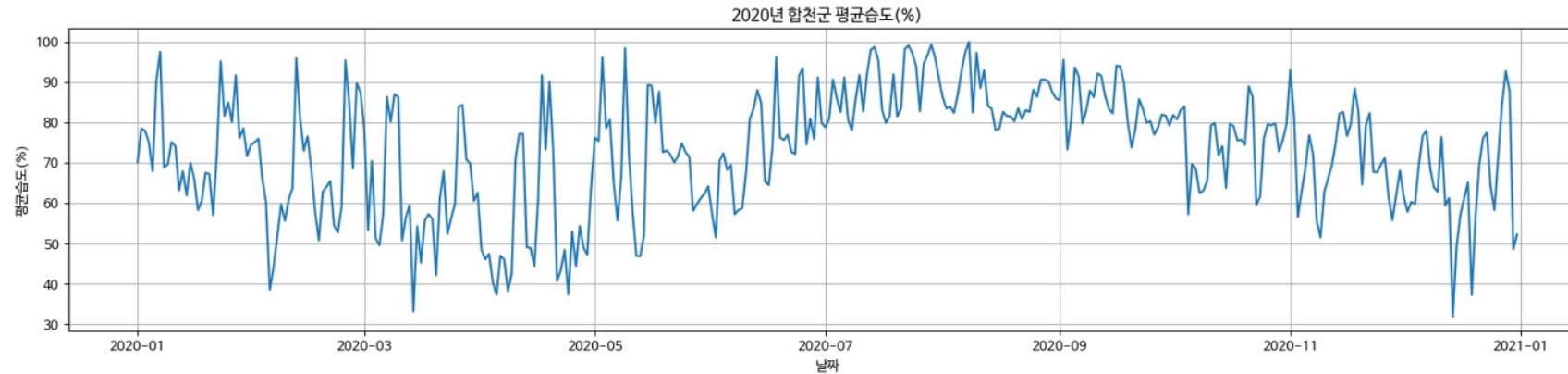
Exploratory Data Analysis

합천군 기상정보 - 습도

2000년 1월 1일부터 2023년 4월 6일까지의 합천군 일일 평균습도 그래프

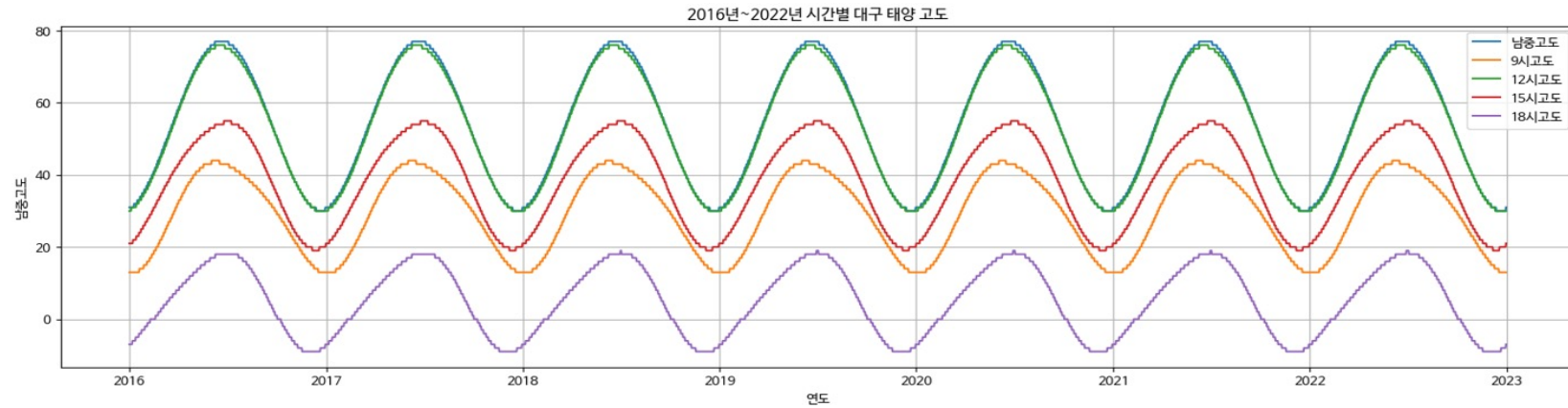


2020년의 이상치 원인을 파악하기 위한 2020년 합천군 일일 평균습도 그래프

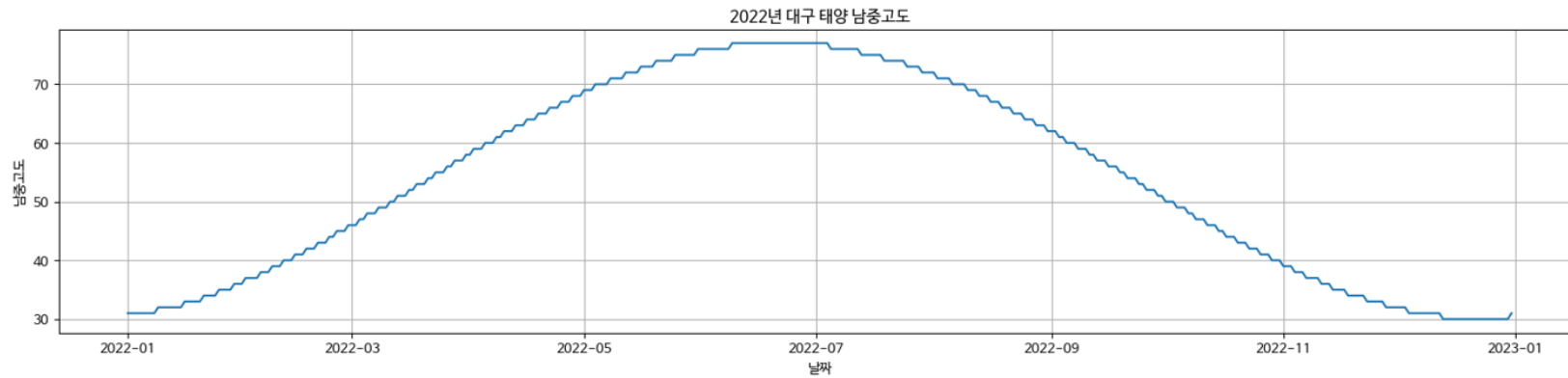


태양고도정보 - 남중고도

2016년부터 2022년까지의 대구광역시 태양 고도 그래프



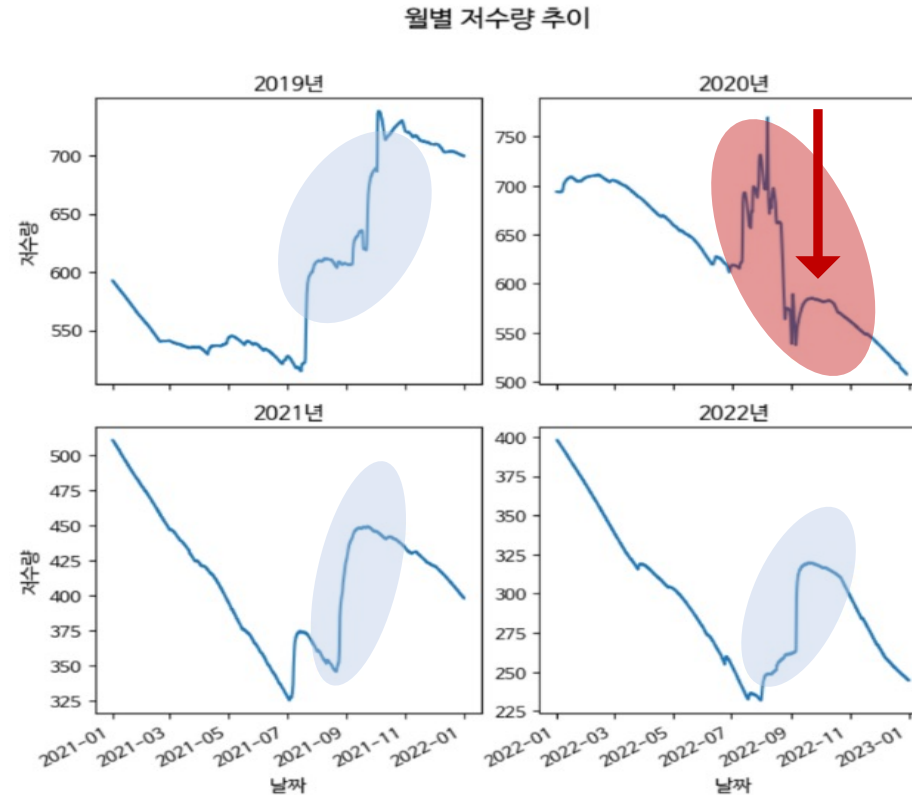
1년 간 남중고도의 변화를 파악하기 위한 2022년 대구광역시 태양 남중고도 그래프



Exploratory Data Analysis

특이사항

홍수기를 대비하여
6~9월 저수량 감소,
이후 **갈수기**에 다시
증가



갈수기를 대비하여
증가해야 할 저수량이
급격히 감소

수자원 공사 답변: 2020년 태풍 '바비'의 예상 세력과 실제 세력 간 차이로 인한 결과

태풍이 강력한 위력을 지닐 것으로 예측되어 방류량을 늘렸으나,
태풍의 중심이 육지에서 멀리 떨어졌고 빨리 약화되어 저수량을 다시 확보하지 못함.

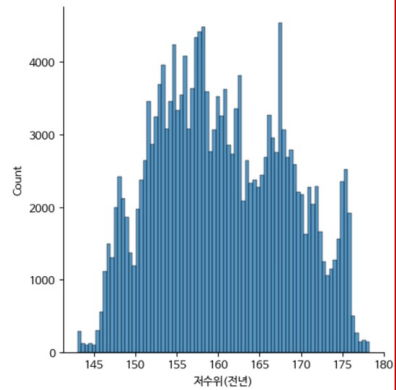
=> 2020년도 이상 추이

Outlier Handling

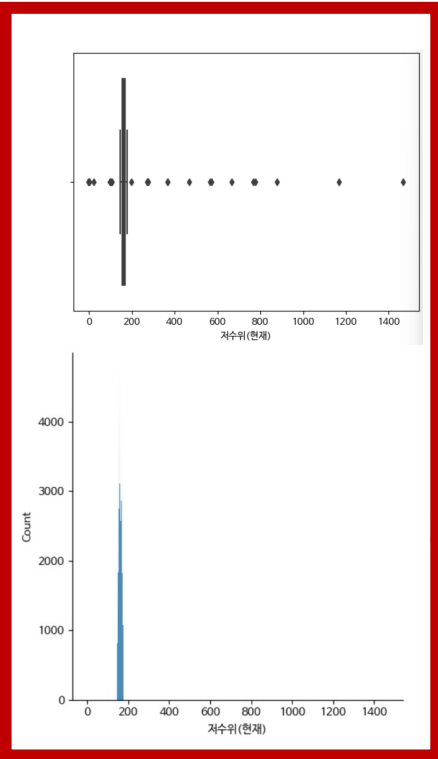
1. 저수량 이상치
2. 과거 이상치

Outlier Handling

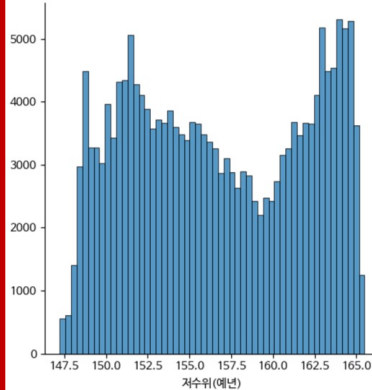
저수위 이상치



전년 저수위



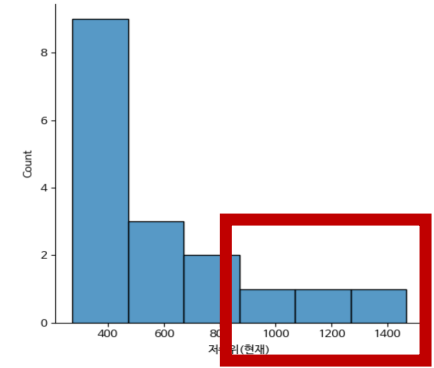
현재 저수위



예년 저수위

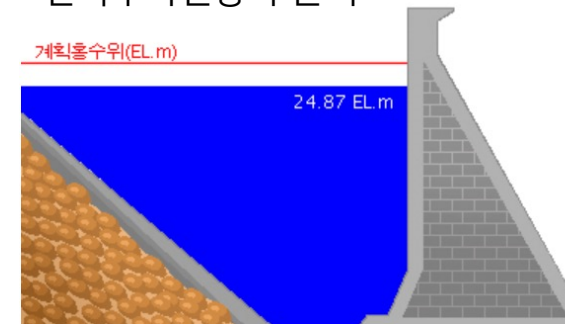
```
df['저수위(현재)'].describe()
```

```
count    184969.000000
mean      160.487997
std        9.719948
min         0.000000
25%       154.580000
50%       160.020000
75%       166.440000
max      1468.080000
Name: 저수위(현재), dtype: float64
```



이상치 판단 기준 $Q3 + 1.5 * IQR$
-> 약 20개의 이상치 확인

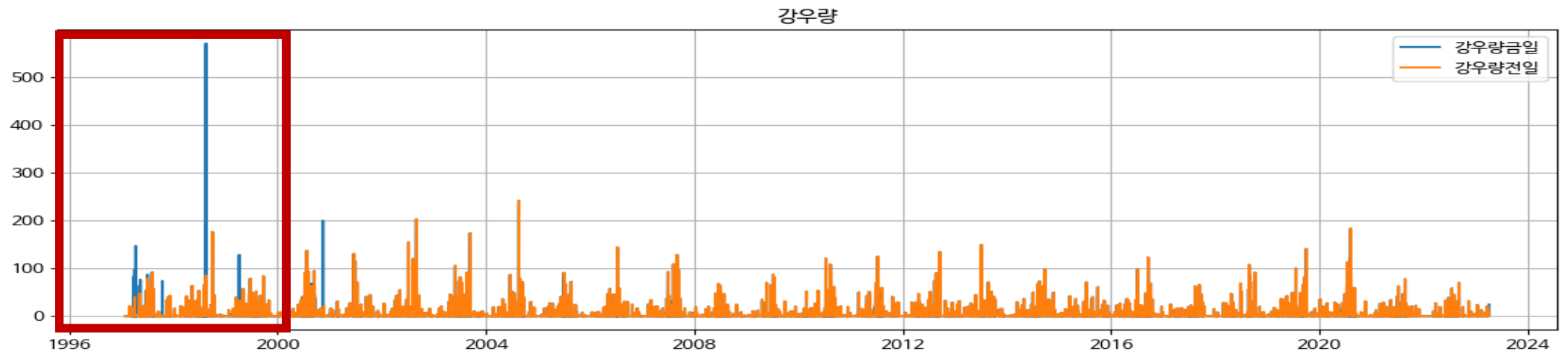
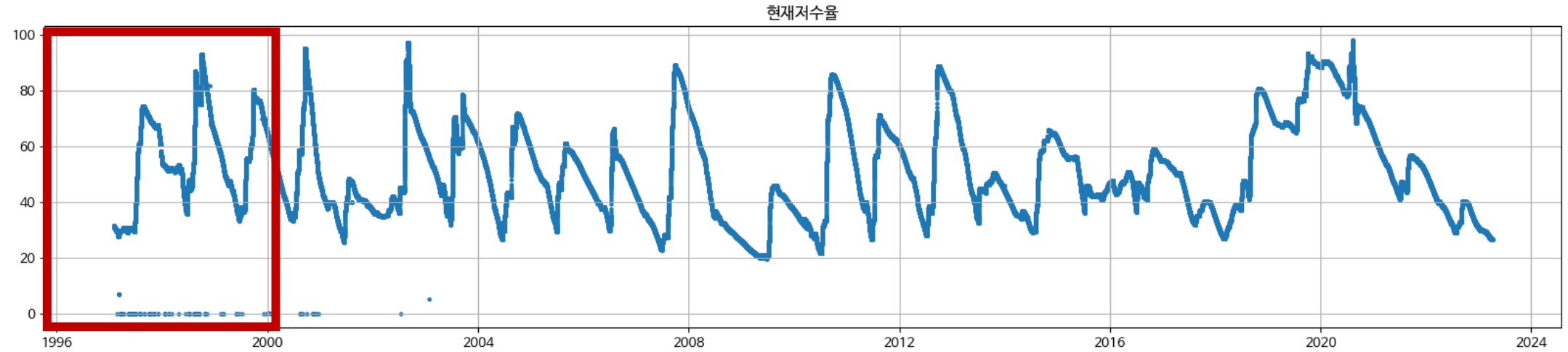
* 한국수자원공사 문의



계획홍수위를 넘는 경우 측정 오류로 간주
이상치 처리: 이상치를 보이는 기간 앞뒤의 값으로 보간

Outlier Handling

과거 이상치



2000년 이전 데이터 삭제 결정

감사합니다

밤밭조

송재현, 노최유하, 송예진, 신다은