

수용가 범주화를 통한 효율적 물 절약 방안제시

임예림 김동규 양명철 정수린 최재훈
중앙대학교 '물 너의 행방을 찾아서'

INDEX


1. 배경 및 목적
2. 데이터 전처리
3. 분석 주요 내용
4. 분석 결과
5. 결론 및 시사점

 적은 강수량

높은 일조량

주요하천에서 멀리 떨어진 지리적 특성



 가뭄으로 인한 수자원 부족

💧 외부재원확보, 물 절약 캠페인 시행, 대형 관정 개발

➡️ 절수 권고가 잘 지켜지지 않아 제한 급수를 실시

“수용가별 물 사용량에 영향을 미치는 요인을 찾으면
맞춤형 물 절약 방안을 제시할 수 있을 것으로 기대됨”

데이터 설명

water_usage_dataset.csv

	Meter_ID	Data	Time	Usage(cumulative)	Usage(hourly)
1	16-005738	20161001	0	30.949	0
2	16-005738	20161001	1	30.949	0
3	16-005738	20161001	2	30.949	0
4	16-005738	20161001	3	30.949	0
5	16-005738	20161001	4	30.949	0



한국 수자원 공사

- › 세대 별 시간 당 수돗물 사용량
- › 관측 기간 : 2016.10.01 ~ 2017.12.31

customer_info.csv

No.	수용가.번호	meter_id	주소	상수도업종	구경	급수상태
1	4.42101E+11	16-005738	팔봉면 호리 418	가정용	15	개전
2	4.42101E+11	16-005271	팔봉면 호리 719-1	가정용	15	개전
3	4.42101E+11	16-005737	팔봉면 호리 496	가정용	15	개전
4	4.42101E+11	16-006148	팔봉면 호리 500	가정용	15	개전
5	4.42101E+11	16-005585	팔봉면 호리 512	가정용	15	개전



한국 수자원 공사

- › 수용가 정보 (주소, 업종)
- › 관측 기간 : 2016.10.01 ~ 2017.12.31

변수 설명

water_usage_dataset.csv

- › Meter_ID : 수도미터 ID 각수용가와 1:1매치
- › DATE : 데이터 취득 일
- › TIME : 데이터 취득 시간
- › Usage(cumulative) : 누적 수돗물 사용량 (m³/시간)
- › Usage(hourly) : 시간별 수돗물 사용량 (m³/시간)

customer_info.csv

- › 수용가 번호 : 각 수용가의 고유 번호
- › Meter_ID : 수도미터 ID 각수용가와 1:1매치
- › 주소 : 수용가의 주소
- › 상수도 업종 : 수용가의 업종 (가정, 일반, 학교)
- › 수용가 번호 : 각 수용가의 고유 번호
- › 급수상태 : 개전/폐전 또는 중지

데이터 처리과정

- 결측치(NA)
- 실시간 사용량의 합
≠ 누적사용량 값
- 측정시점 ≥ 24

1. 오류값 제거


- 상위 0.01%
→ 이상치 제거
- 2016.12.01~2016.12.05
→ 고려할 만한 이슈X
- 스마트 미터기 시스템의 오류
가능성 존재

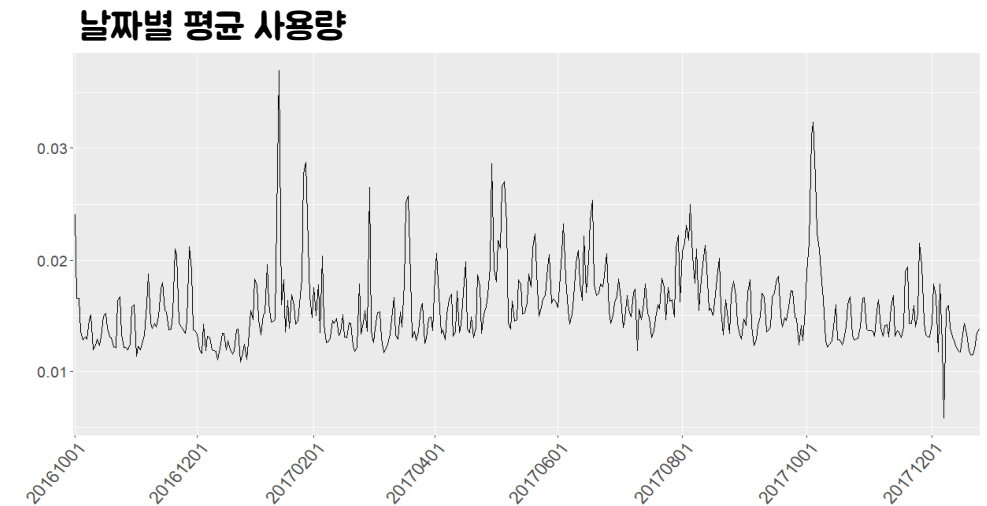
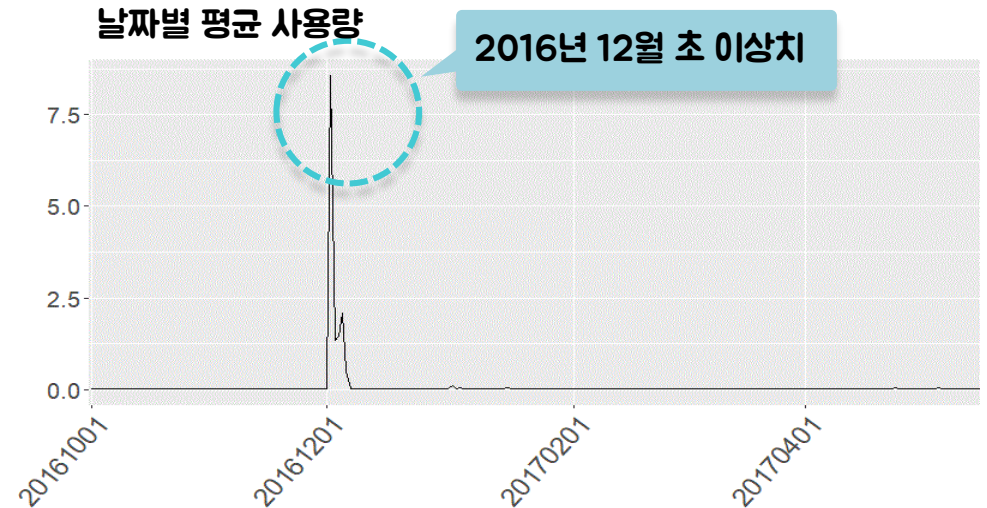
2. 이상치 제거

- 〈기준〉
1. 물 사용 규모
 2. 수용가별 특성과 요인별 패턴
- 네 가지 범주로 구분
 - 초대수용가
 - 일반대수용가
 - 관광수용가
 - 가정수용가

3. 새로운 범주 생성

데이터의 처리과정 - 2. 이상치 제거

- 하루 사용량 합계 중 특정 기간 사용량 급증
일별 패턴 확인 불가
- 
- 이상치 제거 후, 일별 패턴 확인 가능
(2016년 12월 1일 ~ 2016년 12월 5일 제거)



데이터의 처리과정 - 3. 새로운 범주 생성

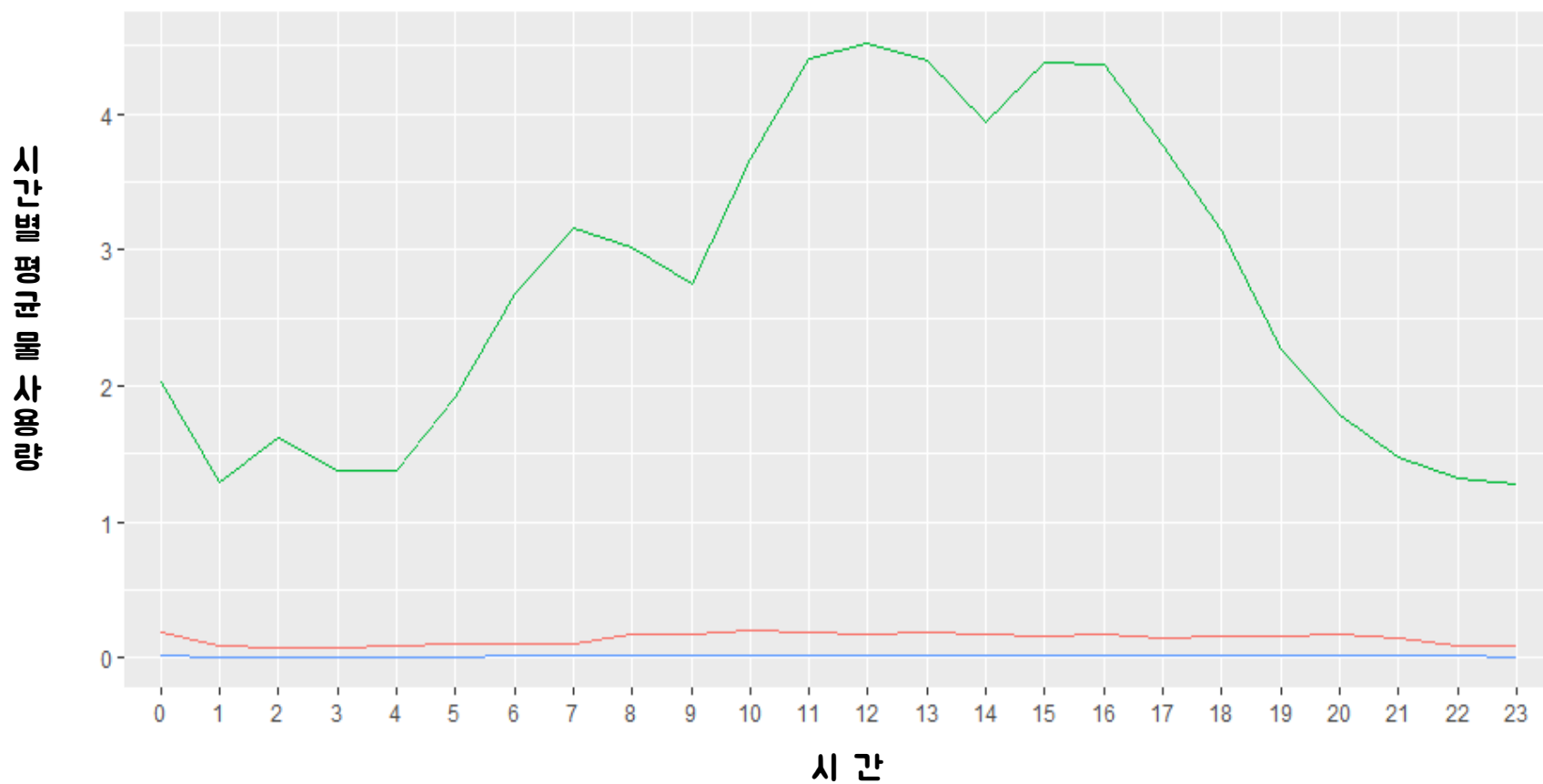
› 범주화가 왜 필요한가?

➡ 데이터 내 수용가들이 이질적인 집단으로 나뉜다는 사실을 발견

**'뚜렷한 물 사용 패턴 차이를 더 잘 분석하기 위해서는
범주를 잘 구분하는 것이 중요하다고 판단된다!'**

데이터의 처리과정 - 3. 새로운 범주 생성

▶ 범주화가 왜 필요한가? ➡ 데이터 내 수용가들이 이질적인 집단으로 나뉜다는 사실을 발견



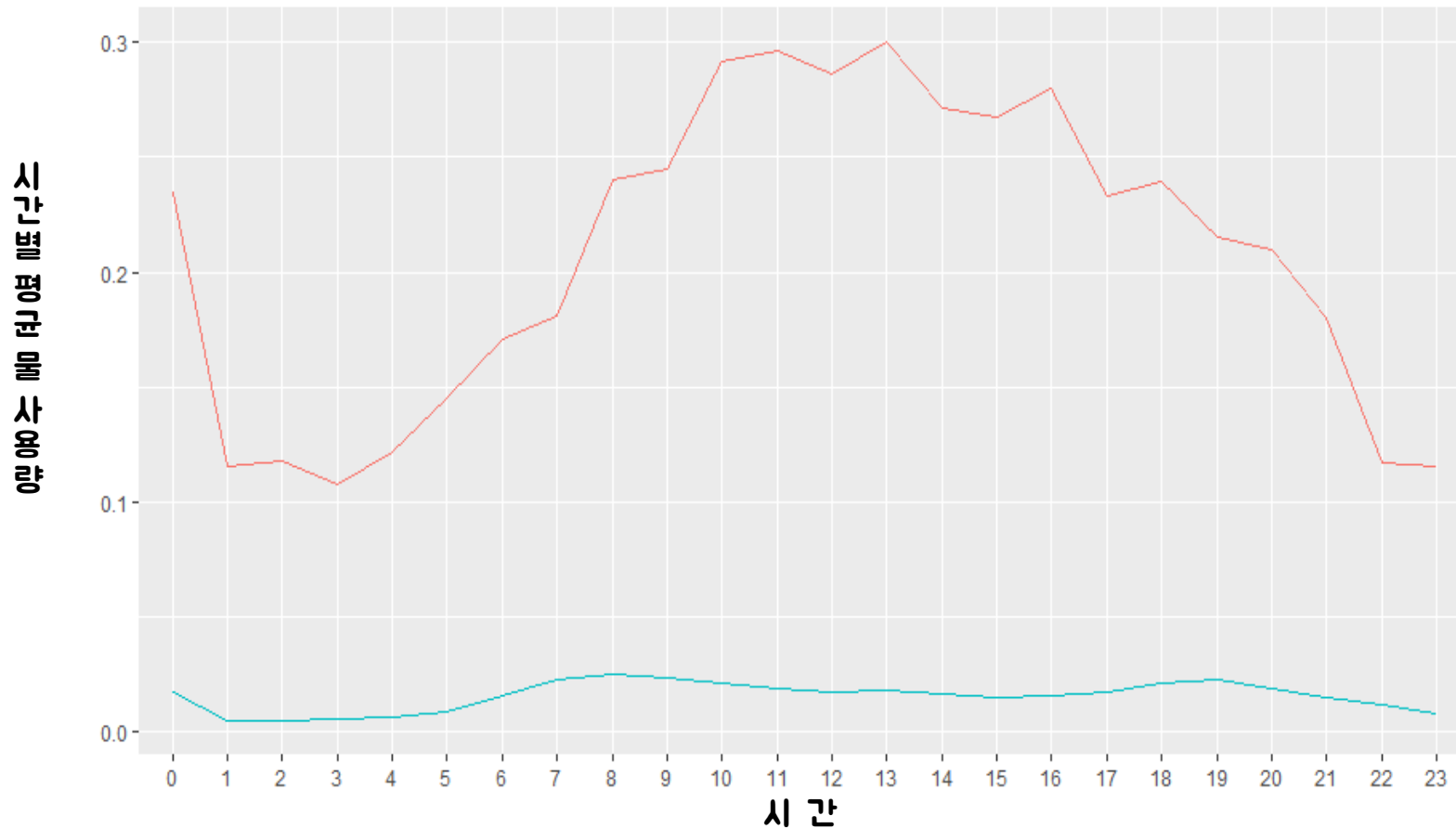
(주)광축 :
하루평균 물사용량 1위
전체 사용량의 약 9%를 차지

수 용 가

상위 40
광 축
그 외

데이터의 처리과정 - 3. 새로운 범주 생성

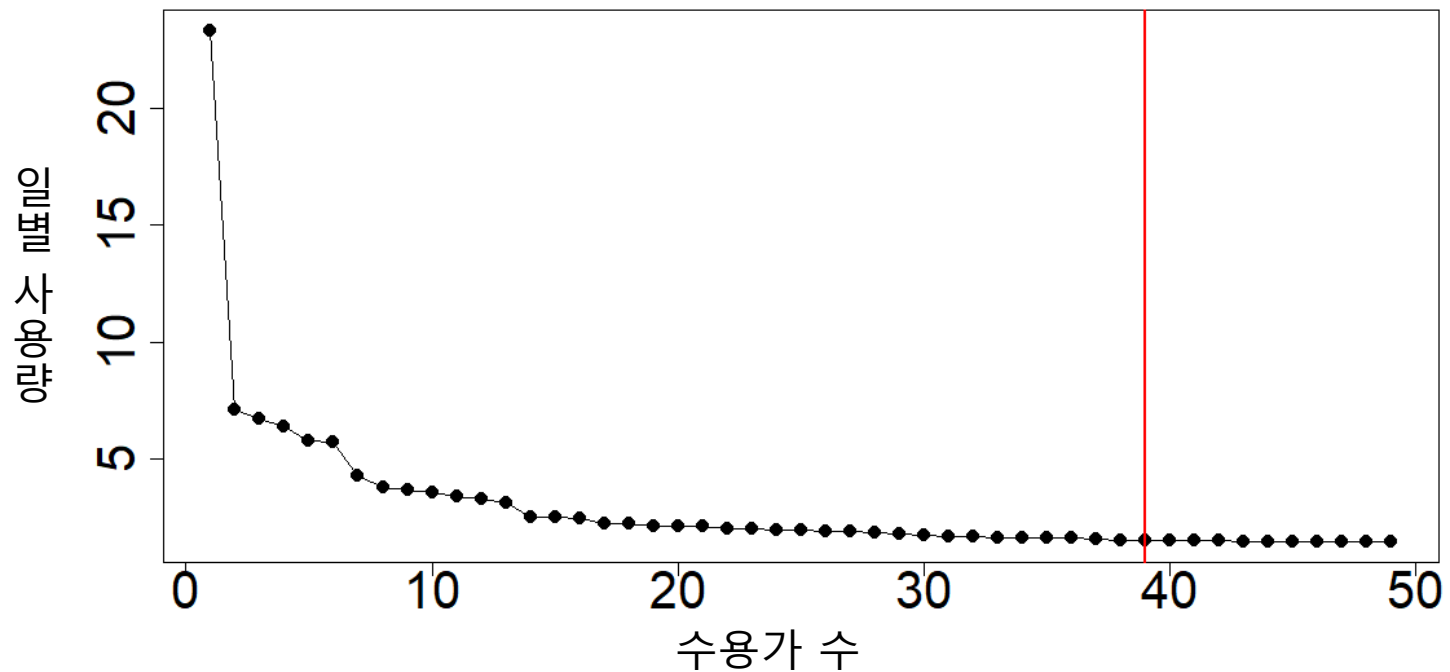
▶ 범주화가 왜 필요한가? ➡ 데이터 내 수용가들이 이질적인 집단으로 나뉜다는 사실을 발견



(주)광축을 제외한 39세대와 그 이외의 물사용량 패턴 차이가 명확
기준 : 하루평균 물사용량 약 1.5톤 이상인가

데이터의 처리과정 - 3. 새로운 범주 생성

▶ 하루 평균 물사용량을 기준으로 상위 40개 추출



상위 40개의 기준 :
하루 평균 물사용량이 약 1.5+ 이상

데이터의 처리과정 - 3. 새로운 범주 생성

- › **초대수용가** : (주)광축 도축장 Meter_ID 16-009003
- › **관광수용가** : 음식점, 숙박업소 등 관광인구에 영향을 받을 수 있는 수용가
- › **일반대수용가** : 대수용가 중 관광수용가 제외 나머지 수용가
- › **가정수용가** : 초대수용가, 일반대수용가, 관광수용가 제외 나머지 수용가

물 사용량
상위 40개 가구



분석기법의 정의

➤ 1. 회귀분석 : 설명변수와 종속변수의 관계를 모델링하는 기법

물 사용량에 영향을 미치는 요인을 탐지할 수 있는가?

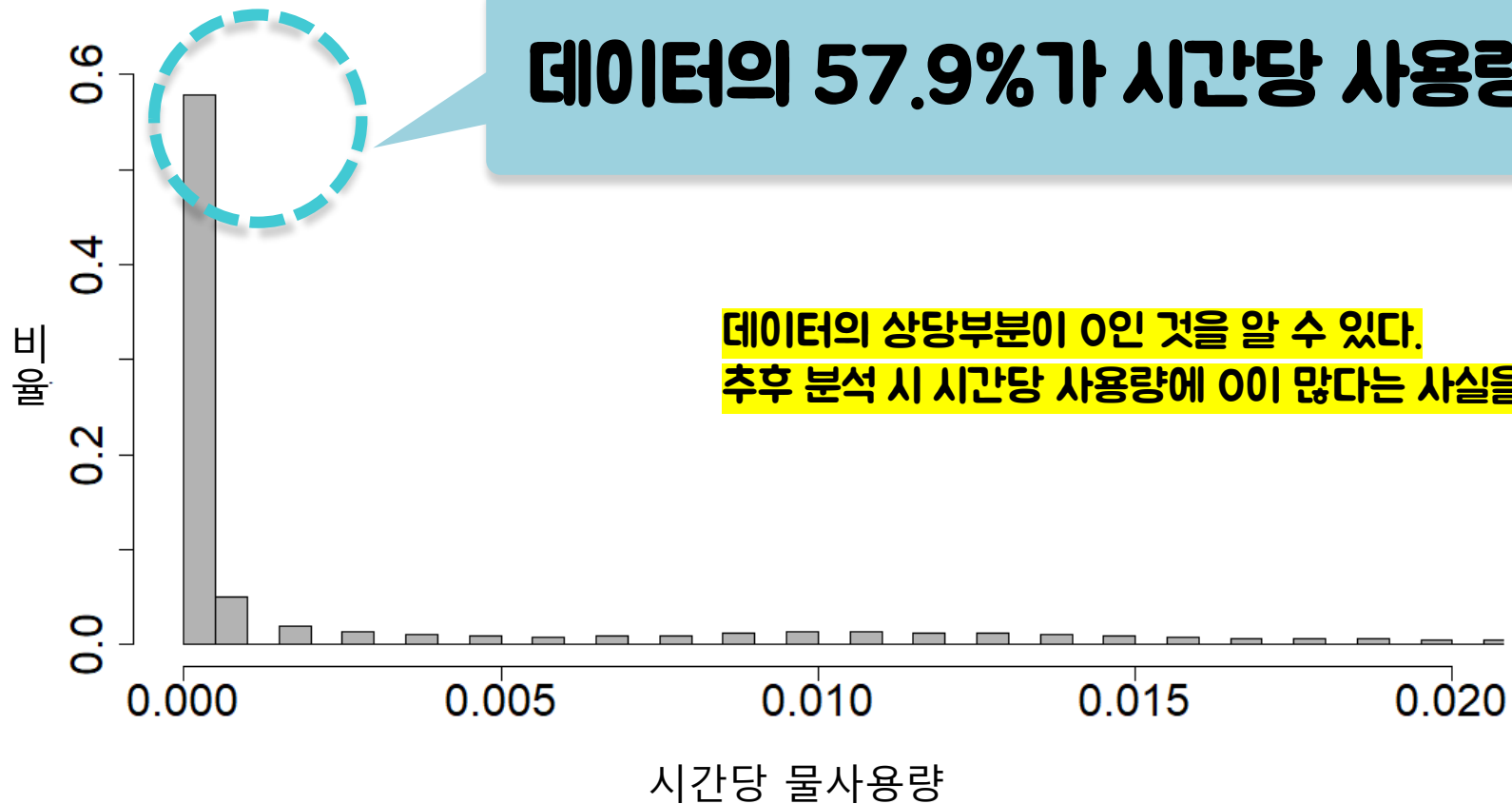
➤ 2. 의사결정나무 및 랜덤 포레스트 : 데이터에 내재된 패턴을 이용, 데이터를 분류/예측하는 방법론

회귀분석에서 살펴본 요인들이 수용가들을 적절하게 분류할 수 있는가?

➤ 3. 시계열 분석 : 과거 자료를 통해 미래를 예측하는 방법론

주어진 데이터로 미래 물 수요량을 예측할 수 있는가?

데이터의 특징



데이터의 상당부분이 0인 것을 알 수 있다.

추후 분석 시 시간당 사용량에 0이 많다는 사실을 고려할 필요가 있다.

회귀분석 > 분위 회귀(Quantile Regression)

분위 회귀(Quantile regression) : 각 분위수(quantile)에서 회귀계수를 추정하는 비모수적인 방법

- › 일반적인 회귀가정으로부터 자유로움 → 비선형이고 극단치가 많을 경우 선호됨
- › 자료는 특성상 사용량이 0인 자료가 절반 이상 → 분위 회귀가 적합

SVM Regression

모형의 과적합(Overfitting)

Ridge 및 LASSO

해석의 복잡성

Linear Regression

회귀가정을 만족시키지 못함

회귀분석 > 분위 회귀(Quantile Regression)

- 계절 더미의 영향 ↑
- 여름에 피서객이 많이 몰리는 관광지의 특성이 반영됨
- 공휴일 회귀계수가 양수

[관광수용가]

- 공휴일, 시간 변수의 영향 ↑
- 일반대수용가에는 사업장이 많이 포함
- 공휴일에 쉬는 특성을 갖는 수용가들이 많음

[일반대수용가]

- 공휴일의 회귀계수가 양수
↓
쉬는 날에 더 많은 물을 사용함
- 빨래와 같은 가사노동을 하기 때문인 것으로 추측됨

[가정수용가]

의사결정나무와 랜덤 포레스트

변수설명

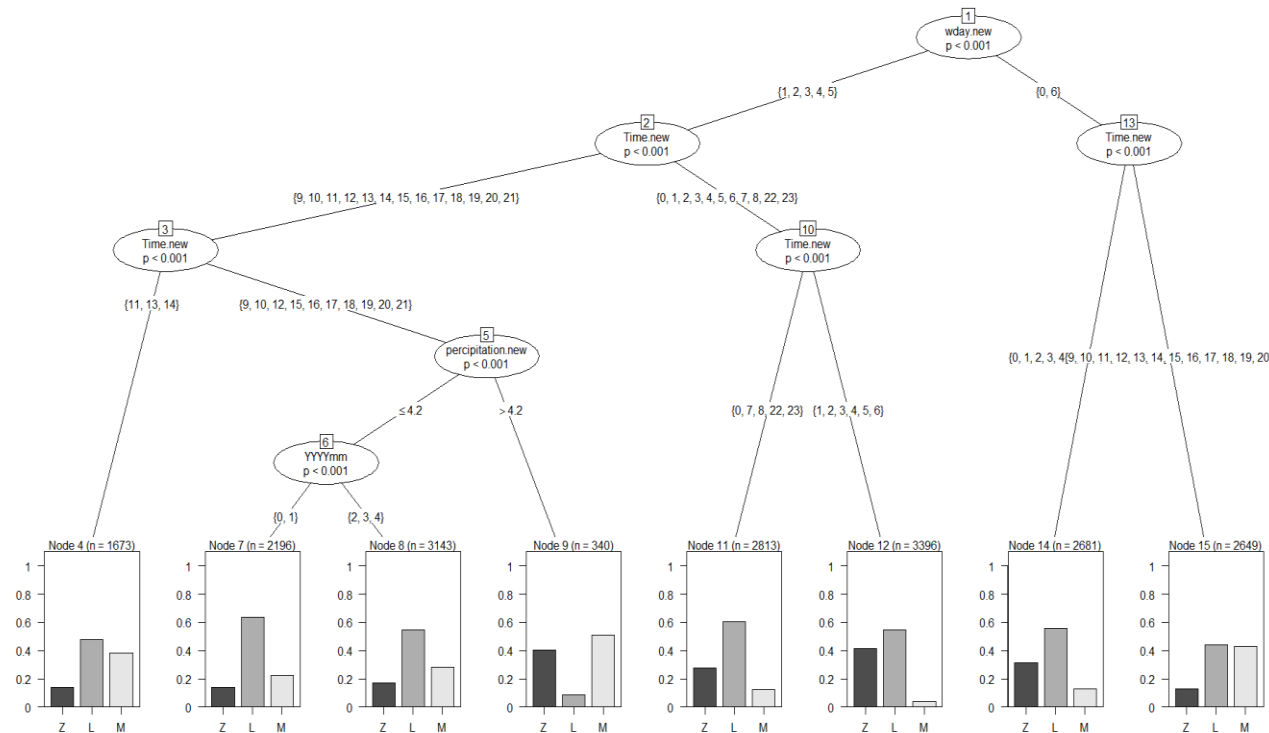
- › 중위수를 기준으로 집단을 나누어 분석진행
시간별 물 사용량이 0인집단(Z), 중위수보다 작은집단(L), 중위수보다 큰 집단(M)
- › 사용한 변수 :
시간 변수(0시~23시), 개월 변수, 일별 평균 기온, 일별 평균 강수량, 공휴일 유무 더미변수, 요일 변수, 상수도 사용 여부, 구경, 상수도 종류
 - 분석 초기 단계에서 시간당 기온, 시간당 강수량을 고려해 모델 구성하였으나, 변수값의 변동폭이 커서 분류와 예측력에 한계를 보인다고 판단, 일별평균기온, 일별평균강수량을 사용

의사결정나무와 랜덤 포레스트

모델설명

- ▶ 비모수적 모형, Robust, 해석이 용이
- ▶ 한 개의 node 에서 선택하는 변수를 랜덤으로 3개씩 고르도록 트리를 구성
- ▶ 분할 기준 : p-value
- ▶ 데이터 분할 기준을 4500개로 설정시 뚜렷한 군집 특성 관찰가능

의사결정나무와 랜덤 포레스트 - 관광수용가



예측률 : 54.49%

랜덤 포레스트 사용시 예측률 : 65.11%

Z : 시간별 물 사용량 0인 집단

L : 중위수보다 작은 집단

M : 중위수보다 큰 집단

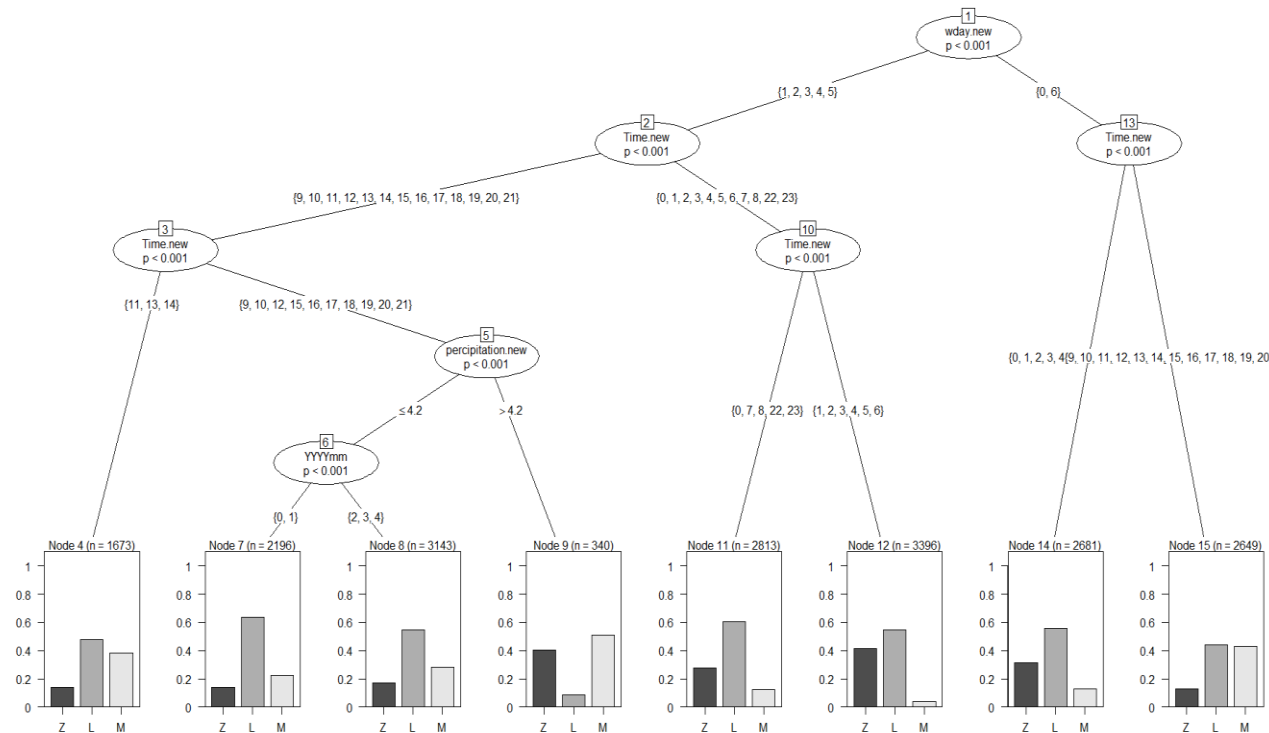
전체 데이터 188888개 중 약 24.21%인 45737개가 0
중위수 : 0.077

주말에는 새벽 제외한 모든 시간에 물 사용량이 높다.

관광수용가는 일반대수용가에 비해 물 사용량의 중위수가 높다.

관광업소 물관리에 대한 대책 마련 필요

의사결정나무와 랜덤 포레스트 - 일반대수용가



Z : 시간별 물 사용량 0인 집단

L : 중위수보다 작은 집단

M : 중위수보다 큰 집단

전체 데이터 230174개 중 약 25%인 58359개가 0
중위수 : 0.066

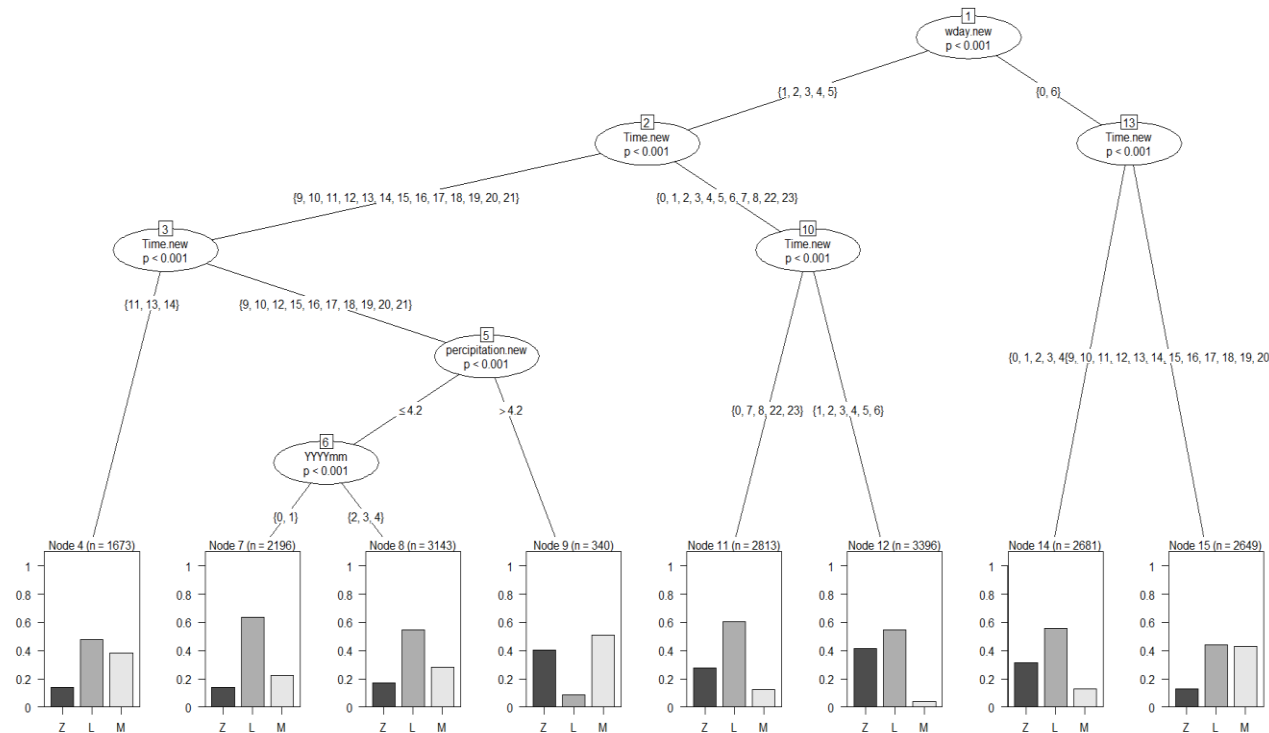
상위 node에 공휴일 변수가 존재하므로
일반대수용가의 물 사용량에는 공휴일이 큰 영향을 준다.

비공휴일, 강수량 0.9 이하, 새벽 제외, 기온 2.4~21.4
도 이하일 경우 물 사용량이 많다.

예측률 : 34.34%

랜덤 포레스트 사용시 예측률 51.01%

의사결정나무와 랜덤 포레스트 - 가정수용가



Z : 시간별 물 사용량 0인 집단

L : 중위수보다 작은 집단

M : 중위수보다 큰 집단

전체 데이터 15476697개 중 약 58%인 9100870개가 0
중위수 : 0.015

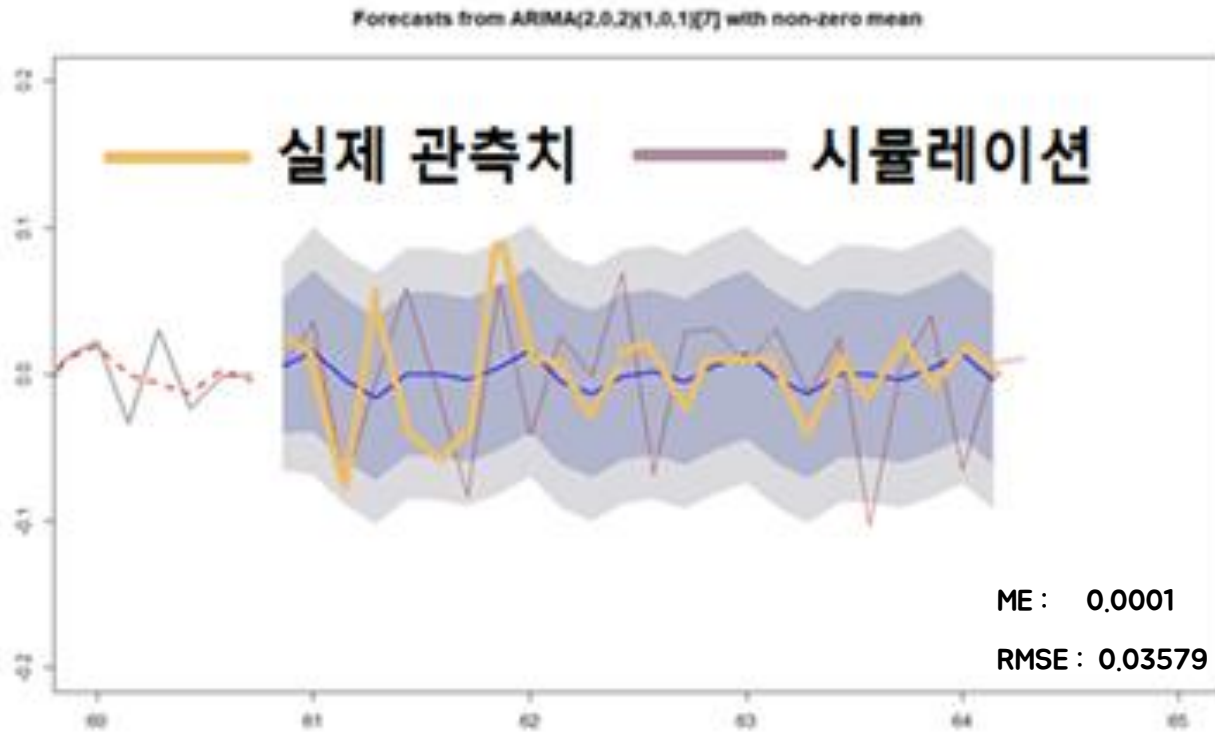
가정수용가 임을 고려하였을 때, 물사용량이
0인 데이터가 다수 포함돼 있기 때문에 분류
에 어려움을 겪은 것으로 보인다.

가정수용가의 경우 군집별 뚜렷한 패턴이 없음

시계열 분석

변동이 비교적 작은 일별 물 사용량 데이터 사용, ARIMA 모형 : 7일마다 반복되는 주기를 고려함

ARIMA 모형 적합 후 실제 관측치와 모형 기반 시뮬레이션을 비교

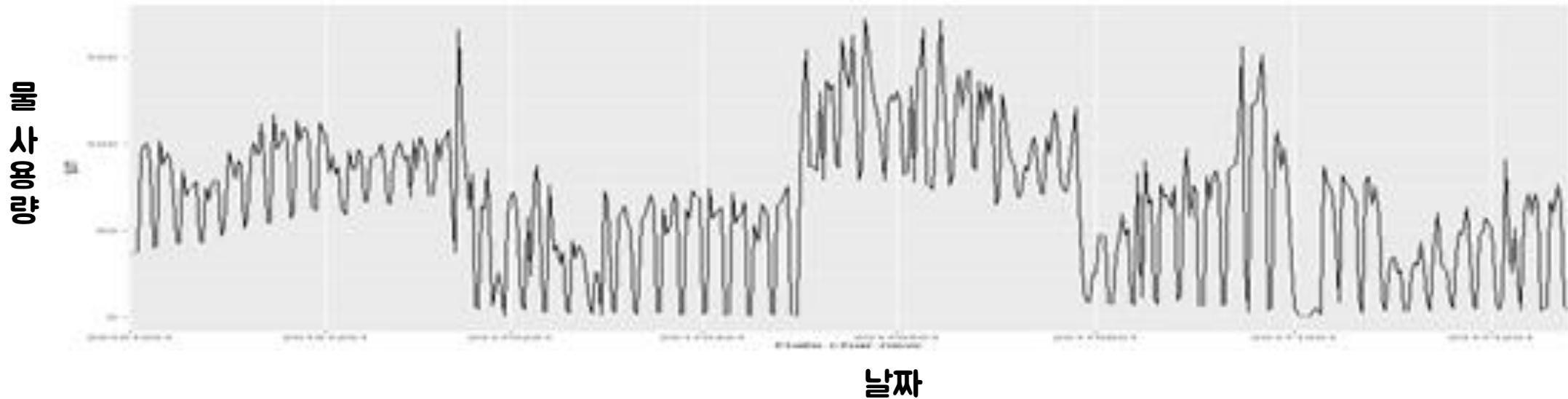


실제 관측치와 시뮬레이션값이 비슷
→ 시뮬레이션이 적절하게 예측하고 있음

과거의 데이터를 이용해 미래를 예측
수요량에 영향을 주는 요인을
탐지하기에는 부적합

초대수용가의 사용량 분석 - (주) 광축

› 15개월동안 (주)광축이 사용한 일별 물 사용량



› 서산시 팔봉면의 물 사용량 중 9%를 차지

› 물 사용량 등락이 규칙적인 편

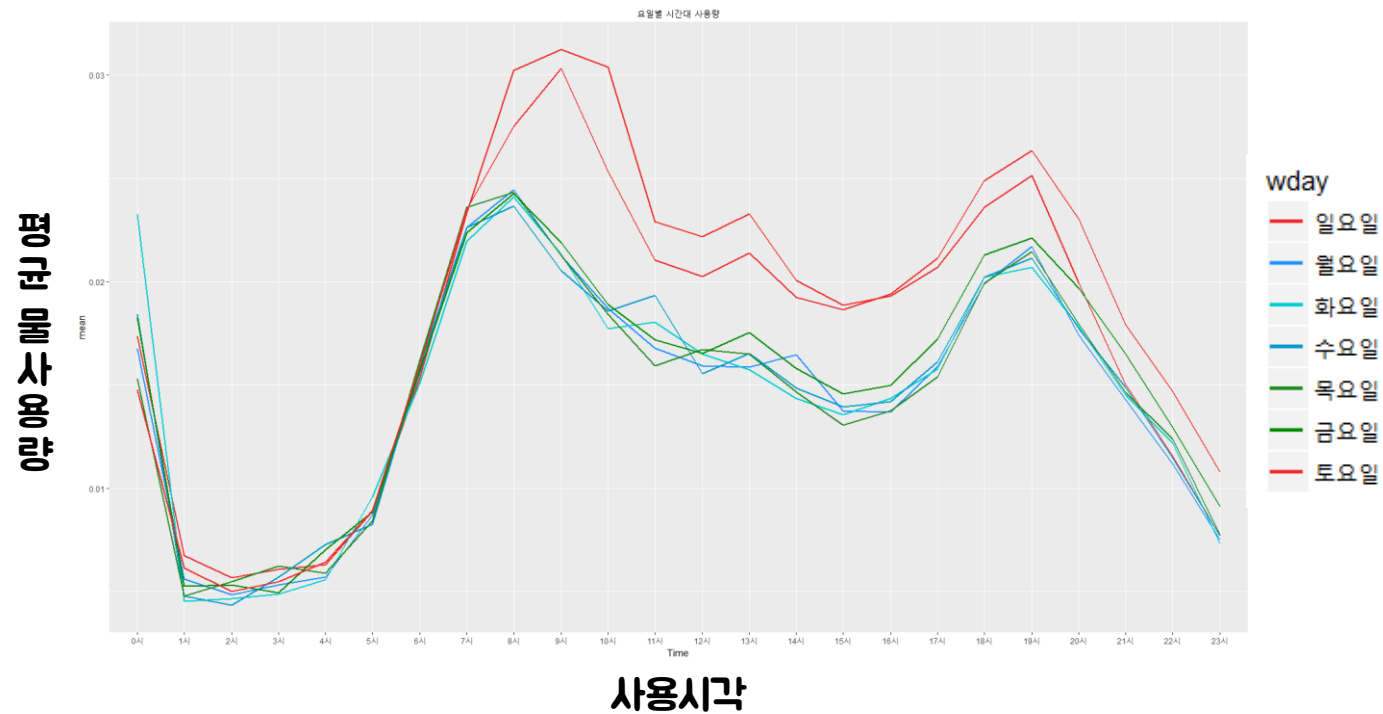
› 특히 여름에 가장 많은 물을 사용

이유 : 위생관리를 강화

물 사용을 효율화 패턴을 고려한 도축장 특별 물관리를 제공하는 것이 효과적일 것

요일효과

▶ 시각별 평균 물사용량은 요일에 따라 다른 양상을 보인다.



토요일과 일요일은
평일에 비해 평균 물사용량이 높다.

맞춤형 물 절약 방안 - 1. (주) 광축

서산시 전체 물 사용량의 9%를 사용

- › 필요성 : 가정의 자발적인 물 사용량 저감을 유도하는 것보다는 물을 많이 사용하는 수용가를 찾아내어 체계적인 물 관리 해결방법을 제공하는 것이 효과적일 것
- › 광축의 물 사용량 5%를 감소시킬 경우 ➡ 서산시 팔봉면의 수도용수를 연간 $1155m^3$ 확보가능
- 외국 사례 : 4년 동안 도축장의 물 감축 프로젝트를 진행 ➡ 물 사용량을 7% 가까이 줄임

› 도축장의 물사용량 감축 방법

1. 움직임 감지를 활용한 물 분사기를 설치한다.
2. 절연 칼을 사용한다.

도축장 물 사용량 중 5%가 칼 소독에 사용되는데, 절연 칼을 사용하면 소독 횟수를 줄일 수 있다.

3. 직원 교육 강화

맞춤형 물 절약 방안 - 2. 스마트 미터기와 연계하여 <스마트 물 절약 앱> 개발

IDEA : 스마트 미터기를 어플과 연동해서 실시간 물관리를 하자!

›스마트 미터기의 역할 : 물 수요 관리 대상을 파악 + 물 수요량을 예측

예상 효과

‘ 각 수용가에 실시간 물 사용량과, 물 사용을 줄일 수 있는 방법을 제공 ’

맞춤형 물 절약 방안 - 2. 스마트 미터기와 연계하여 <스마트 물 절약 앱> 개발

예시

1. 자신의 정보를 입력하면 스마트 미터기 사용 내역을 토대로 일반대수용가, 관광수용가, 가정수용가를 분류
2. 가정 수용가는 공휴일의 효과가 크므로, 공휴일에는 물 사용 절약과 관련된 정보를 더 많이 제공
3. 예상 물 사용량과 이에 따른 물 절약 대응방안 제공
4. 시계열 분석을 통해 추후 1개월 후 물 사용량이 얼마나 변할지 실시간 그래프 제공 -> 가뭄에 대비 가능
5. 시뮬레이션의 예상 결과보다 물 사용량이 많을 시 사용자들에게 알림을 제공

맞춤형 물 절약 방안 - 3. 지역축제와 물 홍보를 연계해 효과를 두 배로!

› 서산시의 내 관광수용가의 물 사용 패턴 -> 여름에 사용량 증가

 피서객을 비롯한 관광객들이 물을 많이 사용

› 해당 지역 축제(ex. 감자 축제)와 결부 -> 물 절약 캠페인을 진행

정부 주도의 물 절약 캠페인에서 벗어나,

지역 주민들이 주도하는 적극적인 물 절약 캠페인을 통해 물 절약 의식을 더욱 고취시킬 수 있다.

› 숙박시설에 절수형 샤워헤드를 설치 -> 관광객들이 절수형 샤워헤드를 사용

절수형 샤워헤드의 효과를 관광객들에게 홍보 서산시를 물 절약을 대표하는 도시로 만든다.