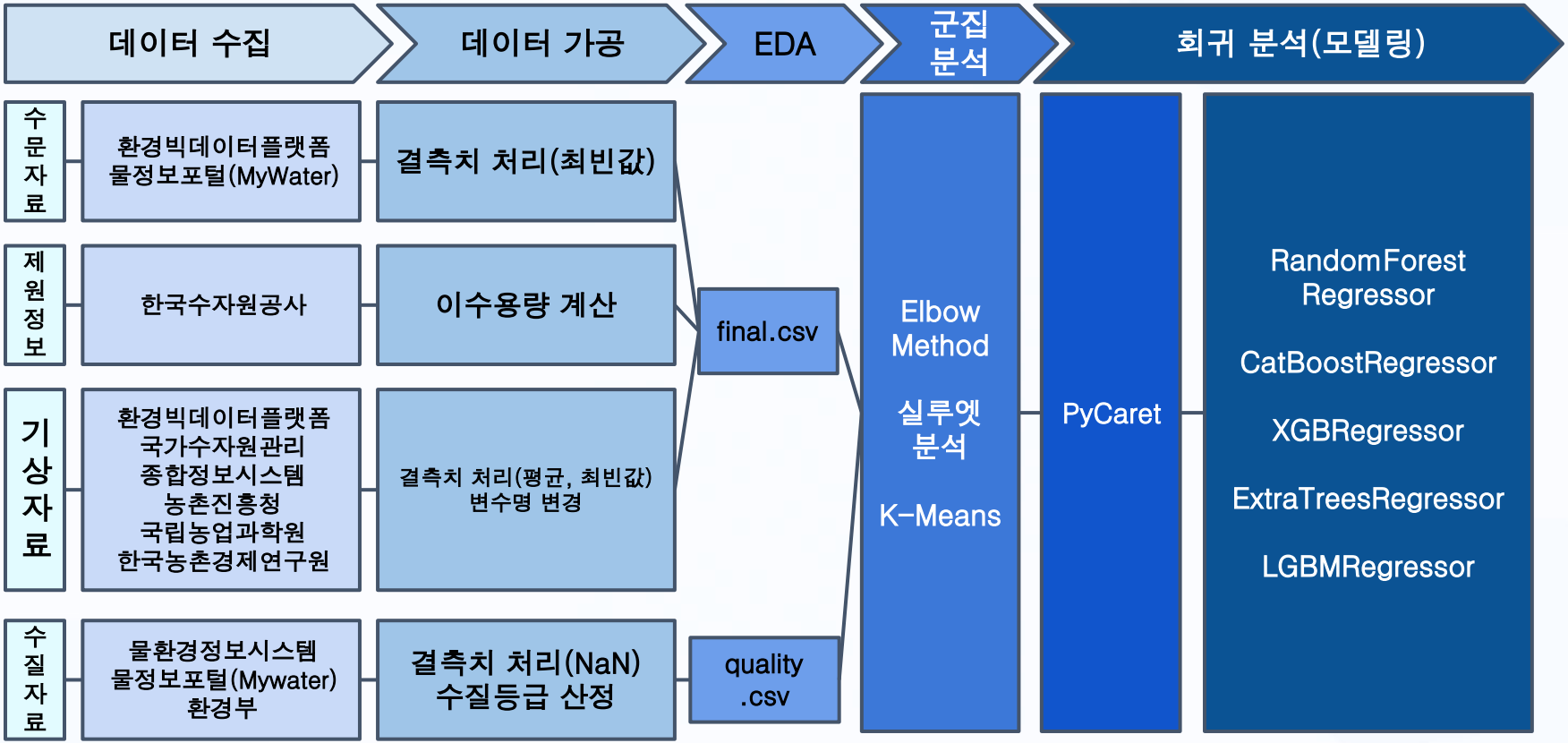
다목적 댐의 상태 평가와 예측을 위한 머신러닝 기반 모델 개발

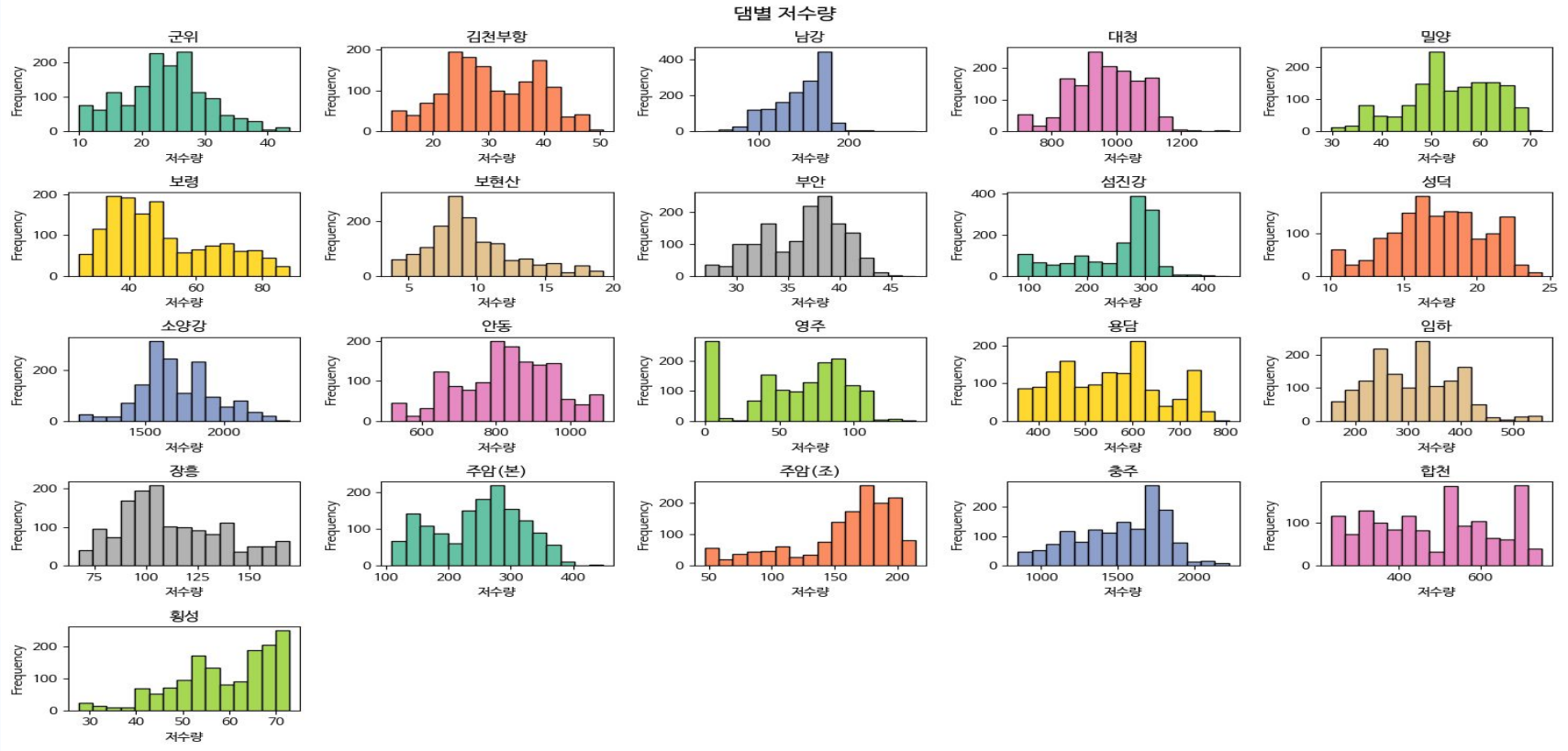
이화여자대학교
박신영
배주원
차수빈

목차

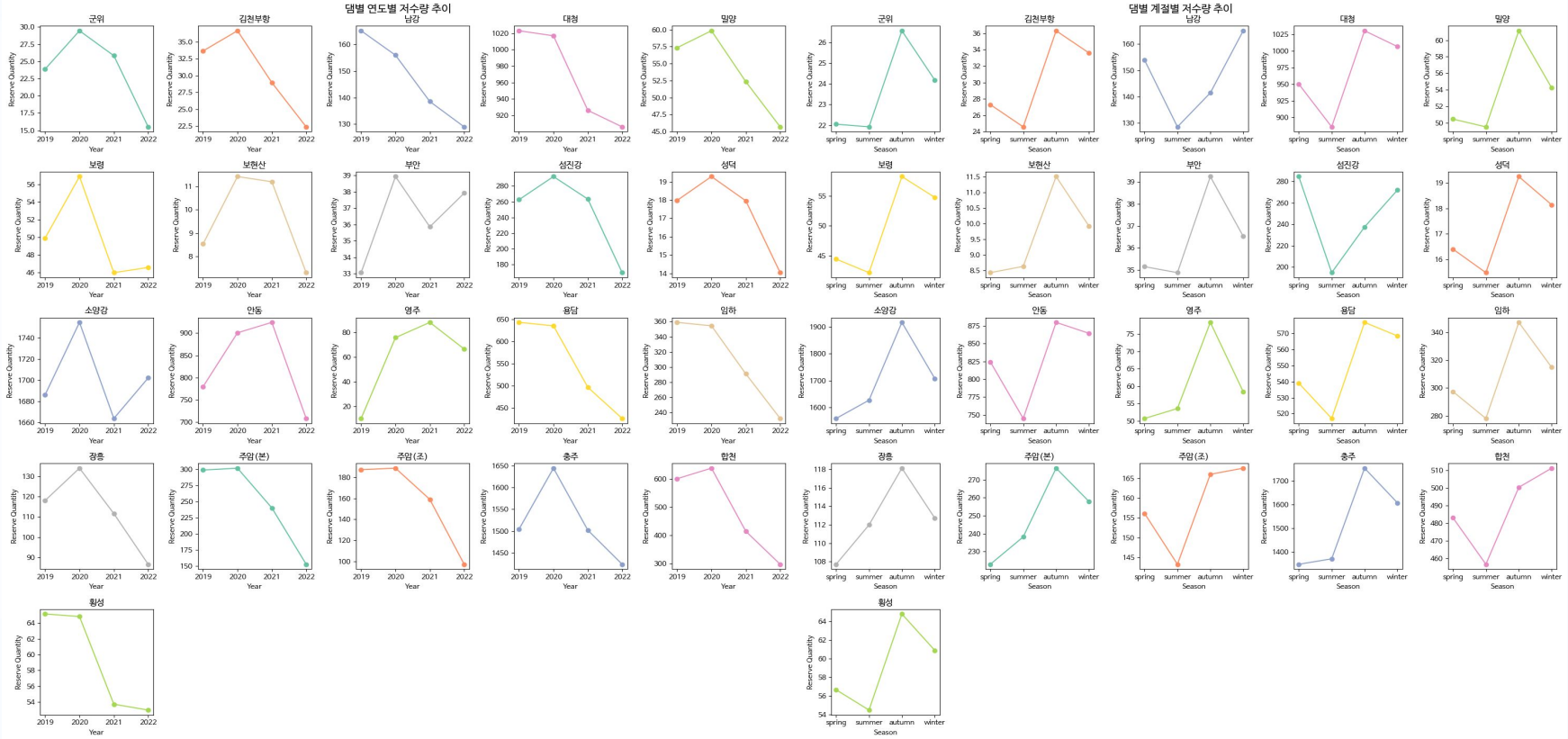
분석 프로세스


데이터 수집
데이터 가공
EDA
군집 분석
회귀 분석(모델링)
수문자료
환경빅데이터플랫폼 물정보포털(MyWater)
결측치 처리(최빈값)
제원정보
한국수자원공사
이수용량 계산
기상자료
환경빅데이터플랫폼 국가수자원관리 종합정보시스템 농촌진흥청 국립농업과학원 한국농촌경제연구원
결측치 처리(평균, 최빈값) 변수명 변경
수질자료
물환경정보시스템 물정보포털(Mywater) 환경부
결측치 처리(NaN) 수질등급 산정
final.csv
quality .csv
Elbow Method
실루엣 분석
K-Means
PyCaret
RandomForest Regressor
CatBoostRegressor
XGBRegressor
ExtraTreesRegressor
LGBMRegressor

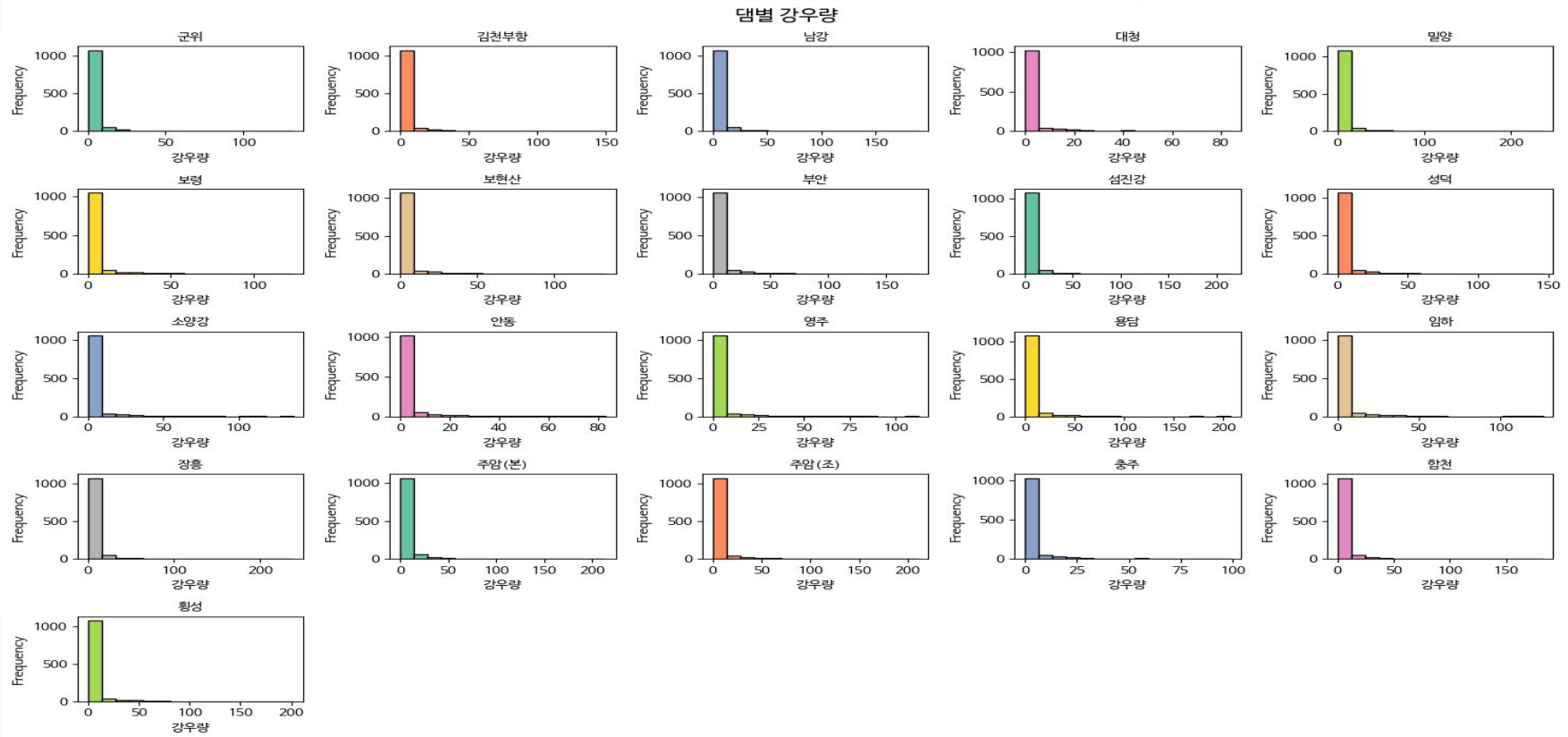
데이터 탐색 - 1) Target 변수



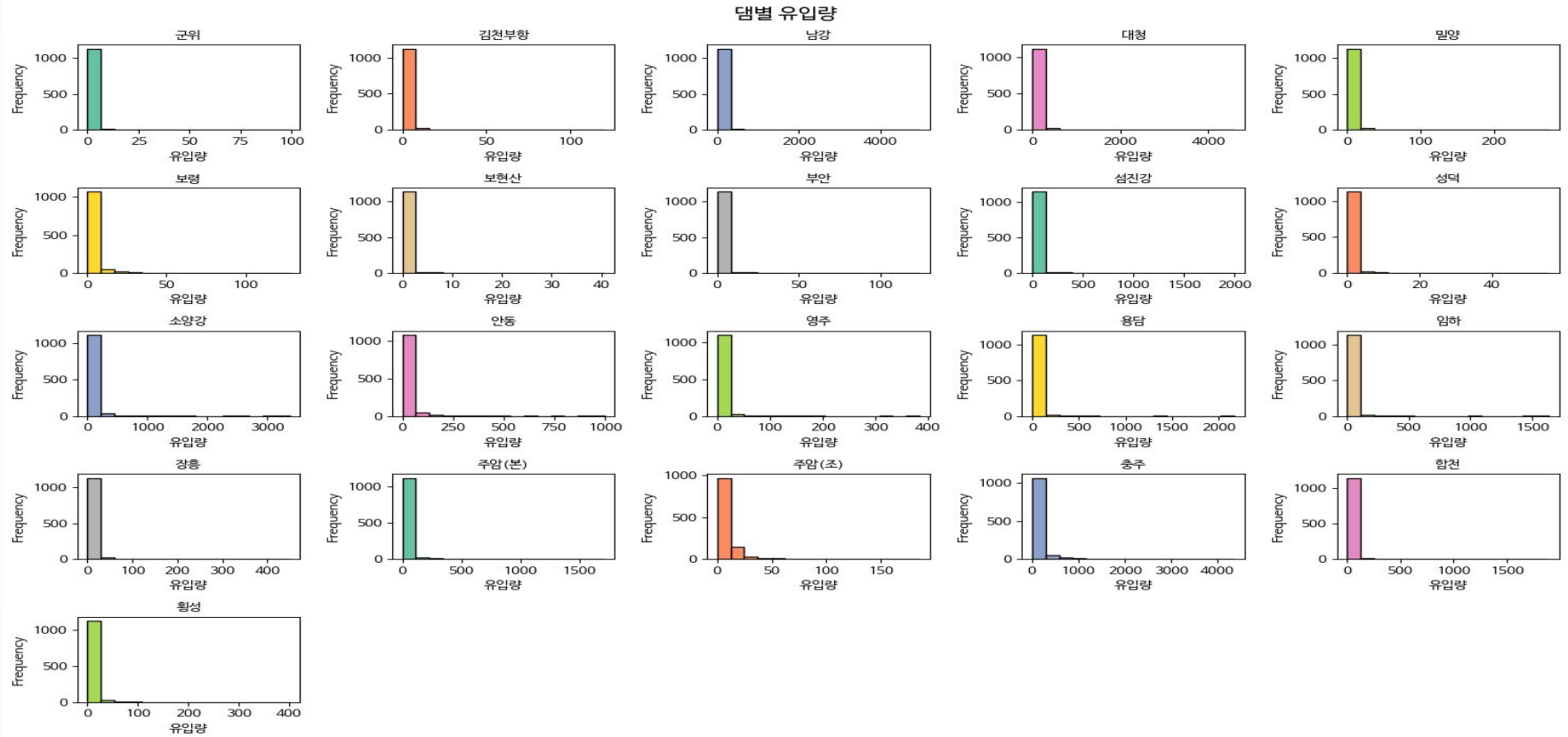
데이터 탐색 - 1) Target 변수



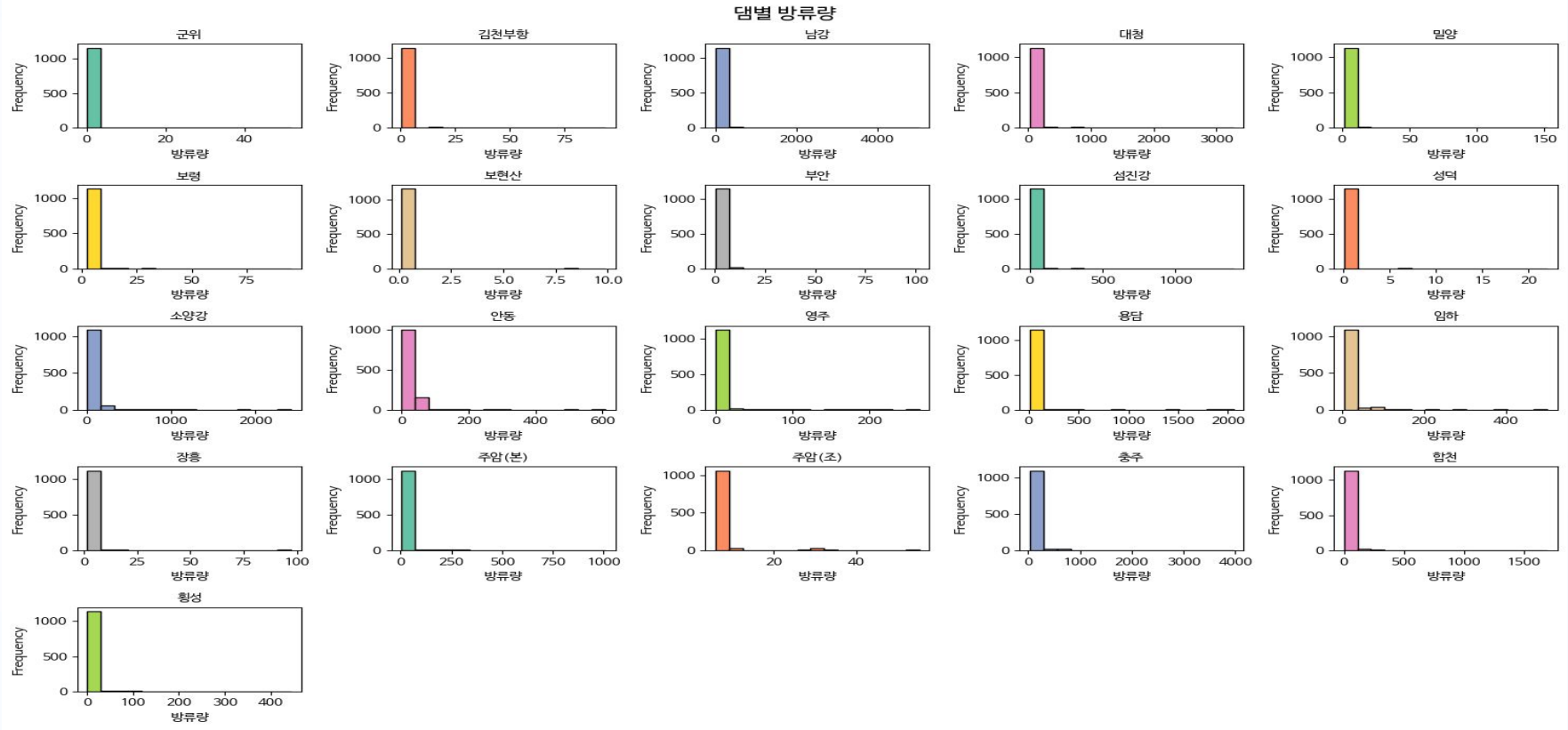
데이터 탐색 - 2) Feature 변수



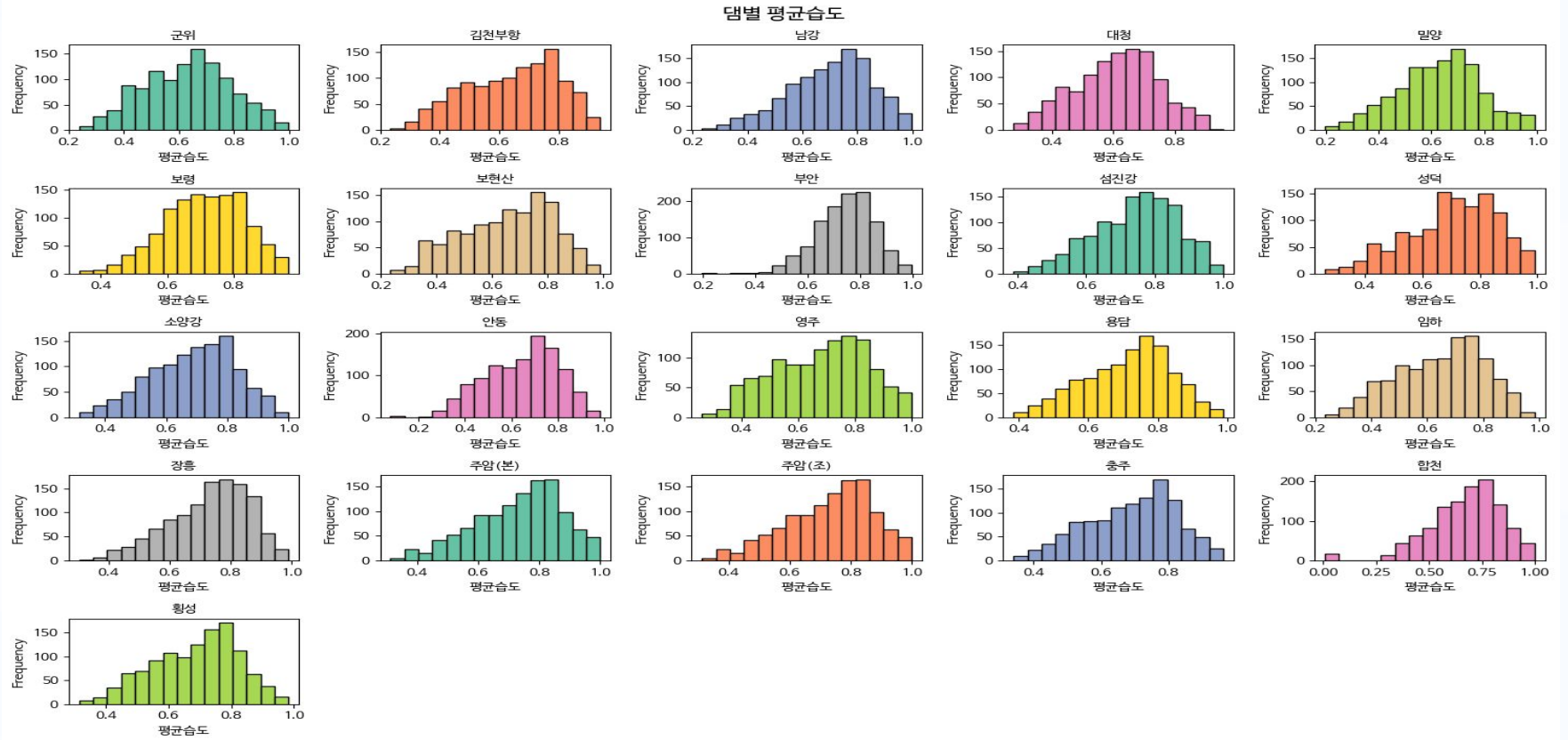
데이터 탐색 - 2) Feature 변수



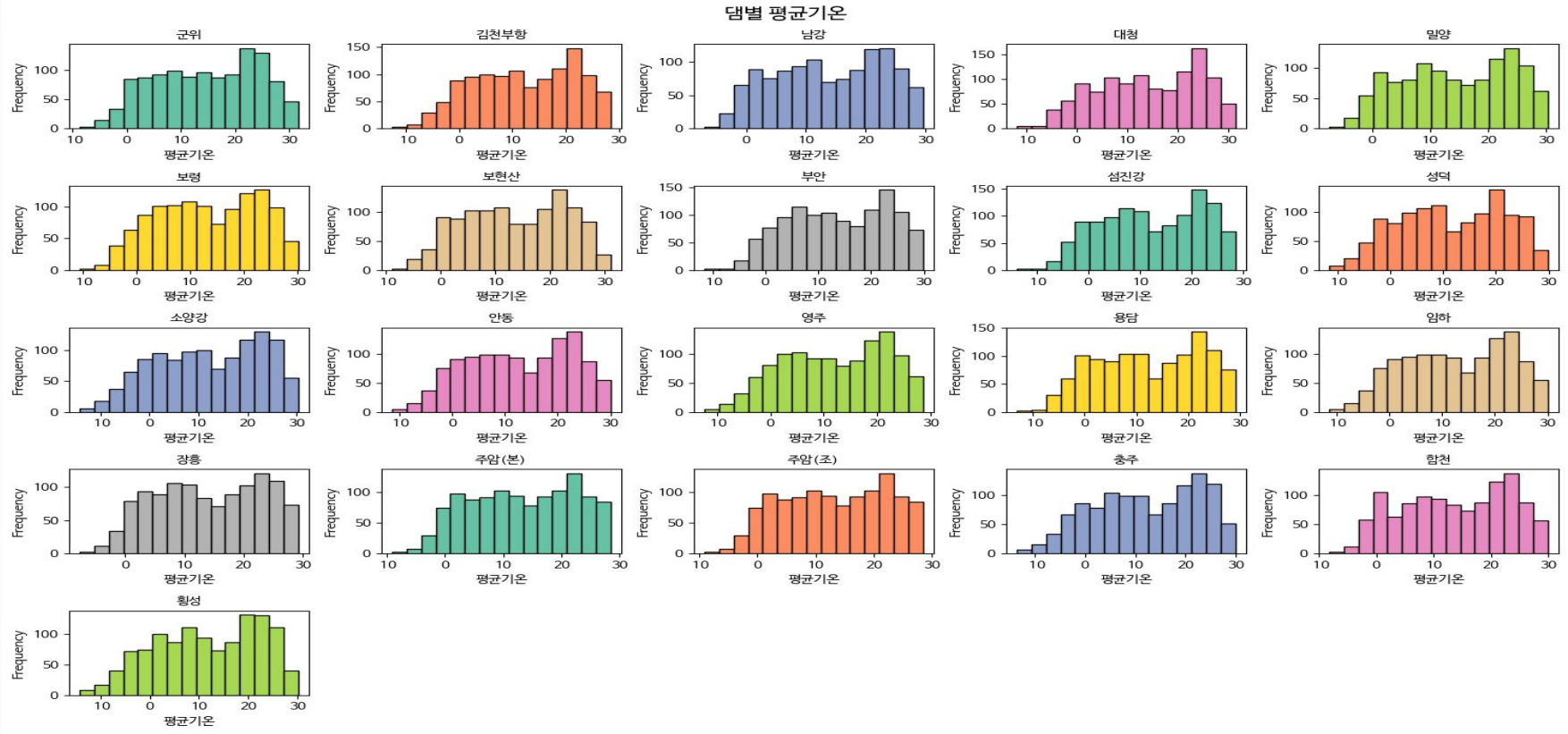
데이터 탐색 - 2) Feature 변수



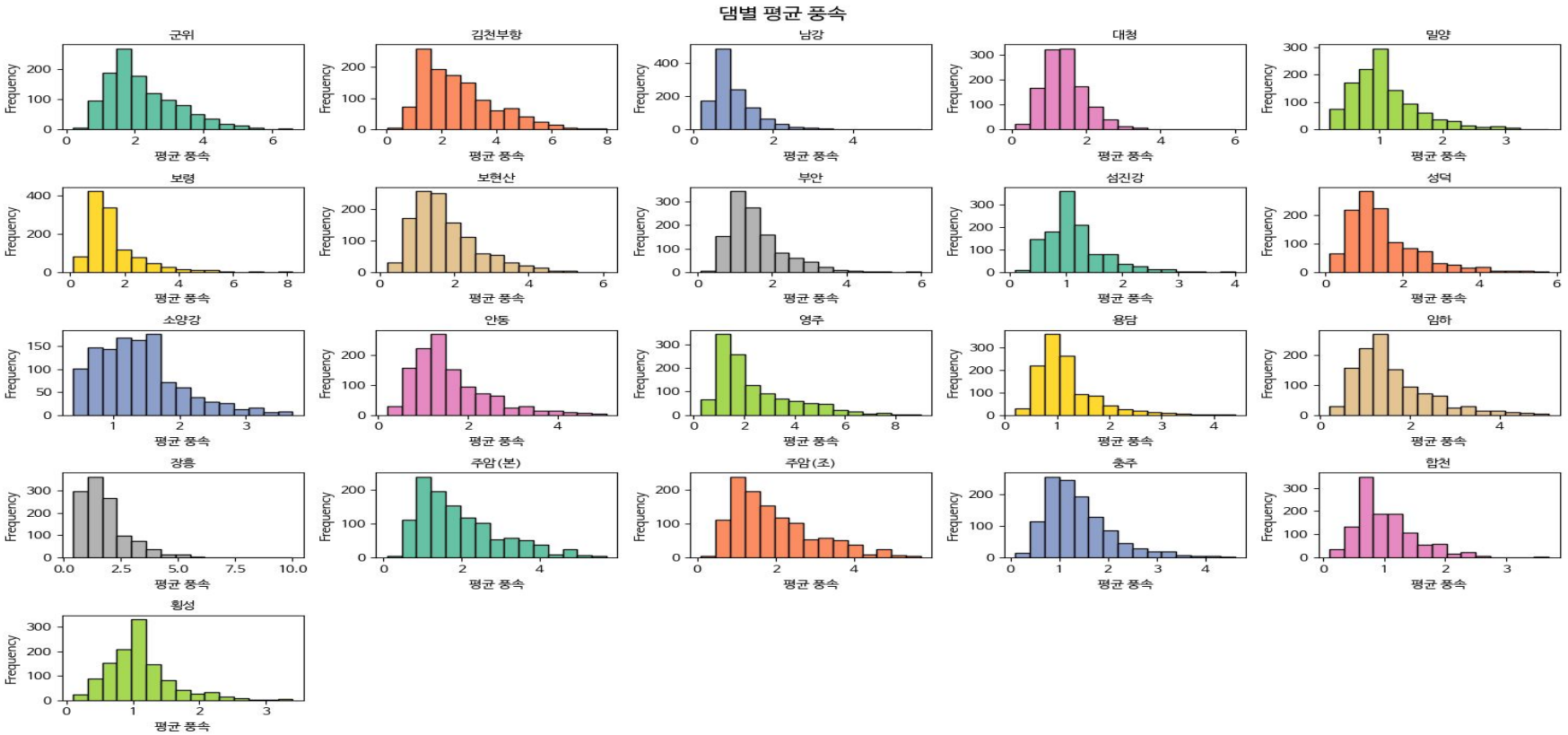
데이터 탐색 - 2) Feature 변수



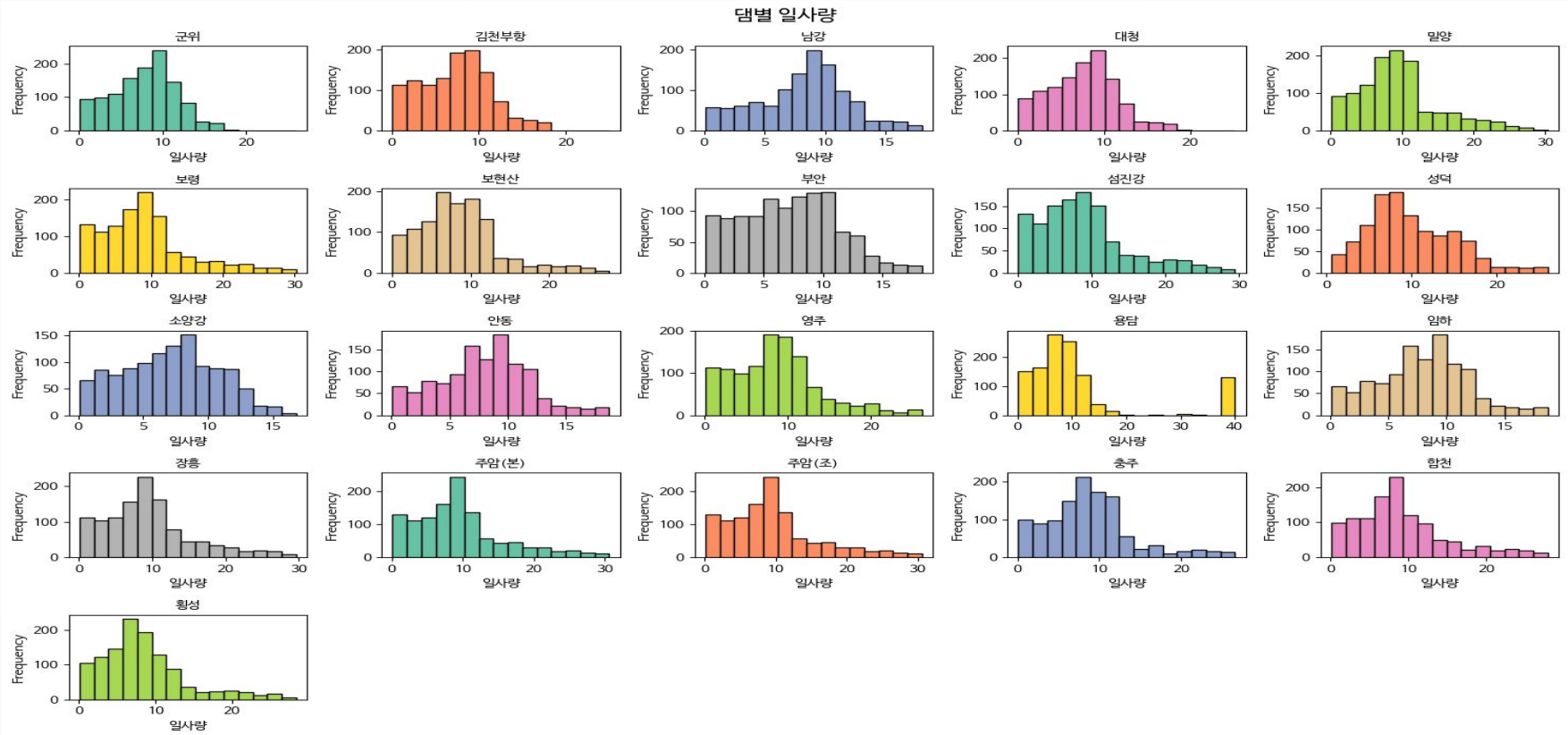
데이터 탐색 - 2) Feature 변수



데이터 탐색 - 2) Feature 변수



데이터 탐색 - 2) Feature 변수



데이터 탐색 결과

스케일과 분포 양상에 차이가 있음

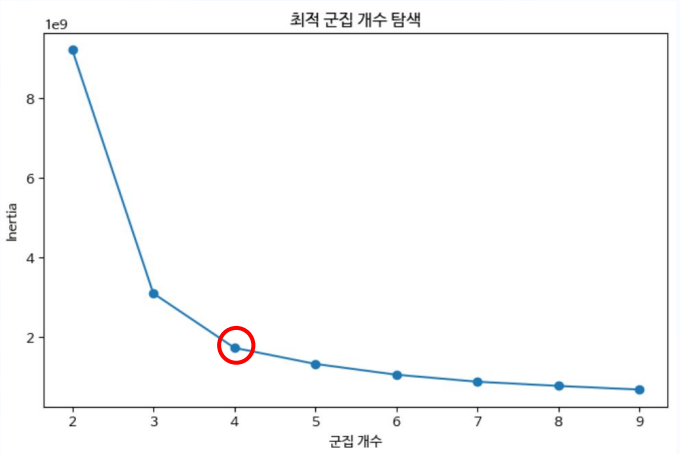
유사한 특성을 가진 댐별로 그룹화하기 위해
군집분석을 수행

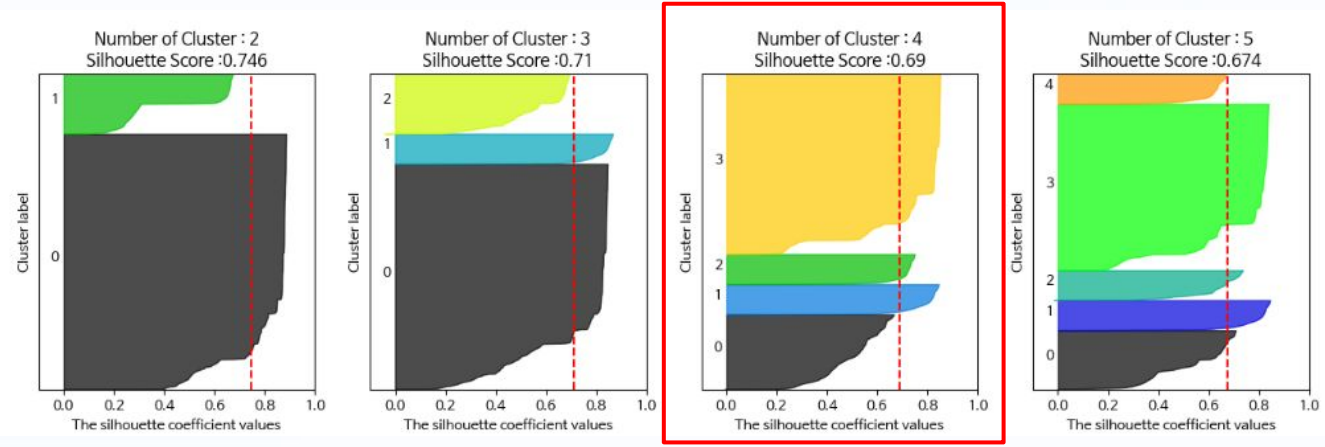
군집분석

⊙ 최적 군집 개수 탐색

- Elbow Method
- 실루엣 계수

Elbow Method ▸



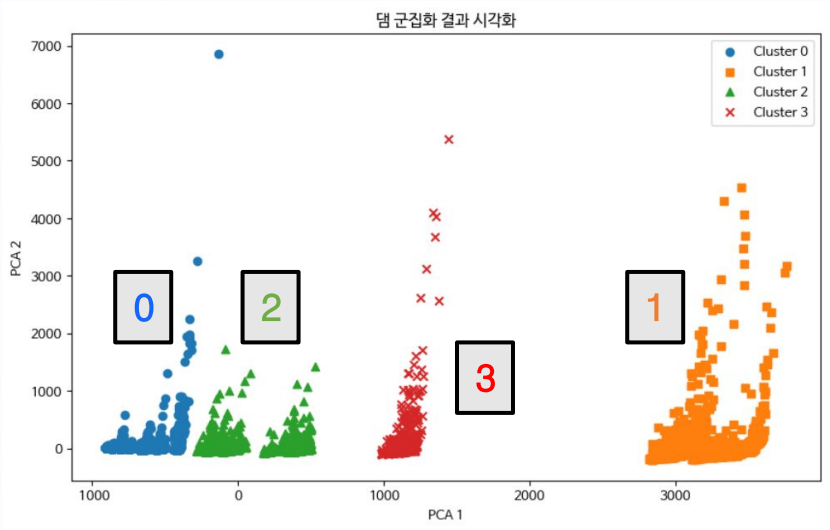


◂ 실루엣 분석

군집분석

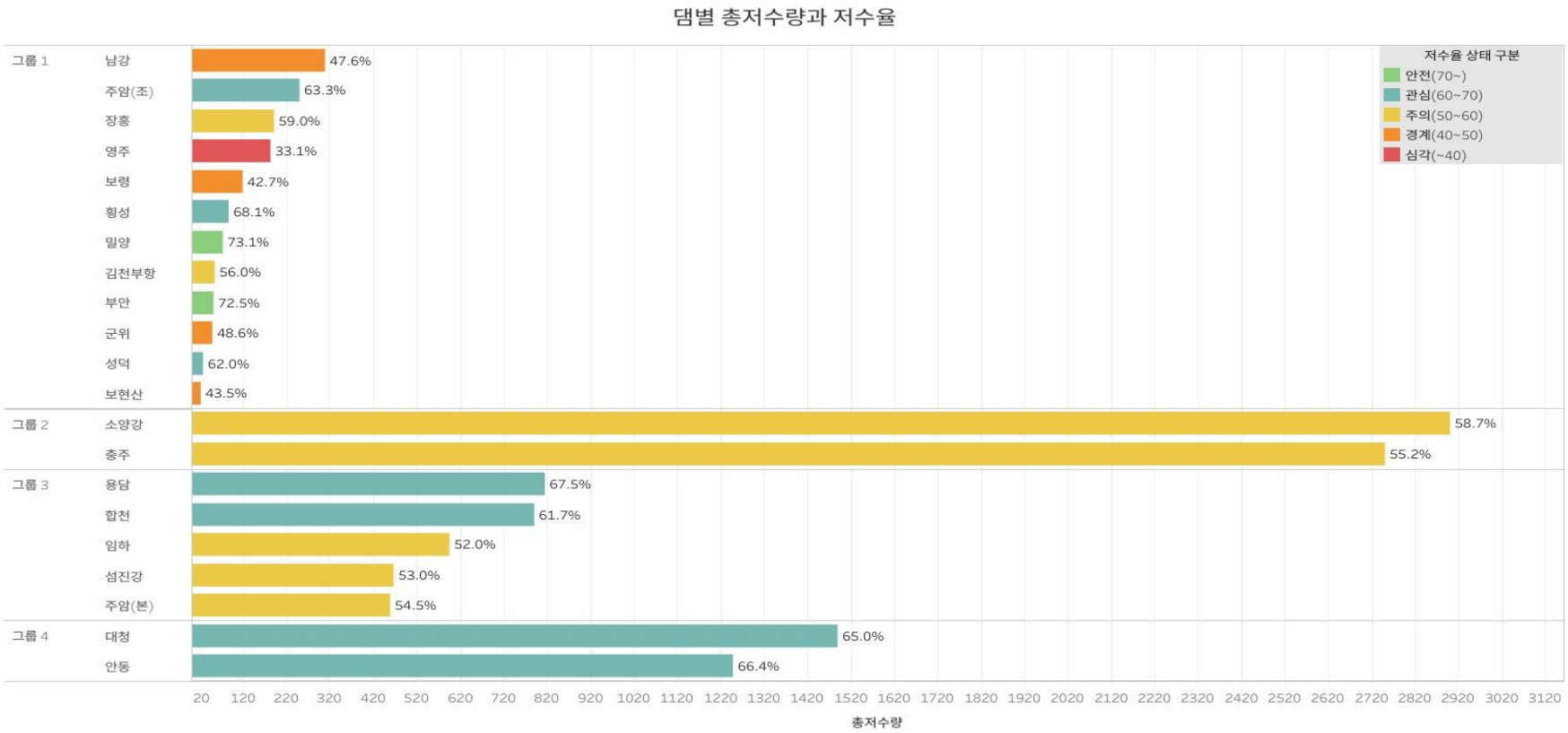
⊙ 군집화 결과

- group 1(Cluster 0) : 군위, 김천부항, 남강, 밀양, 보령, 보현산, 부안, 성덕, 영주, 장흥, 주암(조). 횡성
- group 2(Cluster 1) : 소양강, 충주
- group 3(Cluster 2) : 섬진강, 용담, 임하, 주암(본), 합천
- group 4(Cluster 3) : 대청, 안동



군집분석

Tableau 시각화



군집분석

⊙ 그룹 1

(군위, 김천부항, 남강, 밀양, 보령, 보현산, 부안, 성덕, 영주, 장흥, 주암(조). 횡성)

- 저수량: 낮음
- 저수율: 낮음(약 55%)
- 유입량, 방류량: 낮음
- 평균풍속, 습도는 안정적

⊙ 그룹 2

(소양강, 충주)

- 저수량: 가장 큼(1조 6000억 m^3)
- 저수율: 낮음(약 57%) -> 주의 필요
- 유입량, 방류량: 큼
- 낮은 기온, 적은 강우량과 일사량

⊙ 그룹 3

(섬진강, 용담, 임하, 주암(본), 합천)

- 저수량: 큼
- 저수율: 낮음(60% 미만)
- 2020년 이후로 저수량 감소 추세
- 유입량, 방류량: 작음
- 습도와 일사량 변동 폭이 큼

⊙ 그룹 4

(대청, 안동)

- 저수량: 큼
- 저수율: 안정적(60% 이상)
- 유입량, 방류량: 상당히 큼
- 평균 습도 낮음, 강우량은 평균 정도
- 높은 기온과 큰 풍속
- 일사량 변동 큼

회귀분석(모델링)

⦿ AutoML – PyCaret

– 최적 알고리즘 선정

Group 1

	Model	RMSE	R2
et	Extra Trees Regressor	0.0715	0.9939
catboost	CatBoost Regressor	0.0973	0.9891
rf	Random Forest Regressor	0.1074	0.9860
xgboost	Extreme Gradient Boosting	0.1216	0.9825
dt	Decision Tree Regressor	0.1314	0.9791
lightgbm	Light Gradient Boosting Machine	0.1450	0.9754

Group 2

	Model	RMSE	R2
et	Extra Trees Regressor	57.7325	0.9563
catboost	CatBoost Regressor	68.6755	0.9383
rf	Random Forest Regressor	70.2862	0.9355
xgboost	Extreme Gradient Boosting	75.7080	0.9247
lightgbm	Light Gradient Boosting Machine	79.3183	0.9176

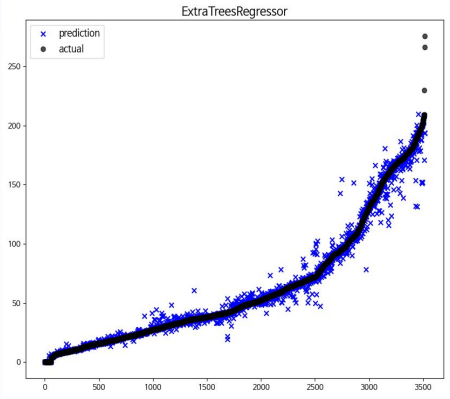
Group 3

	Model	RMSE	R2
et	Extra Trees Regressor	17.7554	0.9878
catboost	CatBoost Regressor	21.5037	0.9823
rf	Random Forest Regressor	23.0638	0.9796
xgboost	Extreme Gradient Boosting	24.6570	0.9767
lightgbm	Light Gradient Boosting Machine	25.3933	0.9754

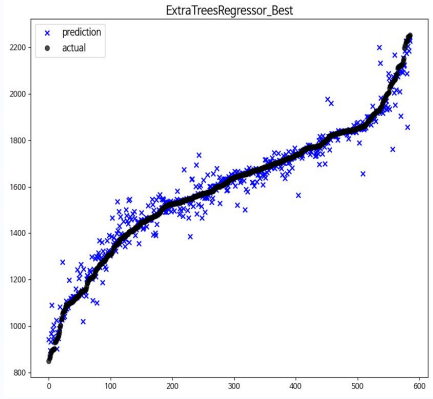
Group 4

	Model	RMSE	R2
et	Extra Trees Regressor	27.6604	0.9559
catboost	CatBoost Regressor	32.4148	0.9400
rf	Random Forest Regressor	33.7301	0.9354
lightgbm	Light Gradient Boosting Machine	35.6573	0.9281
xgboost	Extreme Gradient Boosting	36.0730	0.9260

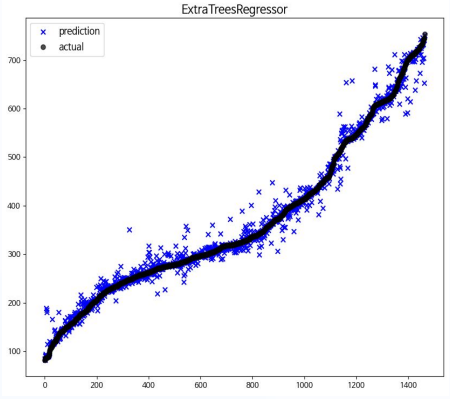
회귀분석(모델링) - 회귀 곡선 시각화



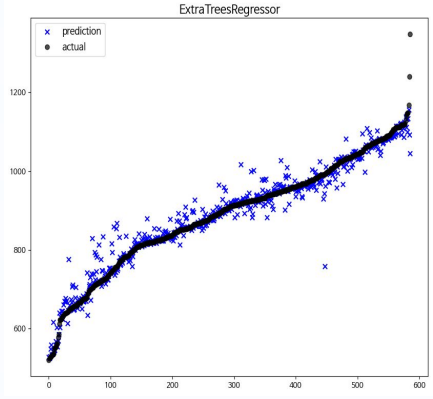
⊙ 그룹 1
(RMSE: 5.072, Adjusted R2: 0.9904)



⊙ 그룹 2
(RMSE: 52.185, Adjusted R2: 0.9651)



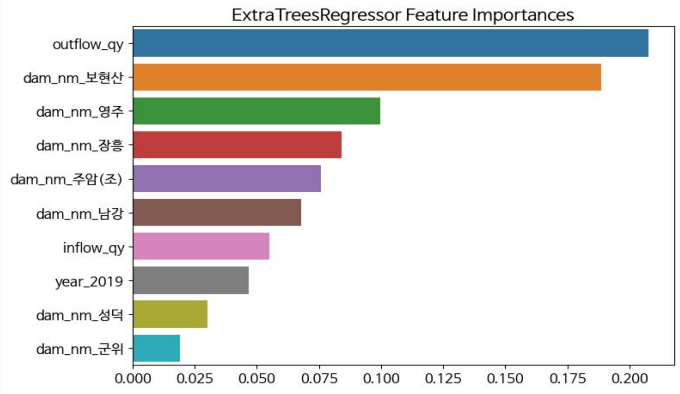
⊙ 그룹 3
(RMSE: 16.460, Adjusted R2: 0.9895)



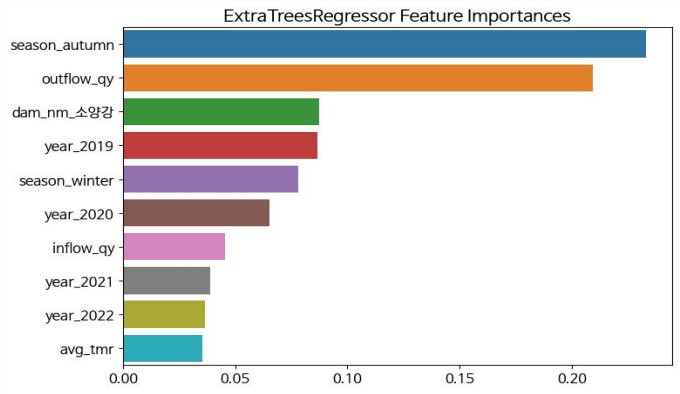
⊙ 그룹 4
(RMSE: 29.469, Adjusted R2: 0.9556)

회귀분석(모델링) - 피처 중요도 시각화

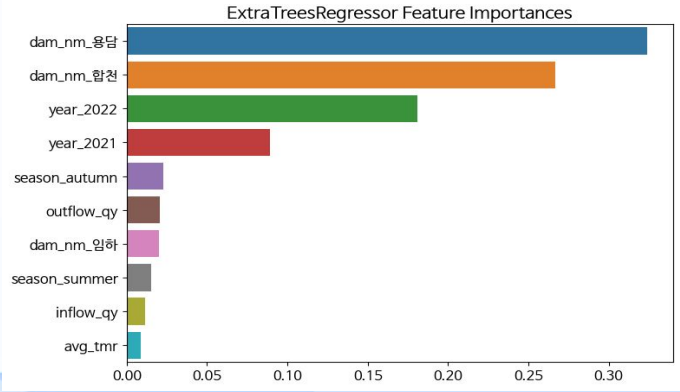
⦿ 그룹 1



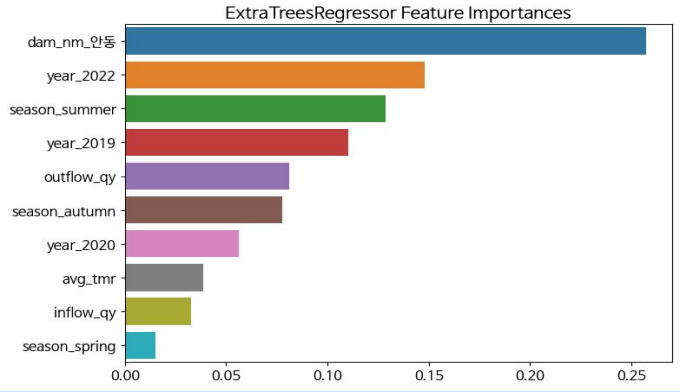
⦿ 그룹 2



⦿ 그룹 3

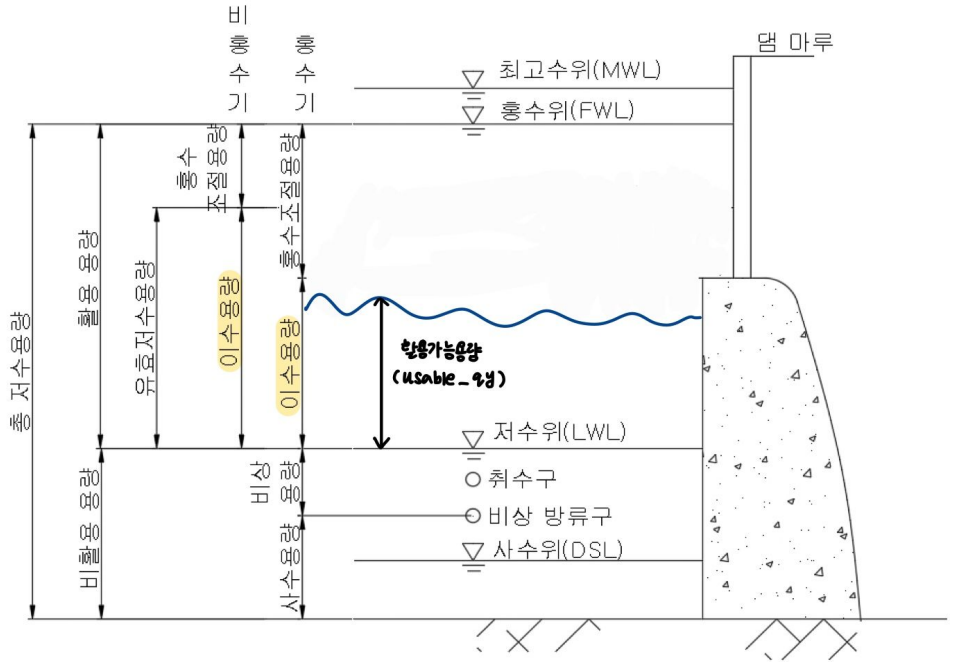


⦿ 그룹 4



결과 및 기대효과

$$\text{efficiency} = \frac{usable\_qy}{maximum\_use\_qy} = \frac{reserve\_qy - unused\_qy}{maximum\_use\_qy}$$



[변수 설명]

tot_qy: 총저수량

valid_qy: 유효저수용량

flood_qy: 홍수조절용량

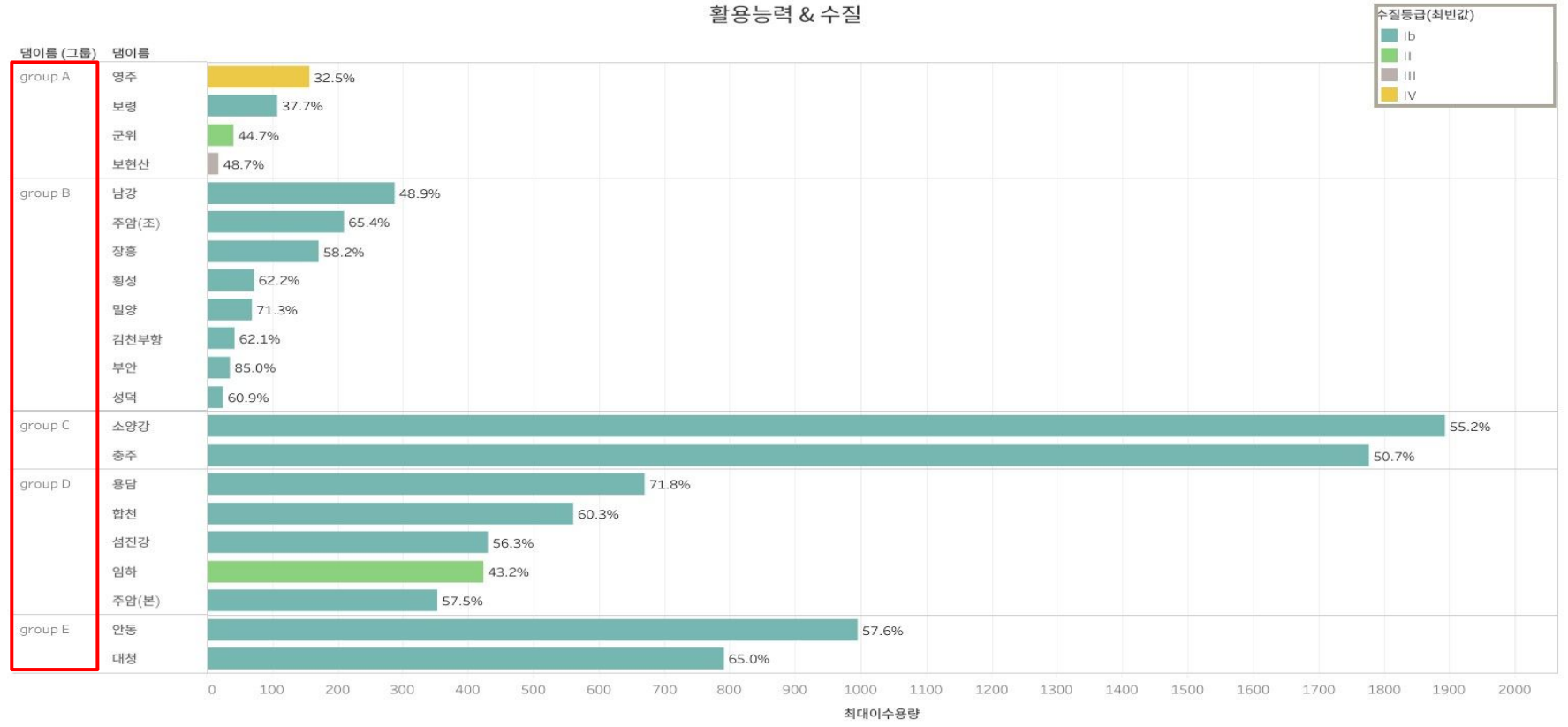
unused_qy: 비활용용량

usable_qy: 활용 가능 용량

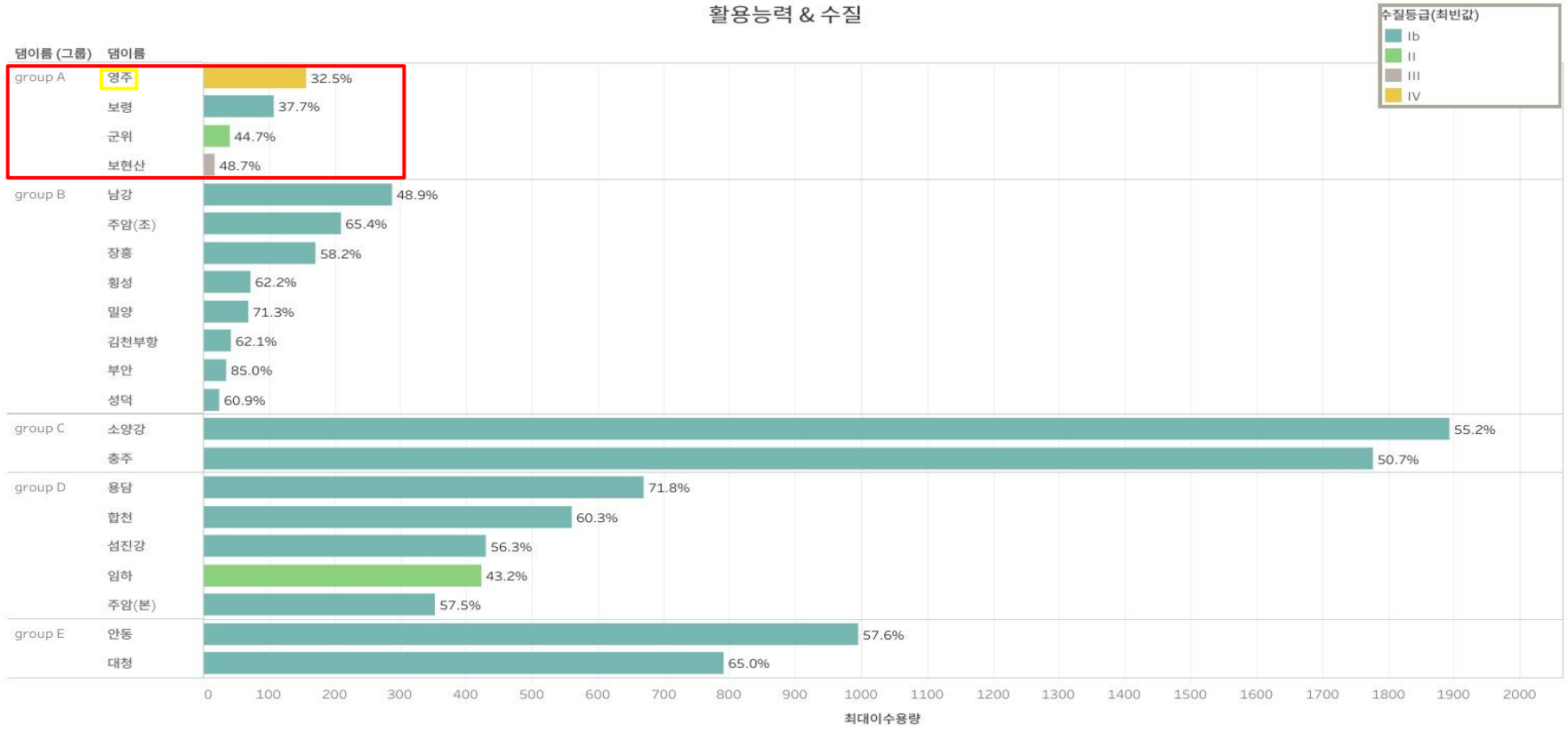
maximum_use_qy: 이수용량

efficiency: 활용 능력

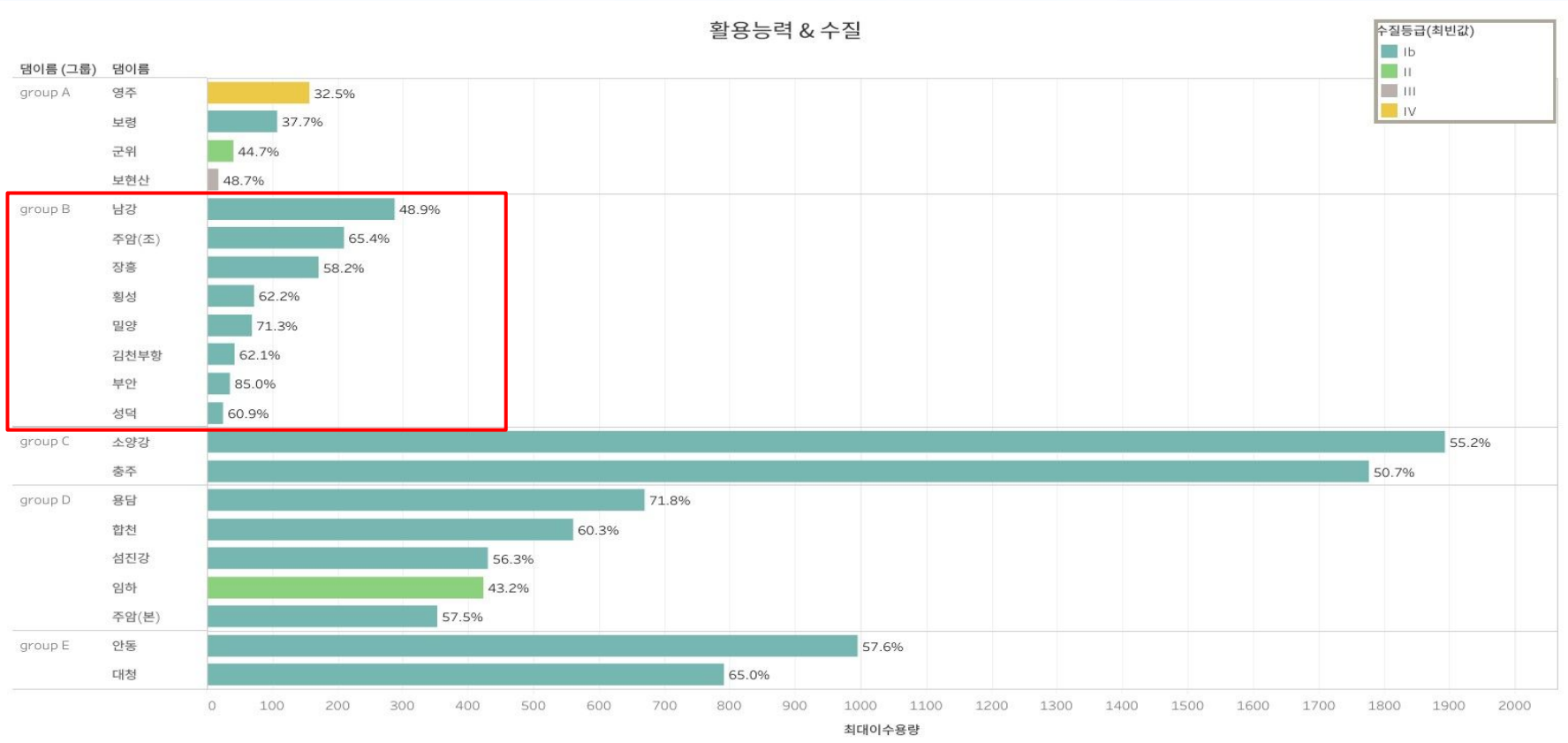
결과 및 기대효과


활용능력 & 수질
수질등급(최빈값)
Ib
II
III
IV
댐이름 (그룹)
댐이름
group A
영주 32.5%
보령 37.7%
군위 44.7%
보현산 48.7%
group B
남강 48.9%
주암(조) 65.4%
장흥 58.2%
횡성 62.2%
밀양 71.3%
김천부항 62.1%
부안 85.0%
성덕 60.9%
group C
소양강 55.2%
충주 50.7%
group D
용담 71.8%
합천 60.3%
섬진강 56.3%
임하 43.2%
주암(본) 57.5%
group E
안동 57.6%
대청 65.0%
0 100 200 300 400 500 600 700 800 900 1000 1100 1200 1300 1400 1500 1600 1700 1800 1900 2000
최대이수용량

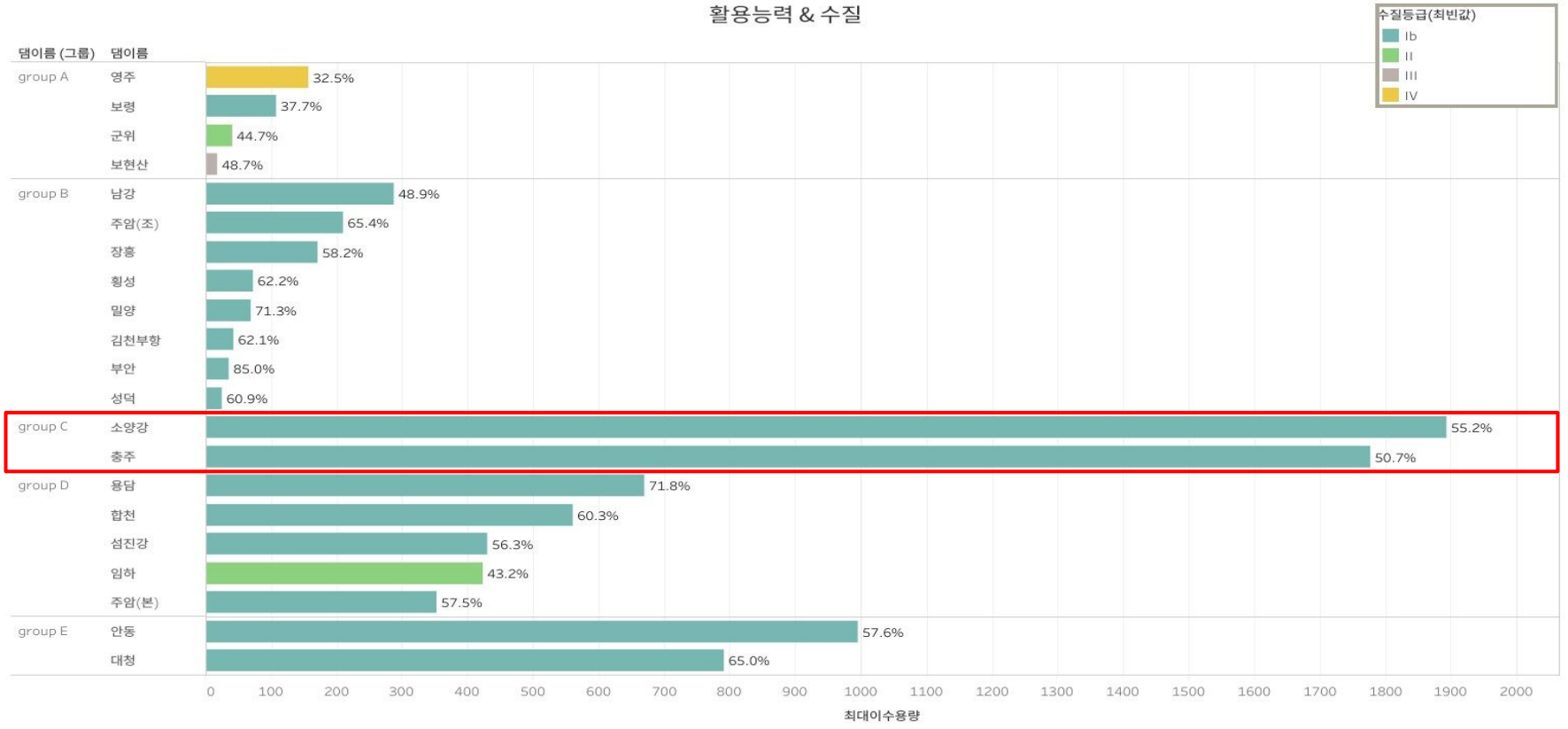
결과 및 기대효과


활용능력 & 수질
수질등급(최빈값)
Ib
II
III
IV
댐이름 (그룹)
댐이름
group A
영주 32.5%
보령 37.7%
군위 44.7%
보현산 48.7%
group B
남강 48.9%
주암(조) 65.4%
장흥 58.2%
횡성 62.2%
밀양 71.3%
김천부항 62.1%
부안 85.0%
성덕 60.9%
group C
소양강 55.2%
충주 50.7%
group D
용담 71.8%
합천 60.3%
섬진강 56.3%
임하 43.2%
주암(본) 57.5%
group E
안동 57.6%
대청 65.0%
0 100 200 300 400 500 600 700 800 900 1000 1100 1200 1300 1400 1500 1600 1700 1800 1900 2000
최대이수용량

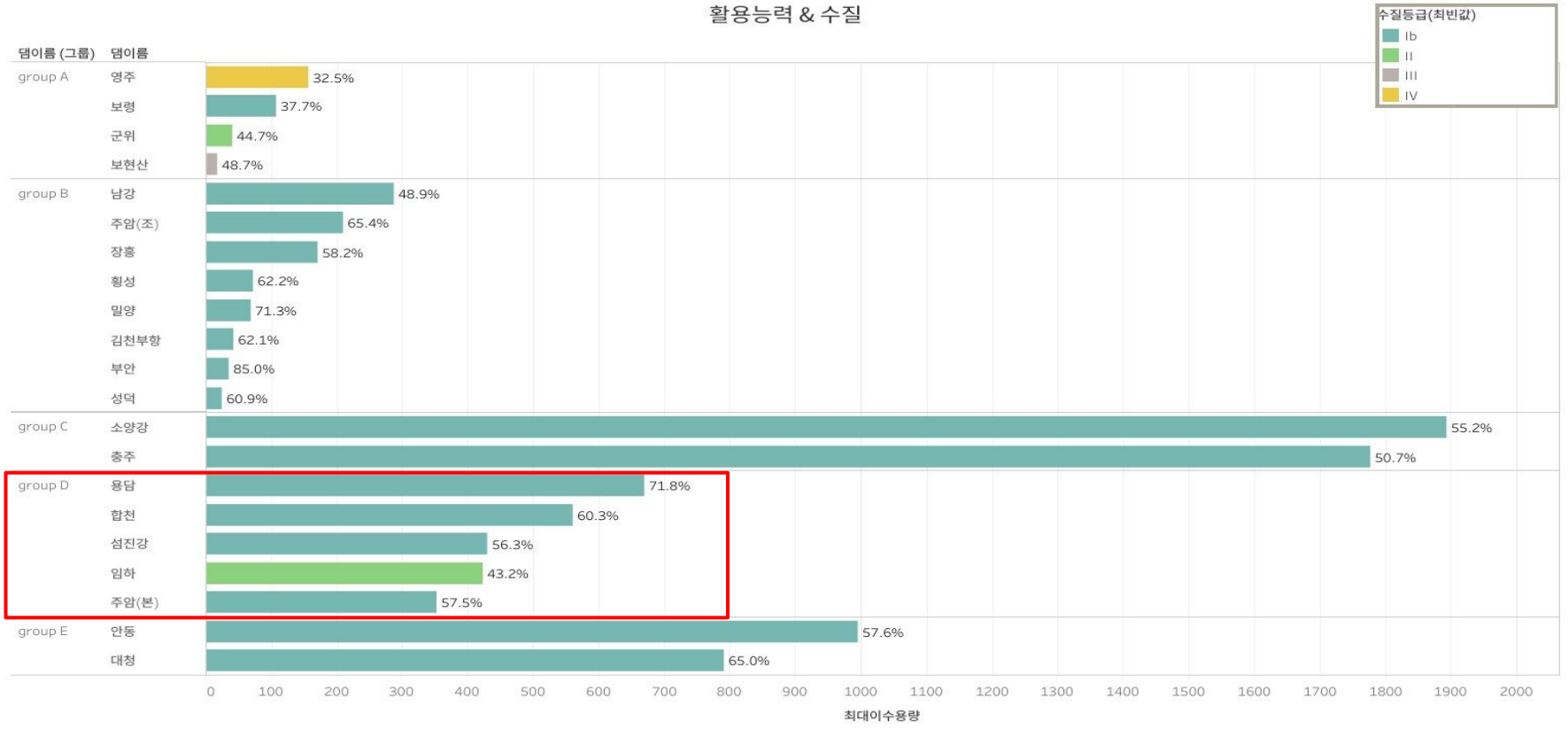
결과 및 기대효과


활용능력 & 수질
수질등급(최빈값)
Ib
II
III
IV
댐이름 (그룹)
댐이름
group A
영주 32.5%
보령 37.7%
군위 44.7%
보현산 48.7%
group B
남강 48.9%
주암(조) 65.4%
장흥 58.2%
횡성 62.2%
밀양 71.3%
김천부항 62.1%
부안 85.0%
성덕 60.9%
group C
소양강 55.2%
충주 50.7%
group D
용담 71.8%
합천 60.3%
섬진강 56.3%
임하 43.2%
주암(본) 57.5%
group E
안동 57.6%
대청 65.0%
0 100 200 300 400 500 600 700 800 900 1000 1100 1200 1300 1400 1500 1600 1700 1800 1900 2000
최대이수용량

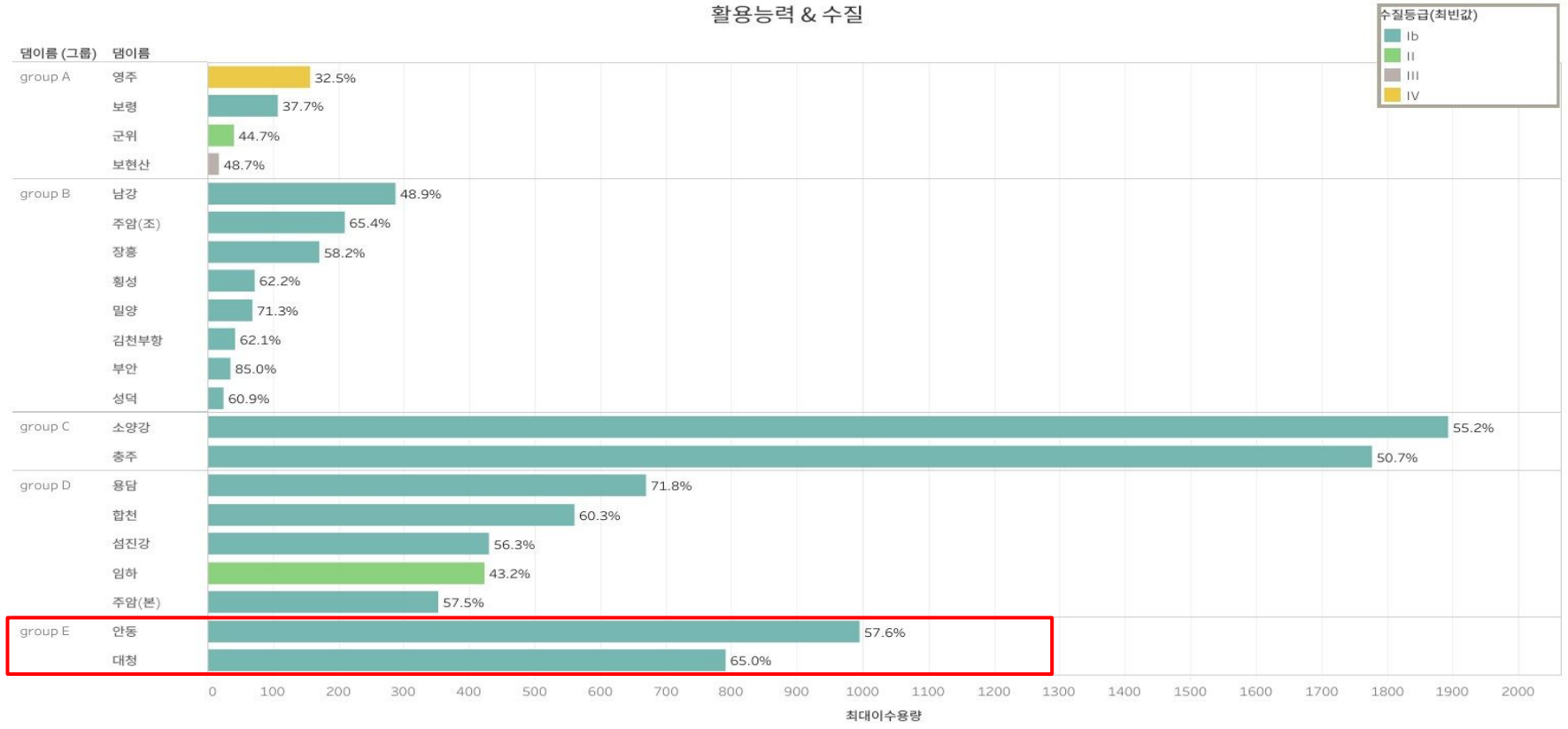
결과 및 기대효과


활용능력 & 수질
수질등급(최빈값)
Ib
II
III
IV
댐이름 (그룹)
댐이름
group A
영주 32.5%
보령 37.7%
군위 44.7%
보현산 48.7%
group B
남강 48.9%
주암(조) 65.4%
장흥 58.2%
횡성 62.2%
밀양 71.3%
김천부항 62.1%
부안 85.0%
성덕 60.9%
group C
소양강 55.2%
충주 50.7%
group D
용담 71.8%
합천 60.3%
섬진강 56.3%
임하 43.2%
주암(본) 57.5%
group E
안동 57.6%
대청 65.0%
0 100 200 300 400 500 600 700 800 900 1000 1100 1200 1300 1400 1500 1600 1700 1800 1900 2000
최대이수용량

결과 및 기대효과


활용능력 & 수질
수질등급(최빈값)
Ib
II
III
IV
댐이름 (그룹)
댐이름
group A
영주 32.5%
보령 37.7%
군위 44.7%
보현산 48.7%
group B
남강 48.9%
주암(조) 65.4%
장흥 58.2%
횡성 62.2%
밀양 71.3%
김천부항 62.1%
부안 85.0%
성덕 60.9%
group C
소양강 55.2%
충주 50.7%
group D
용담 71.8%
합천 60.3%
섬진강 56.3%
임하 43.2%
주암(본) 57.5%
group E
안동 57.6%
대청 65.0%
0 100 200 300 400 500 600 700 800 900 1000 1100 1200 1300 1400 1500 1600 1700 1800 1900 2000
최대이수용량

결과 및 기대효과


활용능력 & 수질
수질등급(최빈값)
Ib
II
III
IV
댐이름 (그룹)
댐이름
group A
영주 32.5%
보령 37.7%
군위 44.7%
보현산 48.7%
group B
남강 48.9%
주암(조) 65.4%
장흥 58.2%
횡성 62.2%
밀양 71.3%
김천부항 62.1%
부안 85.0%
성덕 60.9%
group C
소양강 55.2%
충주 50.7%
group D
용담 71.8%
합천 60.3%
섬진강 56.3%
임하 43.2%
주암(본) 57.5%
group E
안동 57.6%
대청 65.0%
0 100 200 300 400 500 600 700 800 900 1000 1100 1200 1300 1400 1500 1600 1700 1800 1900 2000
최대이수용량

감사합니다