



22-1 데이터마이닝 기말 프로젝트

건물의 전력량 예측



2 1 2 S T G 0 2 고 정 욱

2 1 2 S T G 0 4 김 이 현

2 1 2 S T G 1 8 예 지 혜

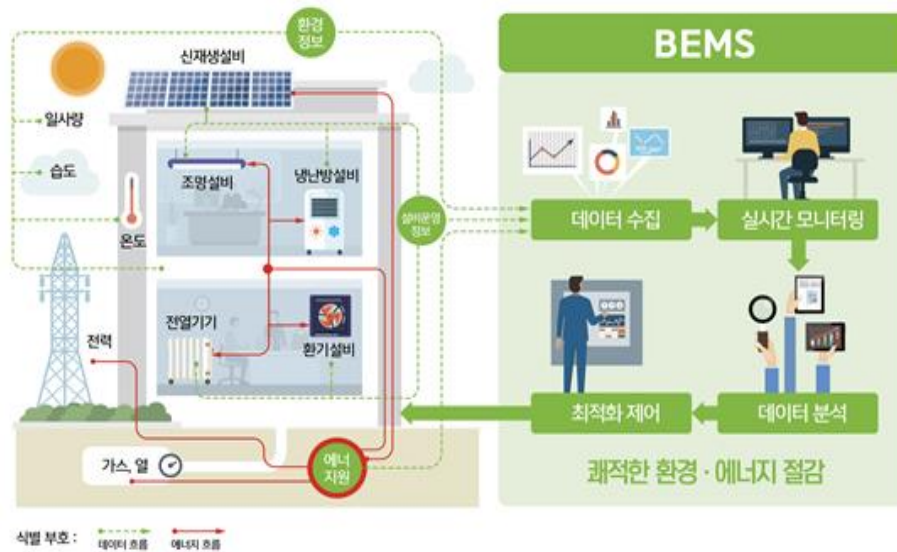
목차

1. 문제 정의
2. 데이터 소개
3. EDA
4. 모델링
5. 결론



1. 문제정의

- 전 세계 에너지 소비량의 약 40%가 건물에서 소비
- 그린뉴딜·탄소중립 실현의 핵심 사업으로 '건물 에너지 관리 시스템(BEMS)'이 주목받고 있음.
- 'BEMS(Building Energy Management System)'는 건물 내 주요 공간·설비에 부착한 센서를 통해 실시간으로 에너지사용 데이터를 수집·분석하고, 에너지소비 절감과 건물의 쾌적한 실내환경 유지에 활용되는 최첨단 ICT 시스템
- ESG경영을 위해 많은 회사들이 에너지 감축에 힘을 쓰고 있는 상황.



출처 : 전기신문
(<https://www.electimes.com>)

"이 건물 전기사용량이 지나치게 많네요". AI·빅데이터로 에너지낭비 막는다

건물·에너지 사용량 분석해
도시단위 '에너지맵' 개발

냉난방 온도·조명 제어 등
기업별 맞춤 솔루션 제공

오대석 기자 | 입력 : 2021.12.15 17:12:23 수정 : 2021.12.15 22:18:31 0

◆ ESG 경영현장 ◆



신재생에너지 전환과 같은 현재의 에너지 정책은 누구를 위한 것일까. 그리고 누가 실천할 수 있는 정책일까.

지난 9월 영국 전력 도매 시장에서 거래된 전기요금(약 70만원)을 넘어섰다. 작년 평균 가격보다 10배, 올해 초보다 4배 치솟았다. 스페인은 1년 새 전기요금에 3배 늘었다. 두 국가만의 문제가 아니다. 정도의 차이는 있지만 독일과 프랑스도 마찬가지다. 갑자기 유럽 지역에서 바람이 불면서 풍력발전이 타격을 받은 게 방아쇠가 됐다. 화석연료로 돌아가자는 얘기가 아니다.



1. 문제정의

분석 주제

- 건물 특성과 건물이 위치한 지역의 기후 및 날씨 변수를 설명하는 변수로 구성된 데이터 세트를 사용하여, 각 건물의 1년 에너지 소비량의 합을 예측.

분석 목표


- 에너지 소비량에 가장 많은 영향을 미치는 원인들을 파악함으로써, 빌딩 관리 측면에서 에너지 효율 개선에 도움이 될 수 있다.
- 미래 에너지 소비량 예측을 통해서 효과적인 전력 사용량 관리에 도움을 줄 것을 기대할 수 있다.
- 최근 기업들은 ESG경영에 투자하고, 국가 차원의 에너지 감축을 권장하고 있는 상황이므로, 나아가 에너지 소비량 감소 정책 입안에도 도움을 줄 수 있다.



2. 데이터 설명

Data

Kaggle WiDS Datathon



Community Prediction Competition

WiDS Datathon 2022

Join the Women in Data Science (WiDS) Datathon

829 teams · 4 months ago

Overview Data Code Discussion Leaderboard Rules Team My Submissions **Late Submission** ...

Overview

Description	WiDS Datathon 2022
Evaluation	In advance of the Women in Data Science (WiDS) Worldwide Conference to be held on March 7, 2022, we invite you to build a team, hone your data science skills, and join us for the 5th Annual WiDS Datathon focused on social impact.
FAQ	
Datathon Phase II: Excellence In Research Award	This year's WiDS Datathon, organized by the WiDS Worldwide team, Stanford University , Harvard University , IACS , and the WiDS Datathon Committee , will address an important way to mitigate the effects of climate change with a focus on energy efficiency. The WiDS Datathon Committee is partnering with experts from many disciplines at Climate Change AI (CCAI) , Lawrence Berkeley National Laboratory (Berkeley Lab) , US Environmental Protection Agency (EPA) , and MIT Critical Data . This year's datathon is open until February 26, 2022. Winners will be announced at the WiDS Conference via livestream, reaching a community of 100,000+ data enthusiasts across more than 85 countries.
Datathon Timeline	
Tutorials And Resources	

- 데이터 출처

Kaggle <WiDS(Women in Data Science) Datathon>

- Climate Change AI (CCAI)와 Lawrence Berkeley National Laboratory (Berkeley Lab)가 협력해서 구축한 데이터셋

<https://www.kaggle.com/c/widsdatathon2022>

2. 데이터 설명

Data

- Data shape = (75742, 64)
- 건물 관련 정보 변수 8개 / 월별 기온 변수 36개 / 날씨 변수 16개

관 측 연 도 코 드	주_코드	주거상업여부	건물_유형	바닥_면적	건축_연도	에너 지_등 급	고 도	최 저 기 온_1 월	평균 기 온_1 월	최 고 기 온_1 월	최 저 기 온_2 월	평균_기온_2월	최 고 기 온_2 월	최 저 기 온_3 월	평균_기온_3월	최 고 기 온_3 월	최 저 기 온_4 월
1	State_1	Commercial	Grocery_store_or_food_market	61242.0	1942.0	11.0	2.4	36	50.5	68	35	50.589286	73	40	53.693548	80	41
1	State_1	Commercial	Warehouse_Distribution_or_Shipping_center	274000.0	1955.0	45.0	1.8	36	50.5	68	35	50.589286	73	40	53.693548	80	41
1	State_1	Commercial	Retail_Enclosed_mall	280025.0	1951.0	97.0	1.8	36	50.5	68	35	50.589286	73	40	53.693548	80	41
1	State_1	Commercial	Education_Other_classroom	55325.0	1980.0	46.0	1.8	36	50.5	68	35	50.589286	73	40	53.693548	80	41
1	State_1	Commercial	Warehouse_Nonrefrigerated	66000.0	1985.0	100.0	2.4	36	50.5	68	35	50.589286	73	40	53.693548	80	41

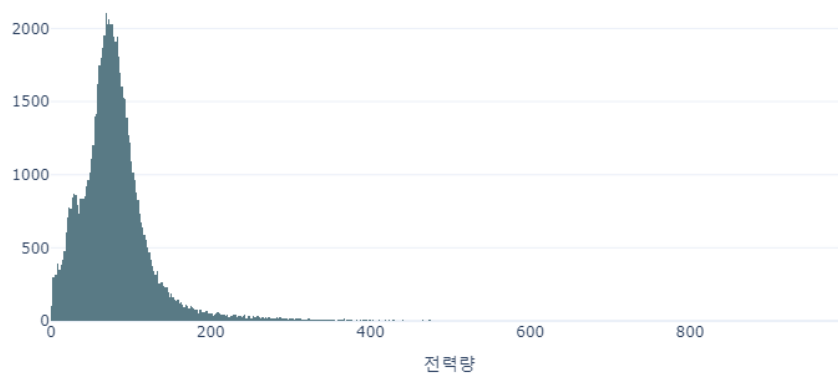
2. 데이터 설명

Target

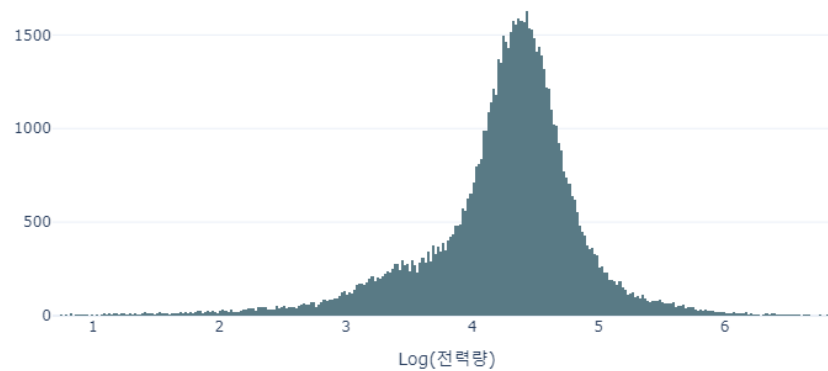
Site EUI(Energy usage Intensity)

: 연간 평방 피트당 에너지양. 1년 동안 건물에 의해 소비된 총 에너지(kBtu 또는 GJ로 측정)를 건물의 총 바닥 면적(평방 피트 또는 평방 미터로 측정)으로 나누어 계산함.

전력량(Target) 분포

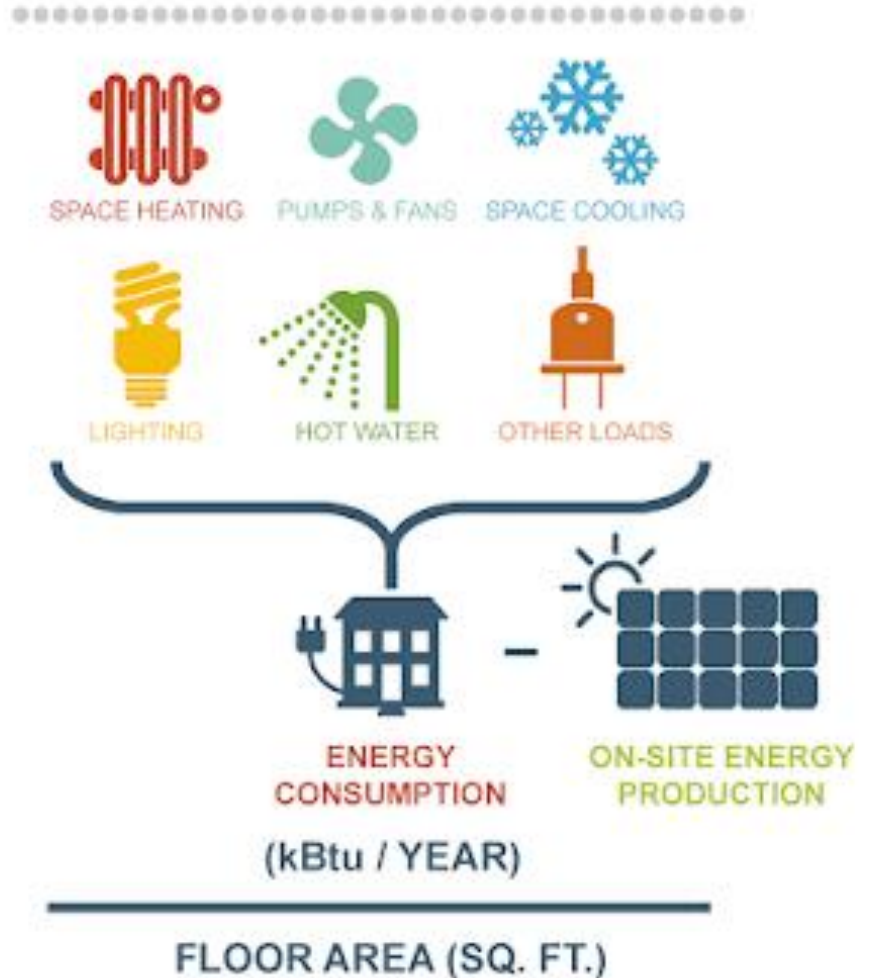


log(전력량) 분포



Site Energy Use Intensity (EUI)

MEASURED FOR 365 DAYS



2. 데이터 설명

Features

구분	변수 개수	변수
건물 정보	8개	<ul style="list-style-type: none">• 관측년도 (에너지 발생량을 관찰한 연도)• 건물 위치(주, State)• 건물 주거 상업 여부(주거용 / 상업용 / 둘 다)• 건물 용도• 건물 면적• 건물 건축 연도• 건물 에너지 등급• 건물 위치의 고도
월별 기온 정보	36개	<ul style="list-style-type: none">• 건물 위치의 최저 기온 (1월 ~ 12월 / 화씨)• 건물 위치의 최고 기온(1월 ~ 12월 / 화씨)• 건물 위치의 평균 기온 (1월 ~ 12월 / 화씨)
날씨 정보	16개	<ul style="list-style-type: none">• 냉각 총 화씨도수, 난방 총 화씨도수• 연간 강수량, 강설량, 적설량• 1년간 평균 기온• 섭씨 -1,-6,-12,-17도 이하의 총 일수• 섭씨 26, 32, 37, 43도 초과의 총 일수• 최대 풍속에 대한 풍향• 최대 풍속 돌풍에 대한 풍향• 최대 풍속• 안개 낀 일 수
기타	1	<ul style="list-style-type: none">• 건물 ID

3. EDA 및 변수 설명

건물 정보 변수

관측 연도 코드

건물 위치 주 코드

건물 주거상업여부

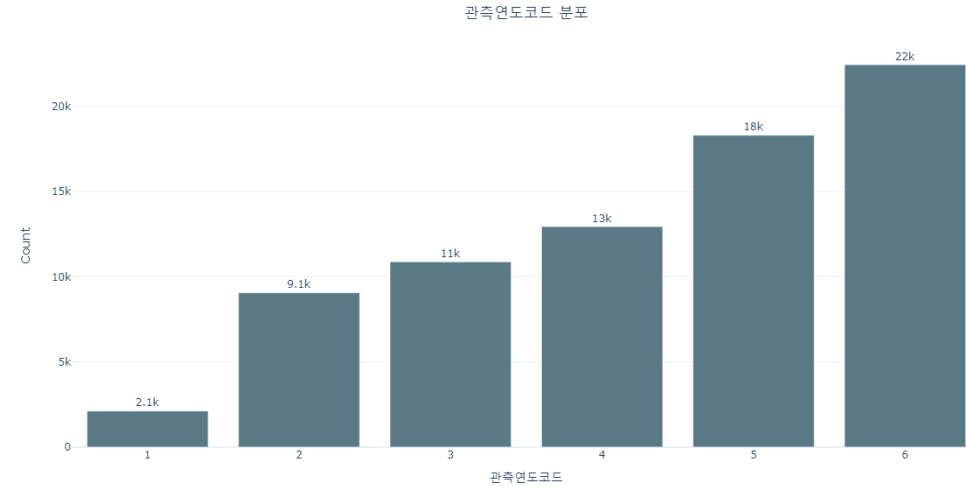
건물 유형

건물 면적

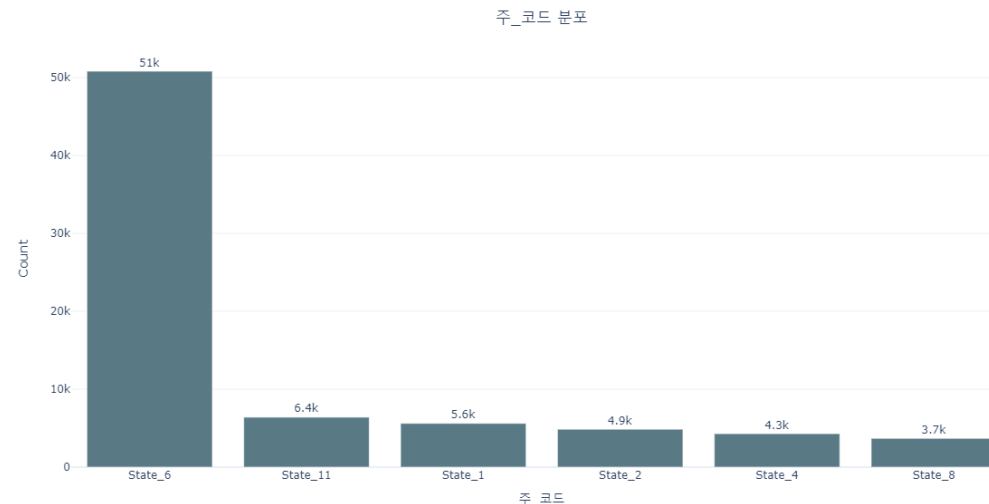
건물 건축 연도

에너지 등급

위치의 고도



건물의 전력 소비량을 관측한 연도를 코드화 한 정보.



건물이 위치한 주(State)를 코드화 한 정보

3. EDA 및 변수 설명

건물 정보 변수

관측 연도 코드

건물 위치 주 코드

건물 주거상업여부

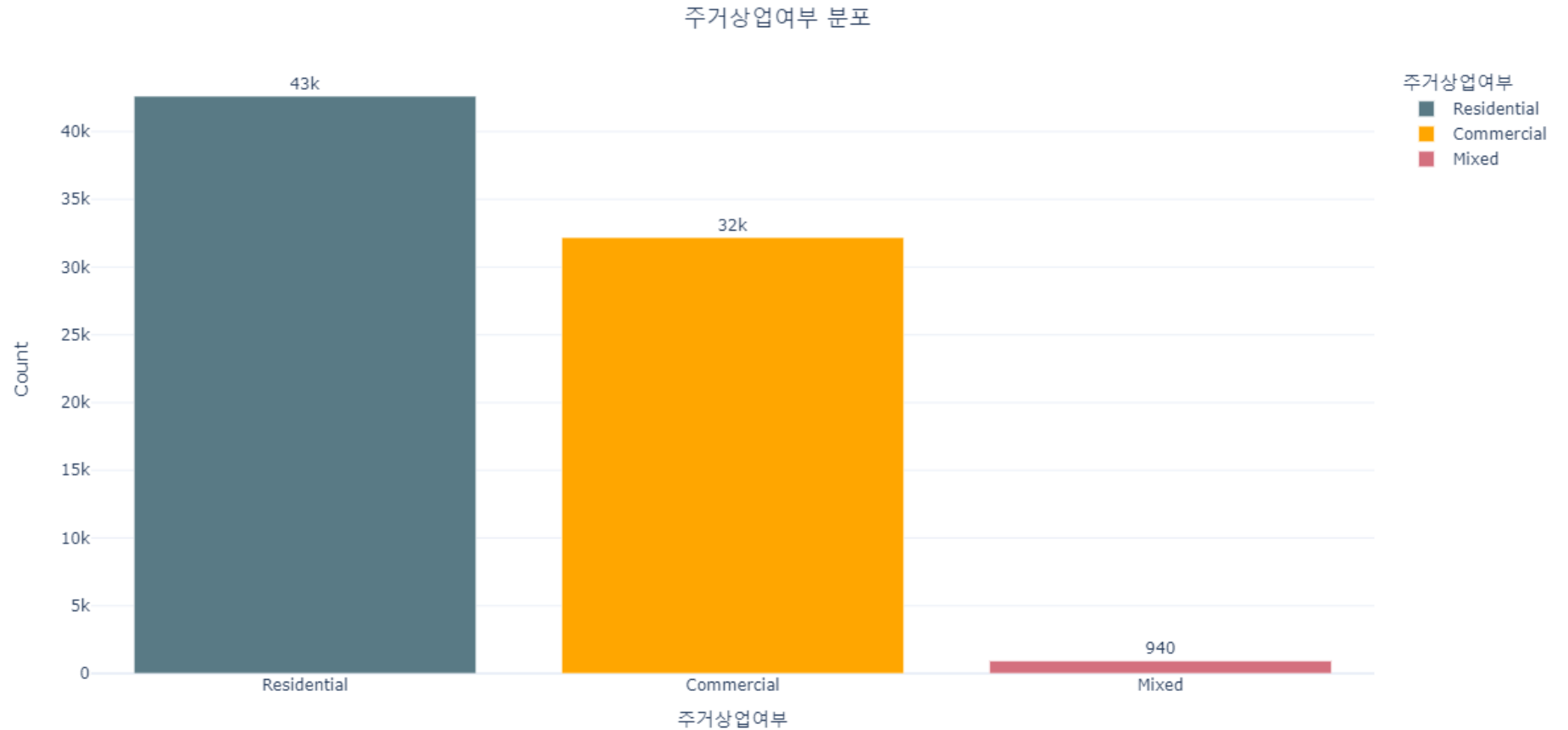
건물 유형

건물 면적

건물 건축 연도

에너지 등급

위치의 고도



건물이 주거용인지, 상업용인지 혹은 둘 다의 기능을 가지고 있는지를 보여주는 변수

3. EDA 및 변수 설명

건물 정보 변수

관측 연도 코드

건물 위치 주 코드

건물 주거상업여부

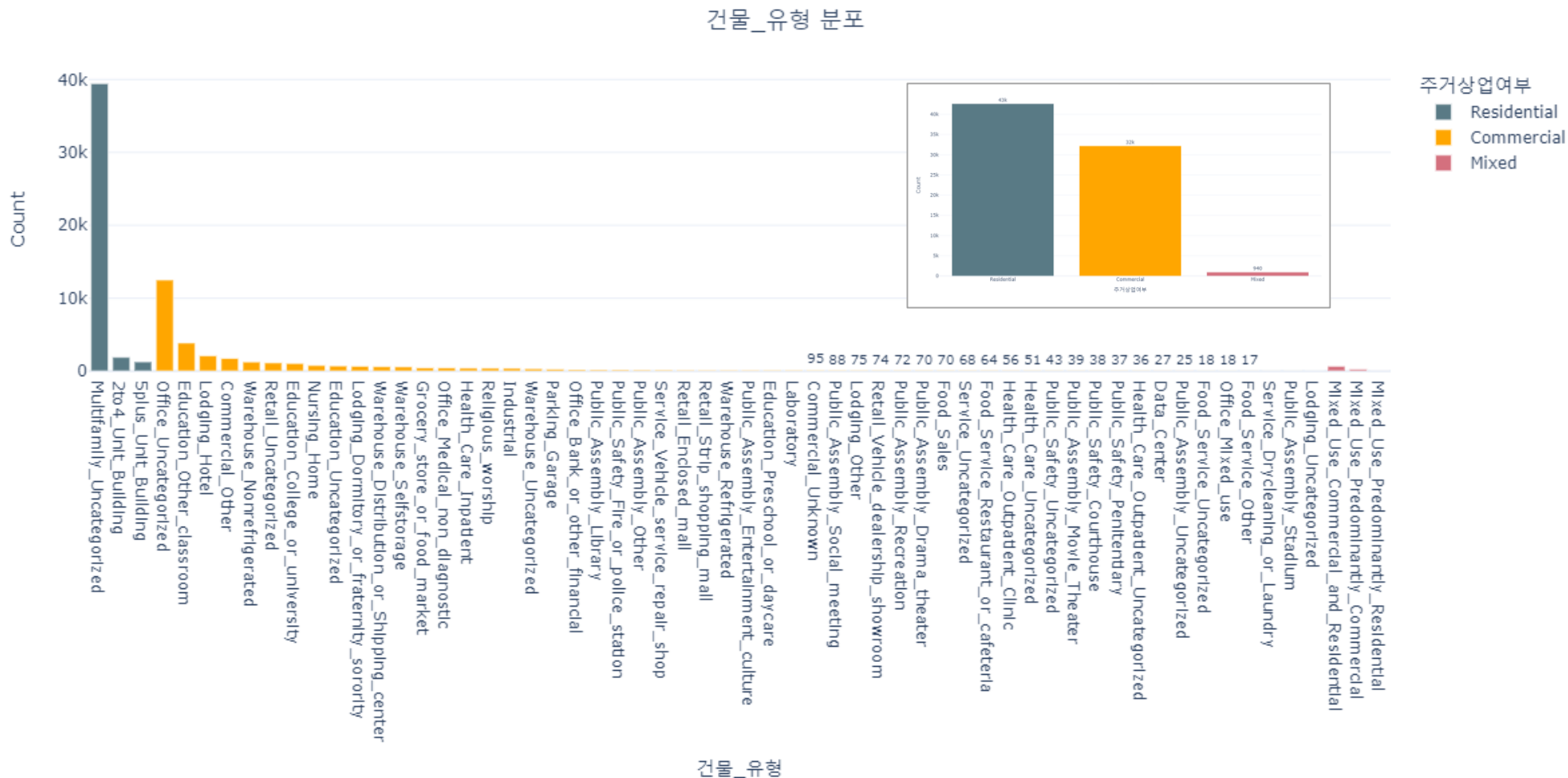
건물 유형

건물 면적

건물 건축 연도

에너지 등급

위치의 고도



총 60개의 건물 유형으로 건물의 용도를 상세하게 세분화해둔 변수

총 60개의 건물 유형으로 건물의 용도를 상세하게 세분화해둔 변수

3. EDA 및 변수 설명

건물 정보 변수

관측 연도 코드

건물 위치 주 코드

건물 주거상업여부

건물 유형

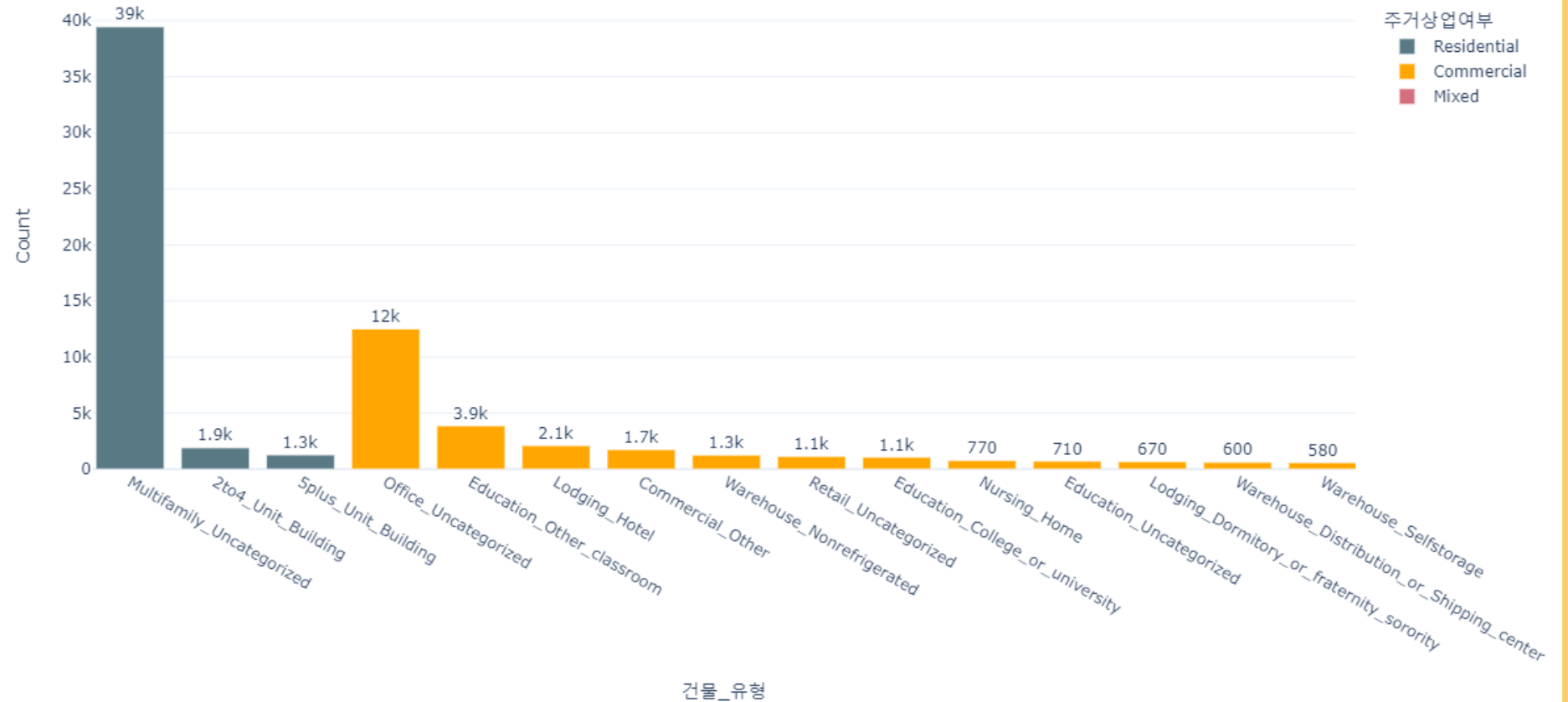
건물 면적

건물 건축 연도

에너지 등급

위치의 고도

건물 유형 분포 : TOP 15개



Top3

1. Multifamily Uncategorized : 주거용 빌라 (미분류)
2. Office Uncategorized : 사무실 (미분류)
3. Education Other Classroom : 교육 기관(학교 교실 등)

3. EDA 및 변수 설명

건물 정보 변수

관측 연도 코드

건물 위치 주 코드

건물 주거상업여부

건물 유형

건물 면적

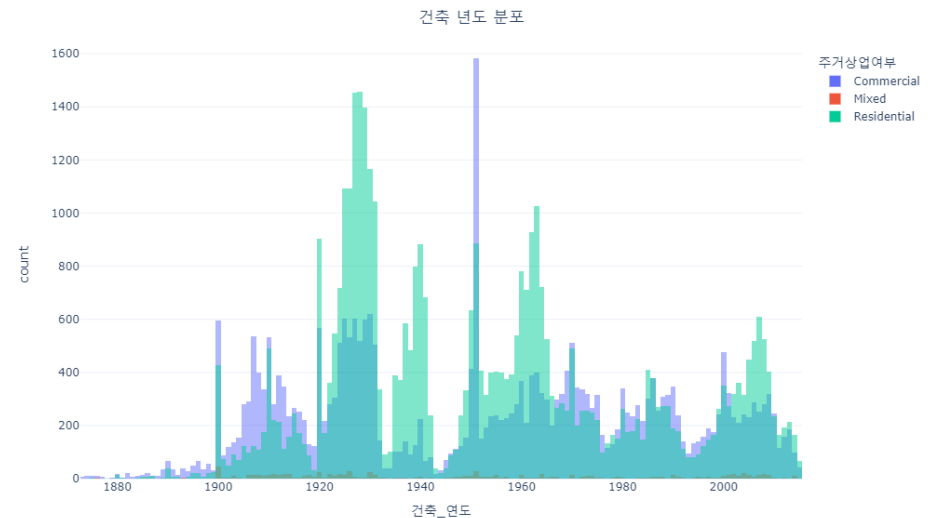
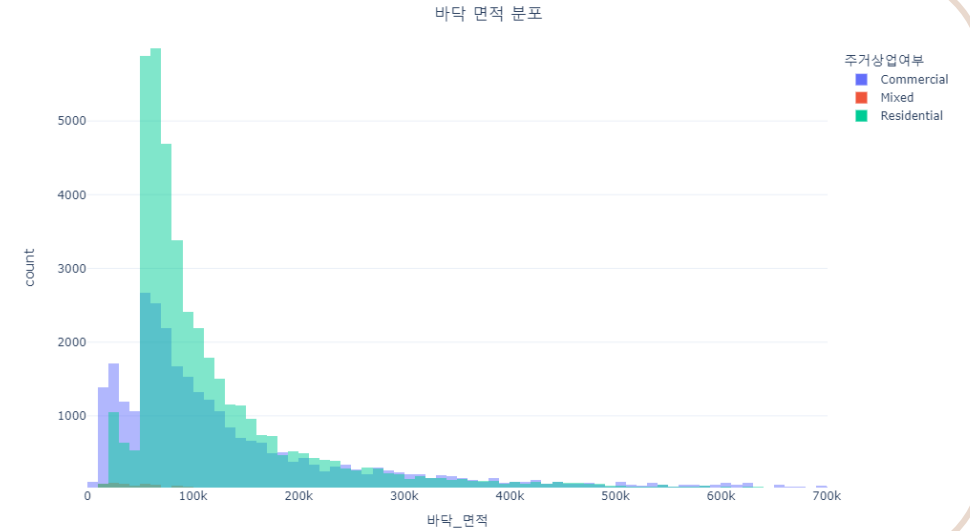
건물 건축 연도

에너지 등급

위치의 고도

- mean 165959.16011
- std 246889.21715
- min 943
- 25% 62371.25
- 50% 91360
- 75% 165938.25
- max 6385382

- mean 1952.306764
- std 37.053619
- min 0
- 25% 1927
- 50% 1951
- 75% 1977
- max 2015



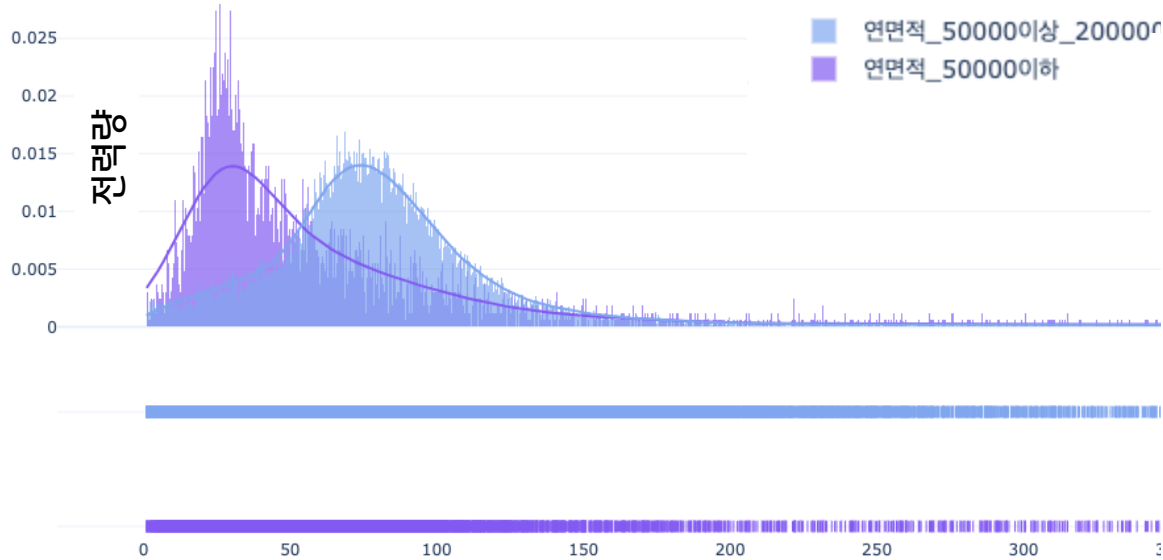
3. EDA 및 변수 설명

건물 정보 변수

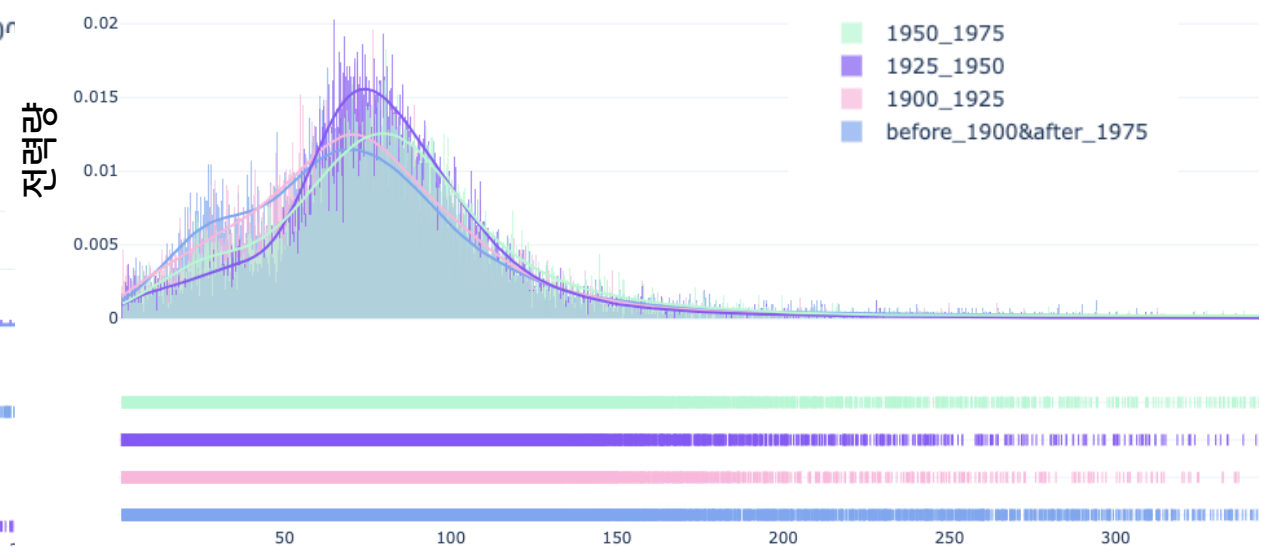
건축년도 그룹화

바닥면적 그룹화

연면적 그룹별 전력량 분포



건축년도 그룹별 전력량 분포



변수 그룹화 이후 전력량 분포 그래프 (좌측-바닥 연면적, 우측-건축년도)

3. EDA 및 변수 설명

건물 정보 변수

관측 연도 코드

건물 위치 주 코드

건물 주거상업여부

건물 유형

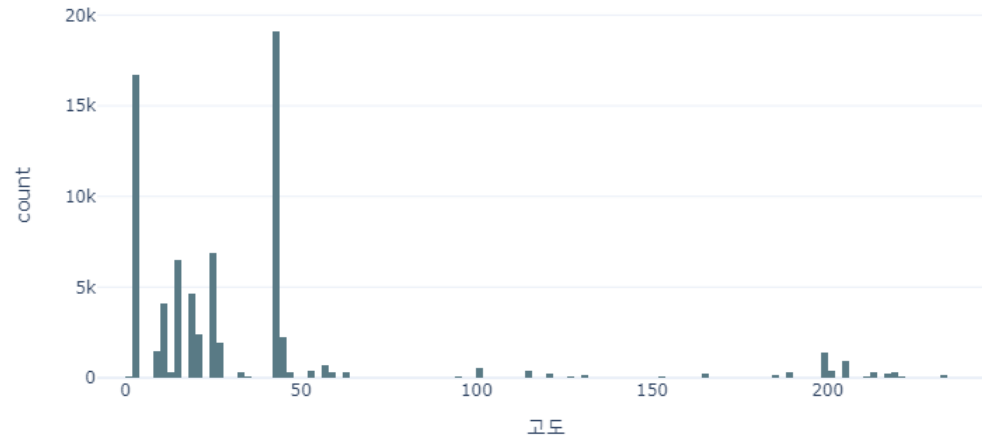
건물 면적

건물 건축 연도

위치의 고도

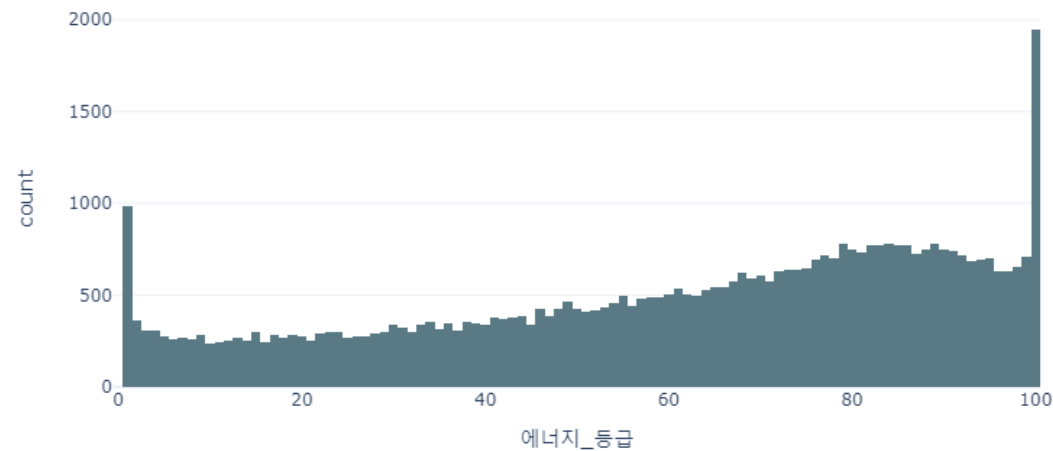
에너지 등급

고도 분포



- mean 39.49185
- std 60.59996
- min -6.4
- 25% 11.9
- 50% 25
- 75% 42.7
- max 1924.5

에너지_등급 분포



결측값 = 26709개

3. EDA 및 변수 설명

건물 정보 변수

관측 연도 코드

건물 위치 주 코드

건물 주거상업여부

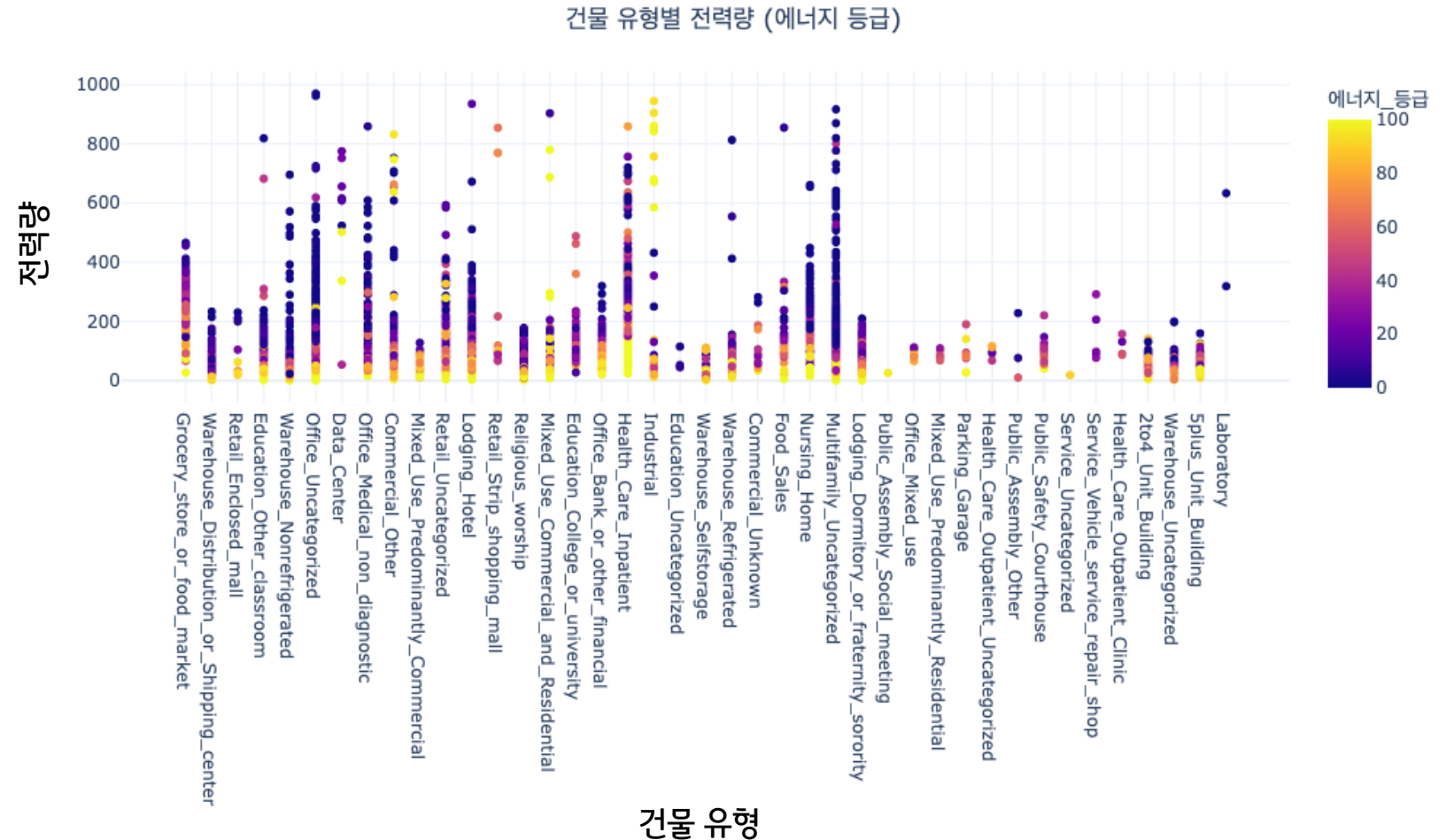
건물 유형

건물 면적

건물 건축 연도

에너지 등급

위치의 고도



에너지 등급 산정 시, 건물의 물리적 특성 (위치 등) / 거주자, 세입자 구성원 / 건물 내에서 사용 되는 디지털 기기 / 건물의 주된 목적 등을 고려해서 산출

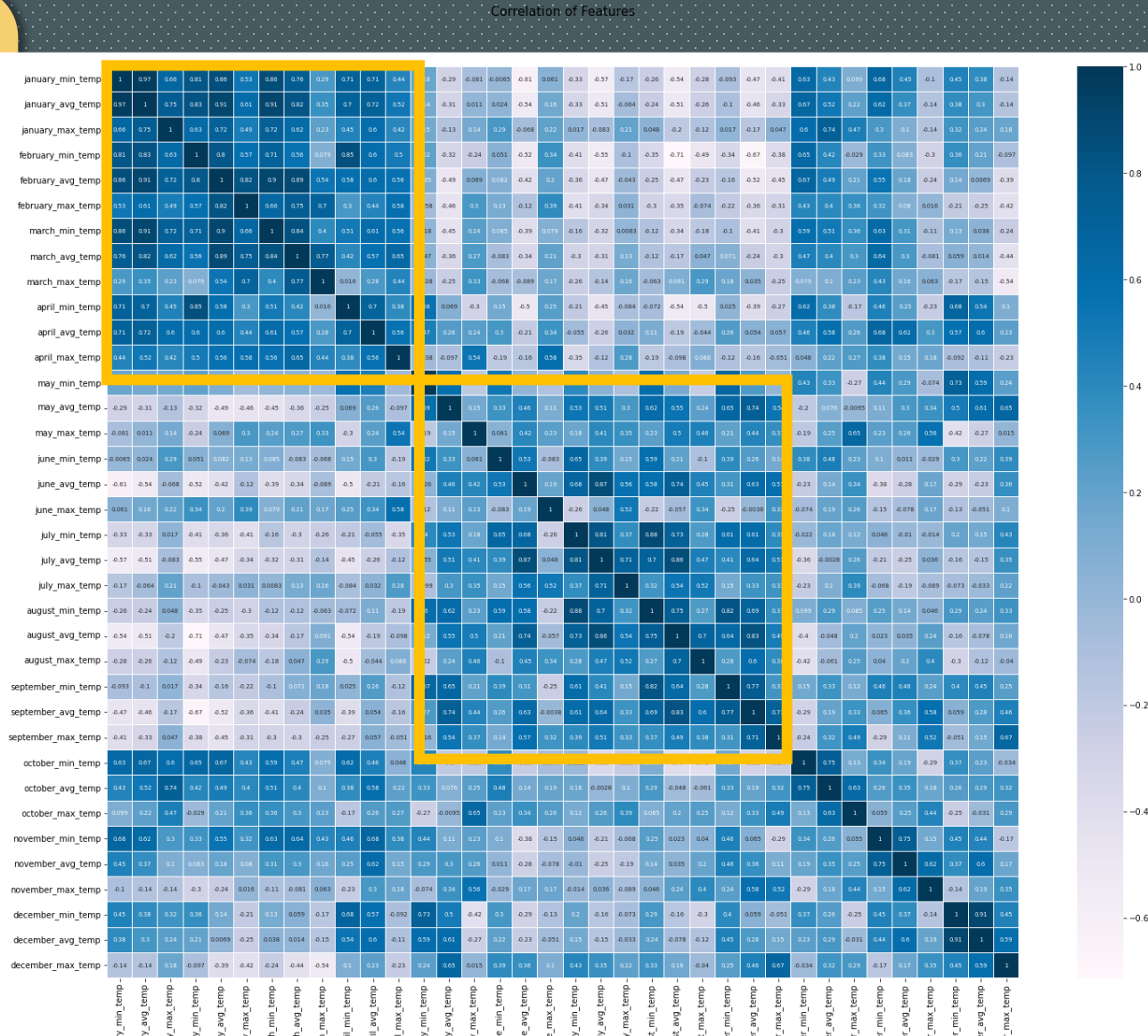
3. EDA 및 변수 설명

월별 기온 변수

1~12월 최저기온

1~12월 최고기온

1~12월 평균기온

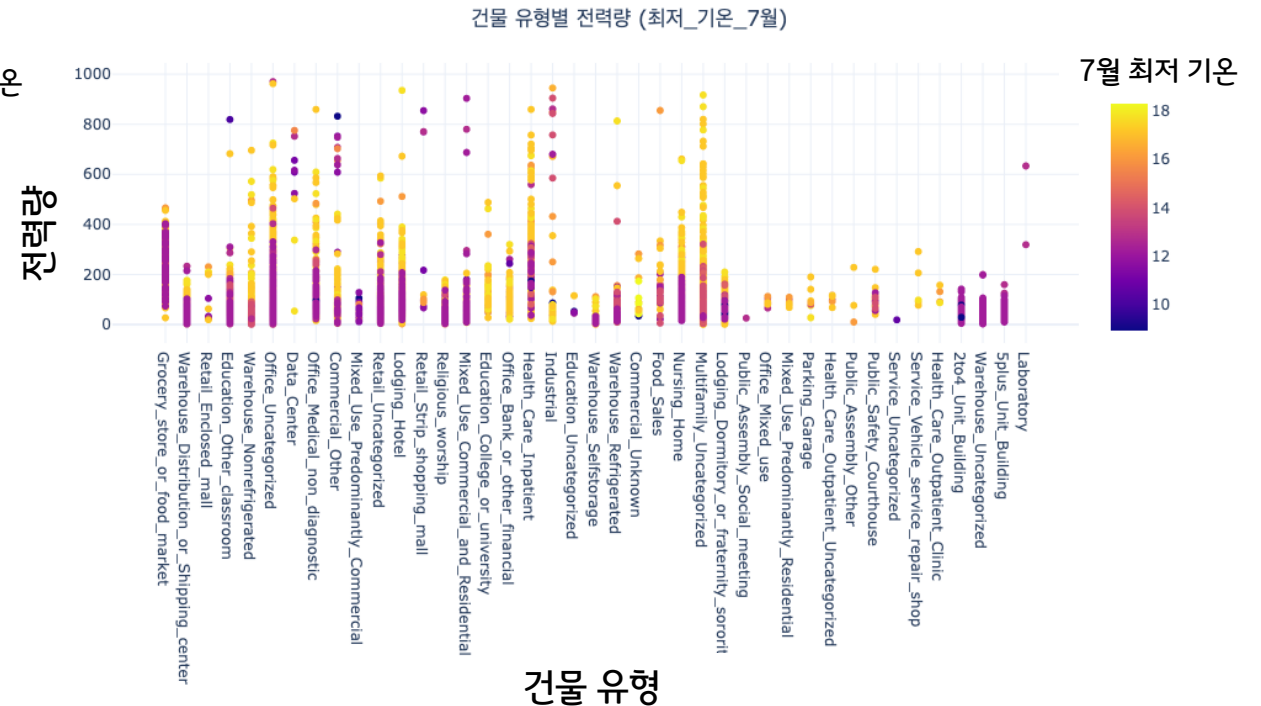
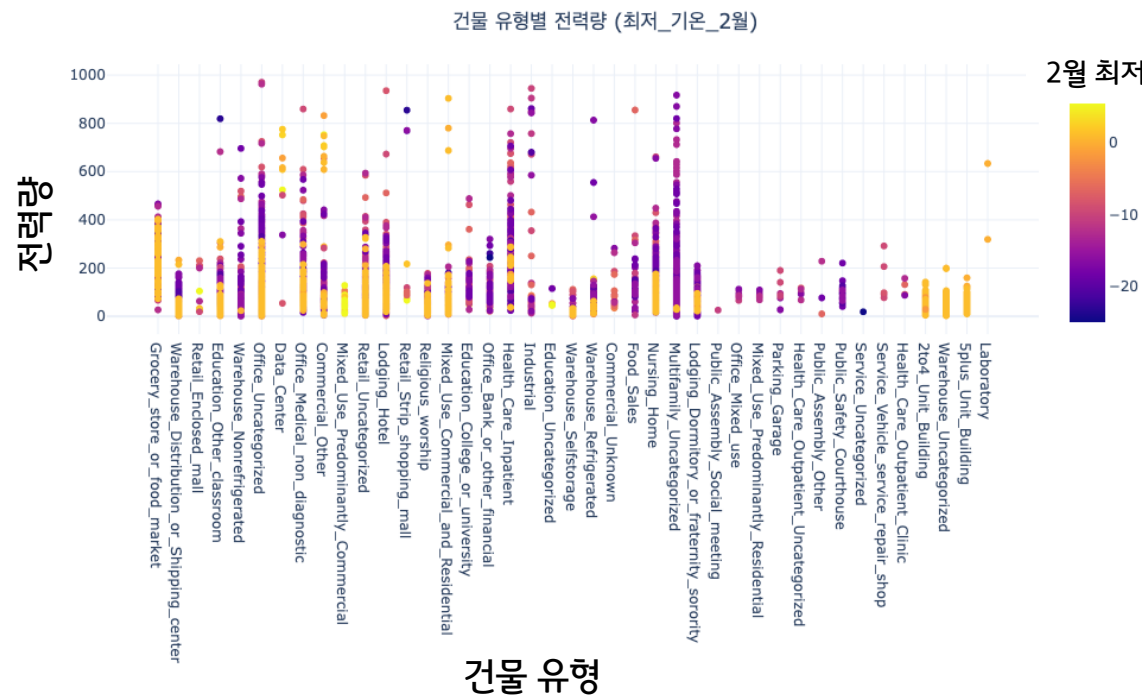


3. EDA 및 변수 설명

월별 기온 변수

2월 최저기온

7월 최저기온



여름/겨울의 최저 기온과 건물 유형별 전력 소비량간의 상관관계를 확인

3. EDA 및 변수 설명

날씨 정보 변수

난방 총 화씨도수

냉각 총 화씨도수

연간 강수량

연간 강설량

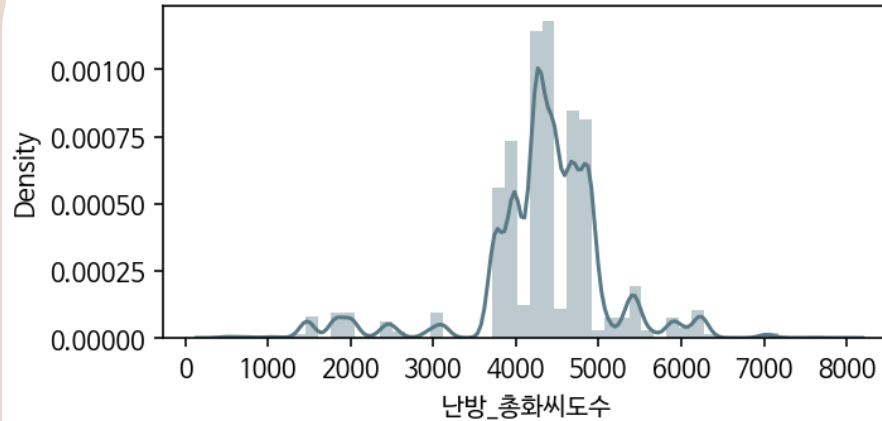
연간 적설량

연간 평균기온

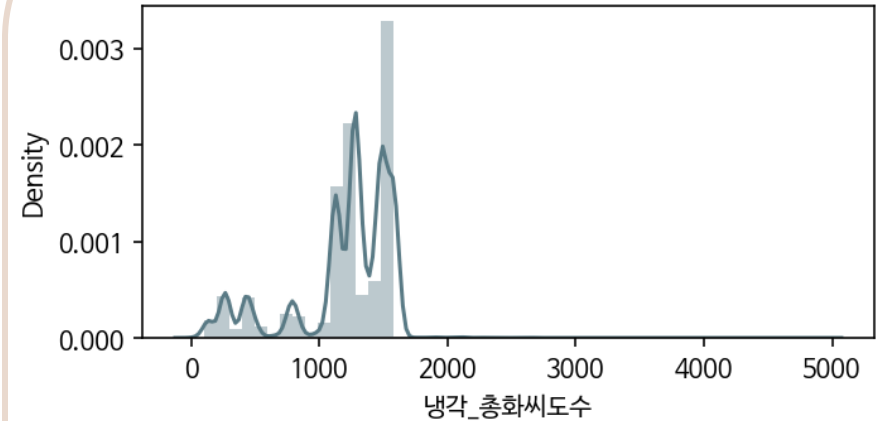
섭씨 -1, -6, -12, -17도
이하의 총 일수

섭씨 26, 32, 37, 43도
초과의 총 일수

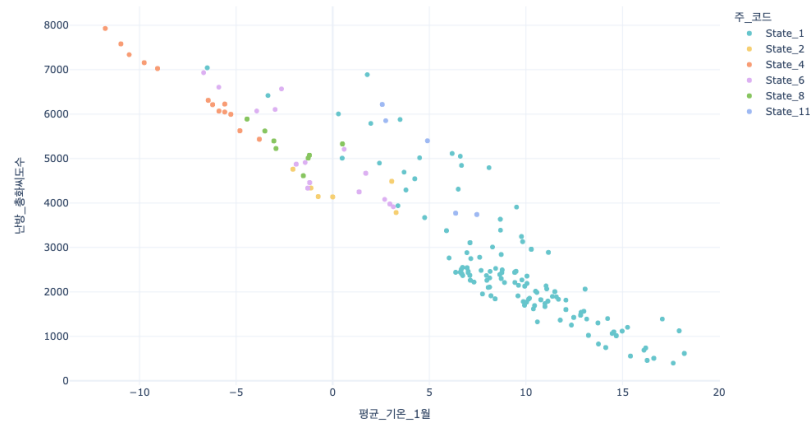
난방_총화씨도수 분포



냉각_총화씨도수 분포



난방이 필요한 화씨도수 총합 (평균_기온_1월)



냉각이 필요한 화씨도수 총합 (평균_기온_8월)



3. EDA 및 변수 설명

날씨 정보 변수

난방 총 화씨도수

냉각 총 화씨도수

연간 강수량

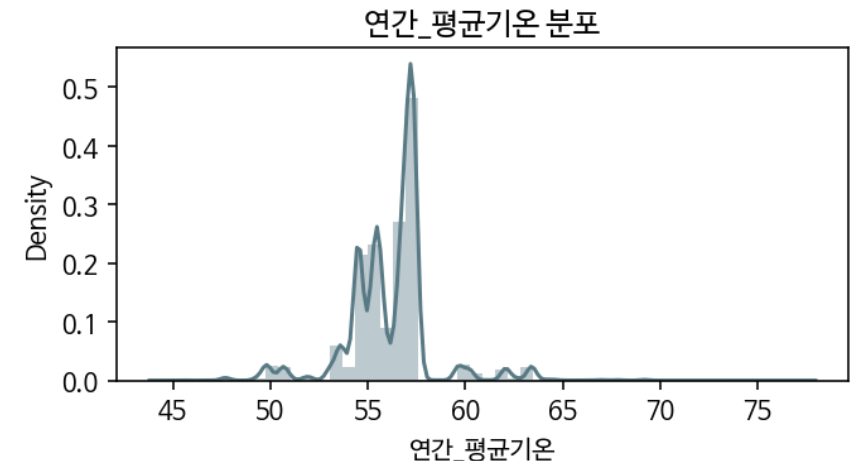
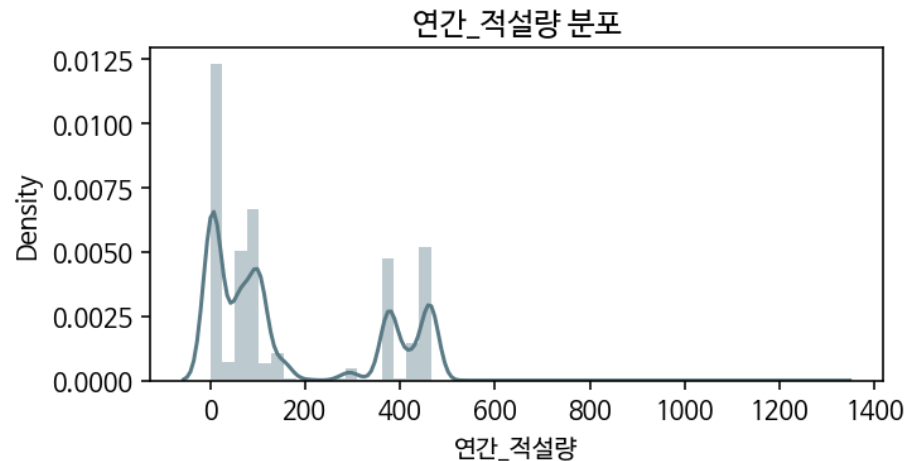
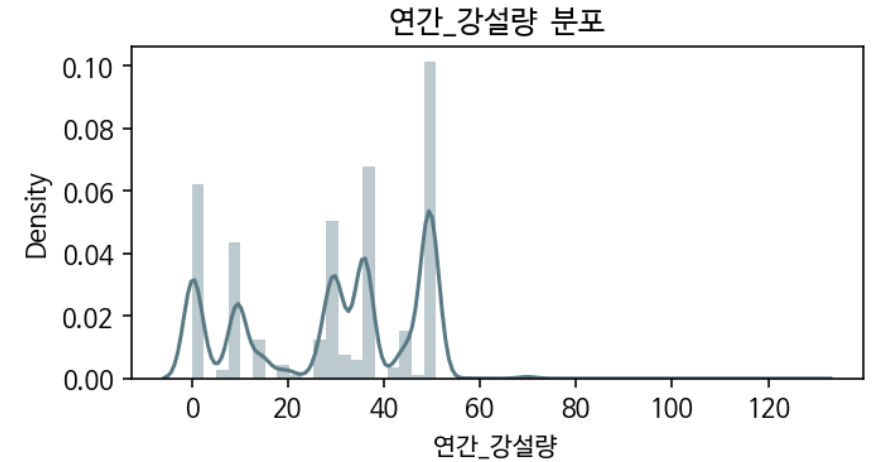
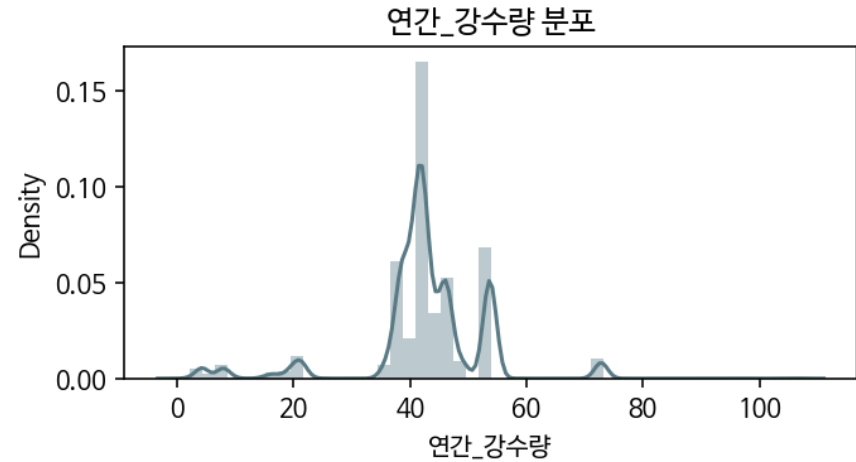
연간 강설량

연간 적설량

연간 평균기온

섭씨 -1, -6, -12, -17도
이하의 총 일수

섭씨 26, 32, 37, 43도
초과의 총 일수



3. EDA 및 변수 설명

날씨 정보 변수

난방 총 화씨도수

냉각 총 화씨도수

연간 강수량

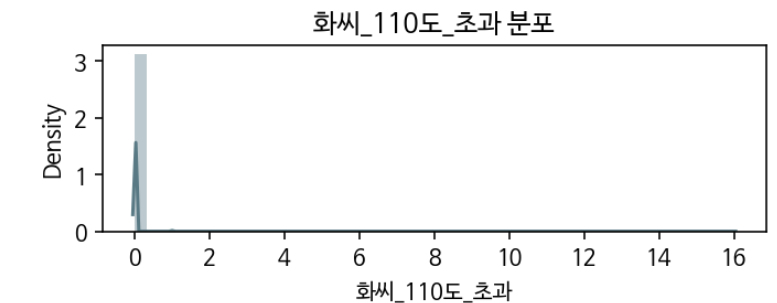
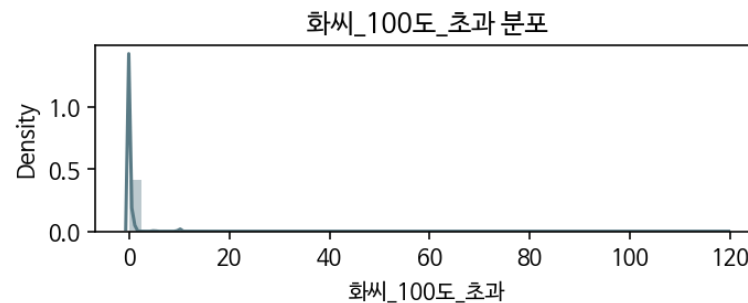
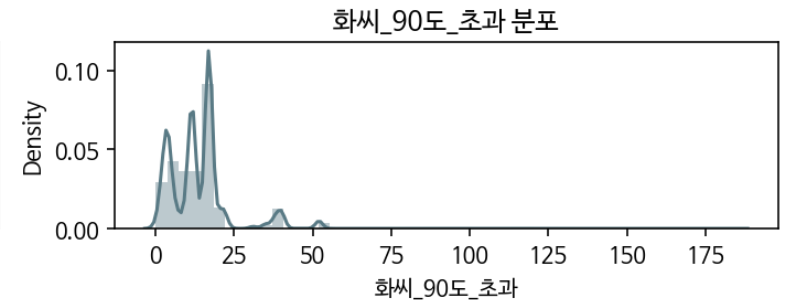
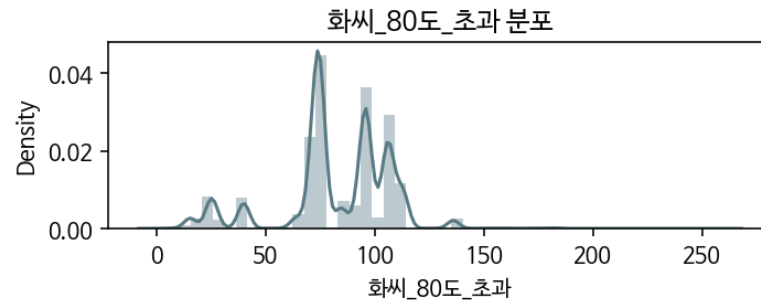
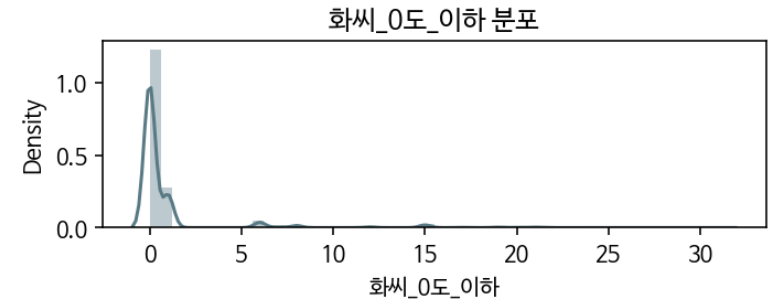
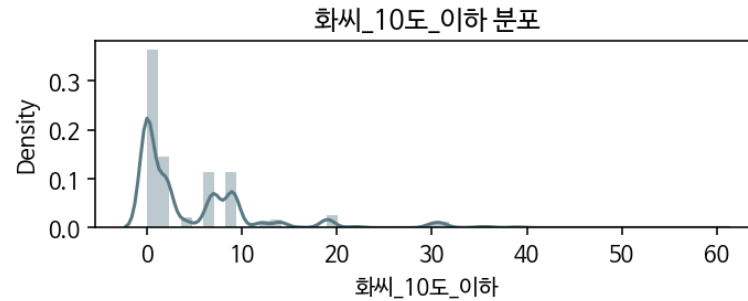
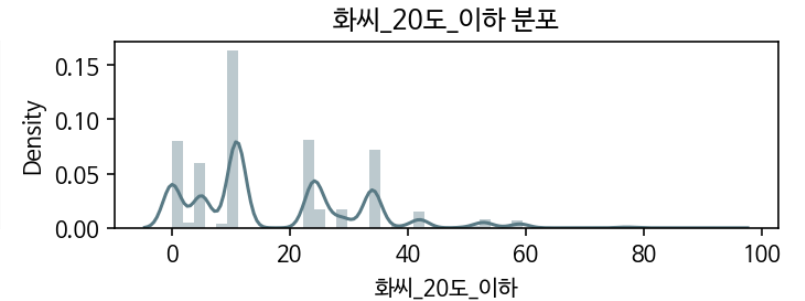
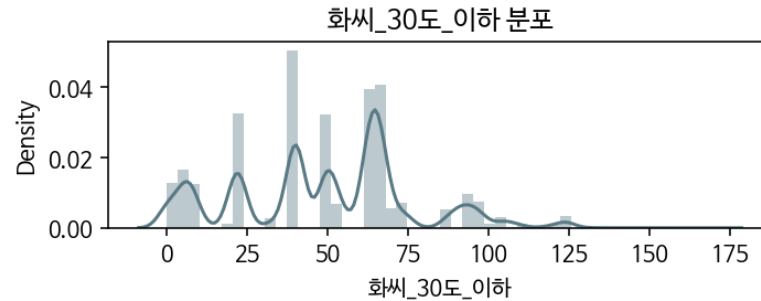
연간 강설량

연간 적설량

연간 평균기온

섭씨 -1, -6, -12, -17도
이하의 총 일수

섭씨 26, 32, 37, 43도
초과의 총 일수

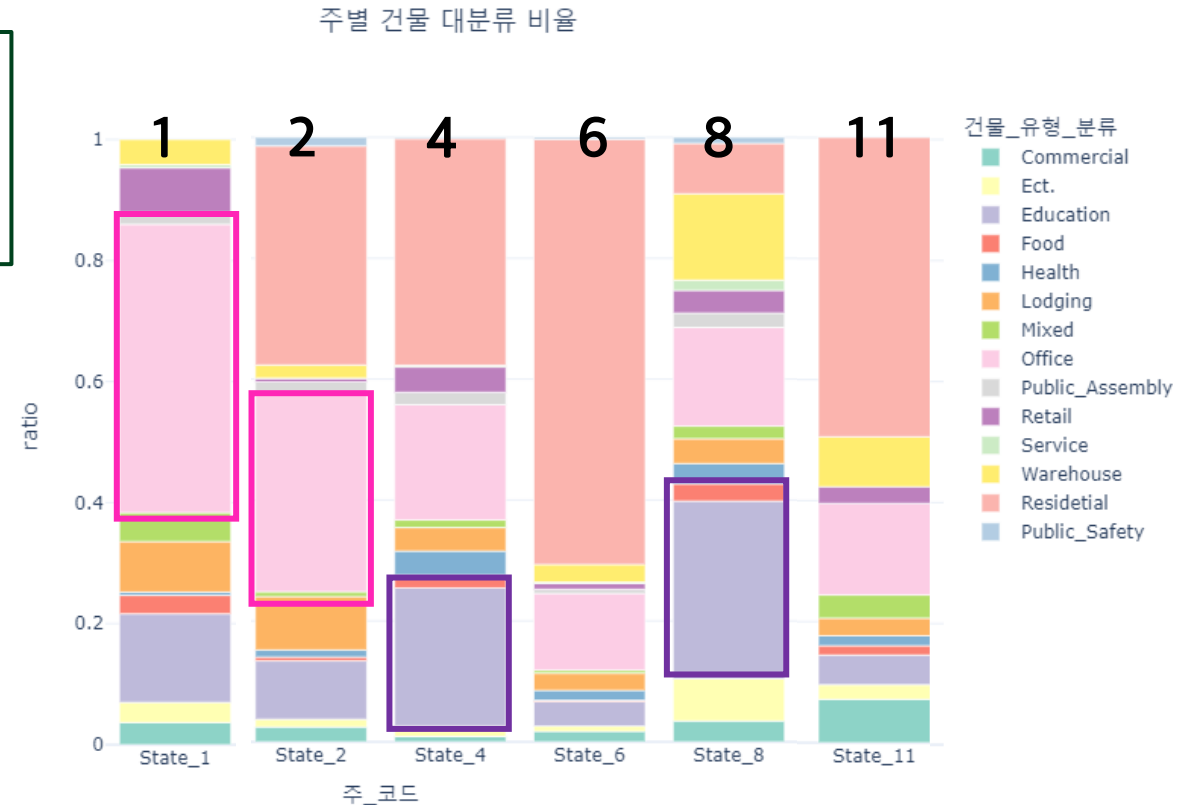
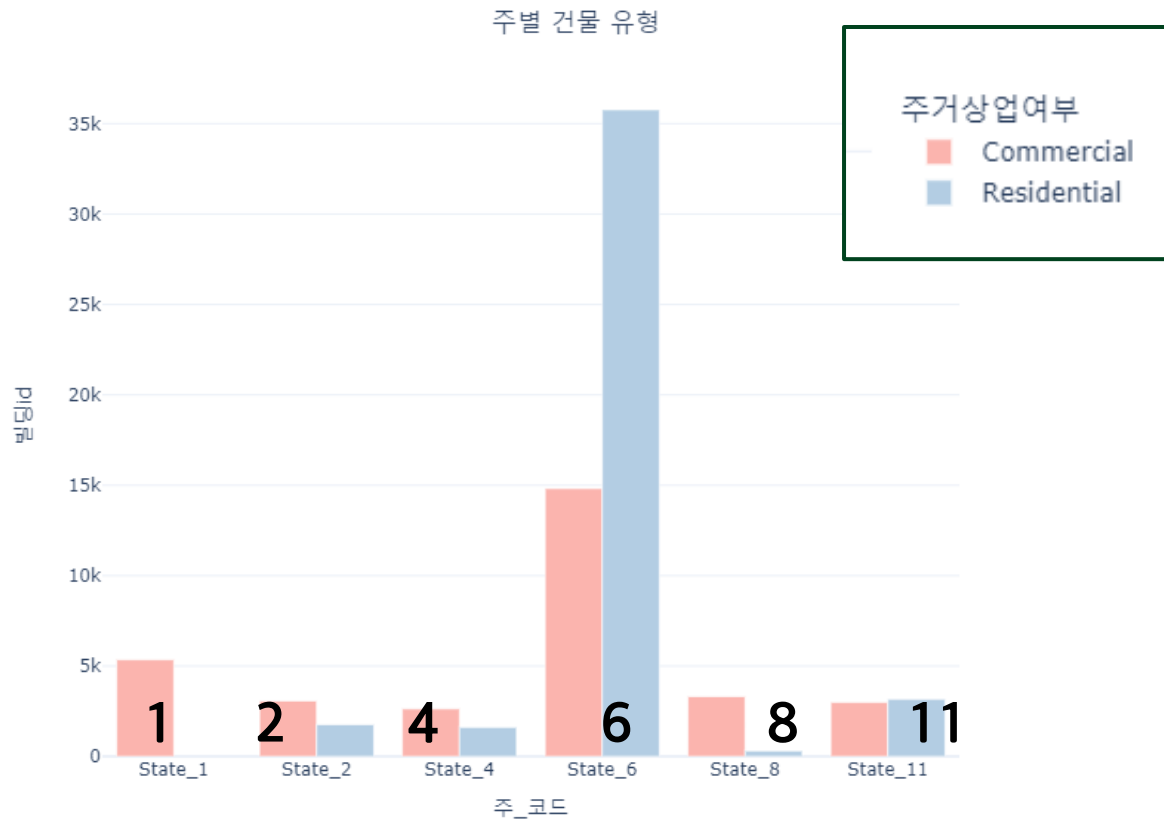




이 지역은 어디까?
마스킹 된 주 코드를 해석해보자

3. EDA 및 변수 설명

주 코드 해석해보기

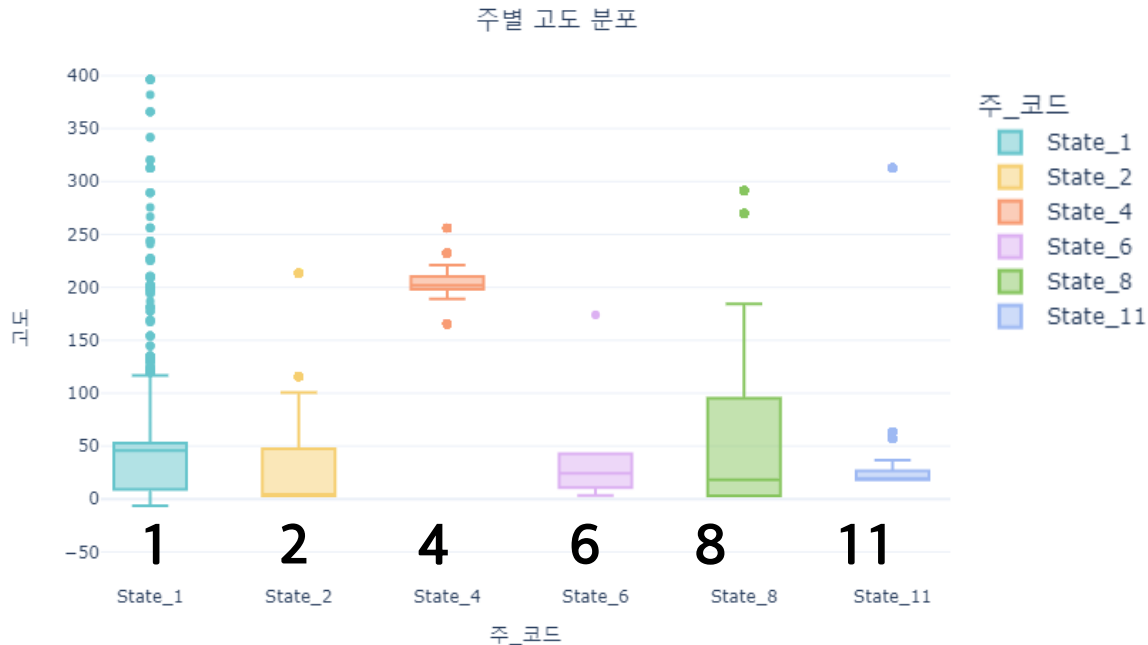


State_6 50840
State_11 6412
State_1 5618
State_2 4871
State_4 4300
State_8 3701

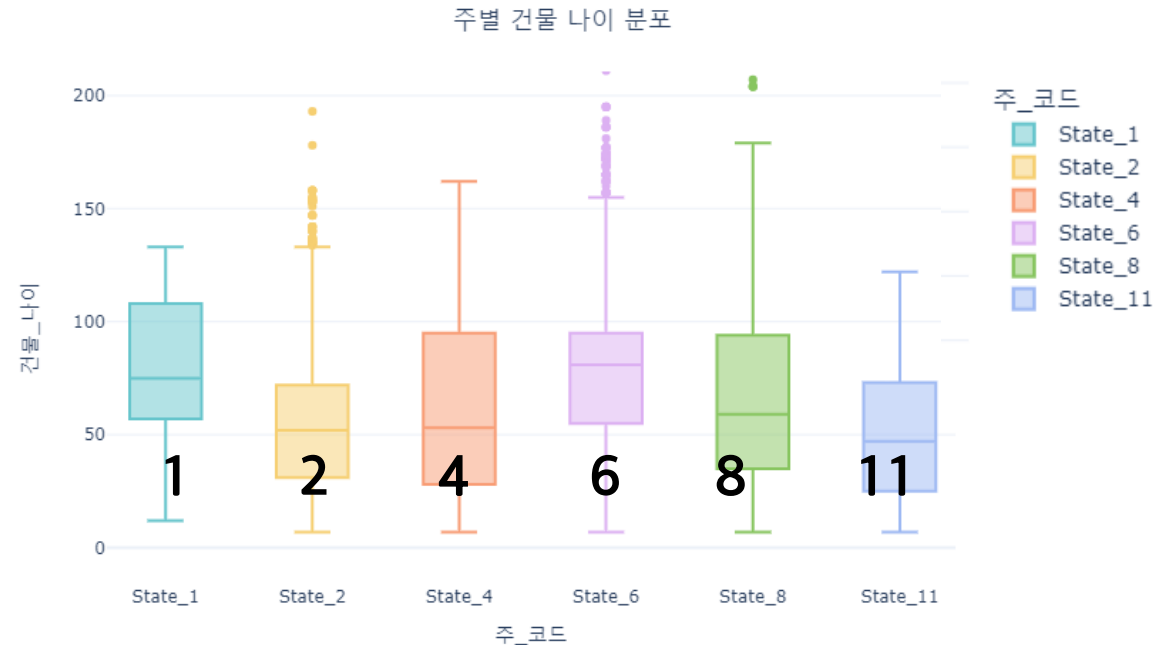
State6은 거주지역, State1과 State8은 상업지구로 볼 수 있다.

3. EDA 및 변수 설명

주 코드 해석해보기



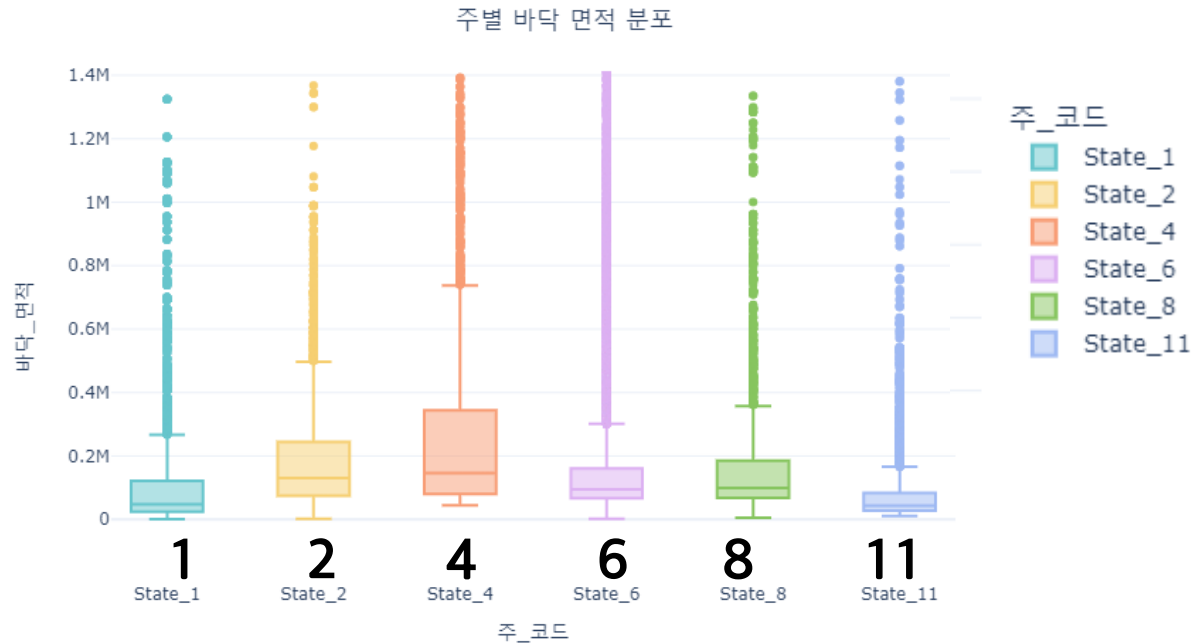
State 4 : 모든 건물이 매우 높은 고도에 위치해 있고, 분산도 작다.
State 1 : 꼬리가 매우 길어, 높은 고도에 위치한 건물도 많이 있다. 높은 산에 마을이 있을 수 있다.
State 6 : 건물 수가 매우 많았지만, 비교적 좁은 범위의 고도 안에 응집되어 있다.



State 2 : 가장 최신 건물이 많다.
State 1 : 비교적 오래된 건물이 많다.

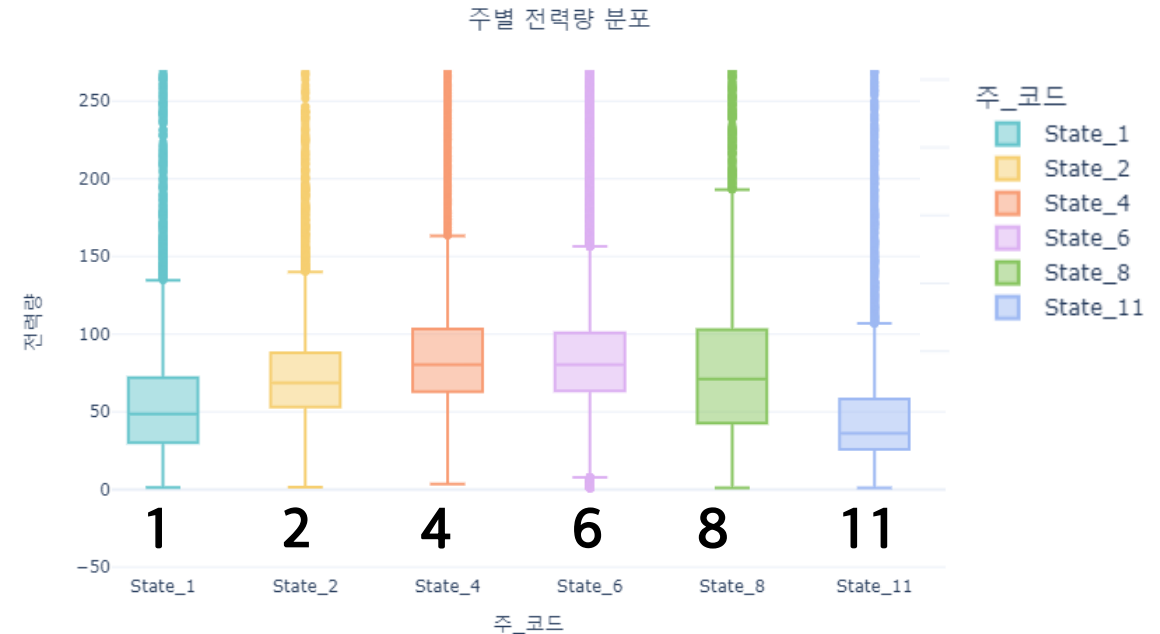
3. EDA 및 변수 설명

주 코드 해석해보기



대부분 꼬리가 긴 분포이다.

State 1, 11 : 작은 건물이 많다.

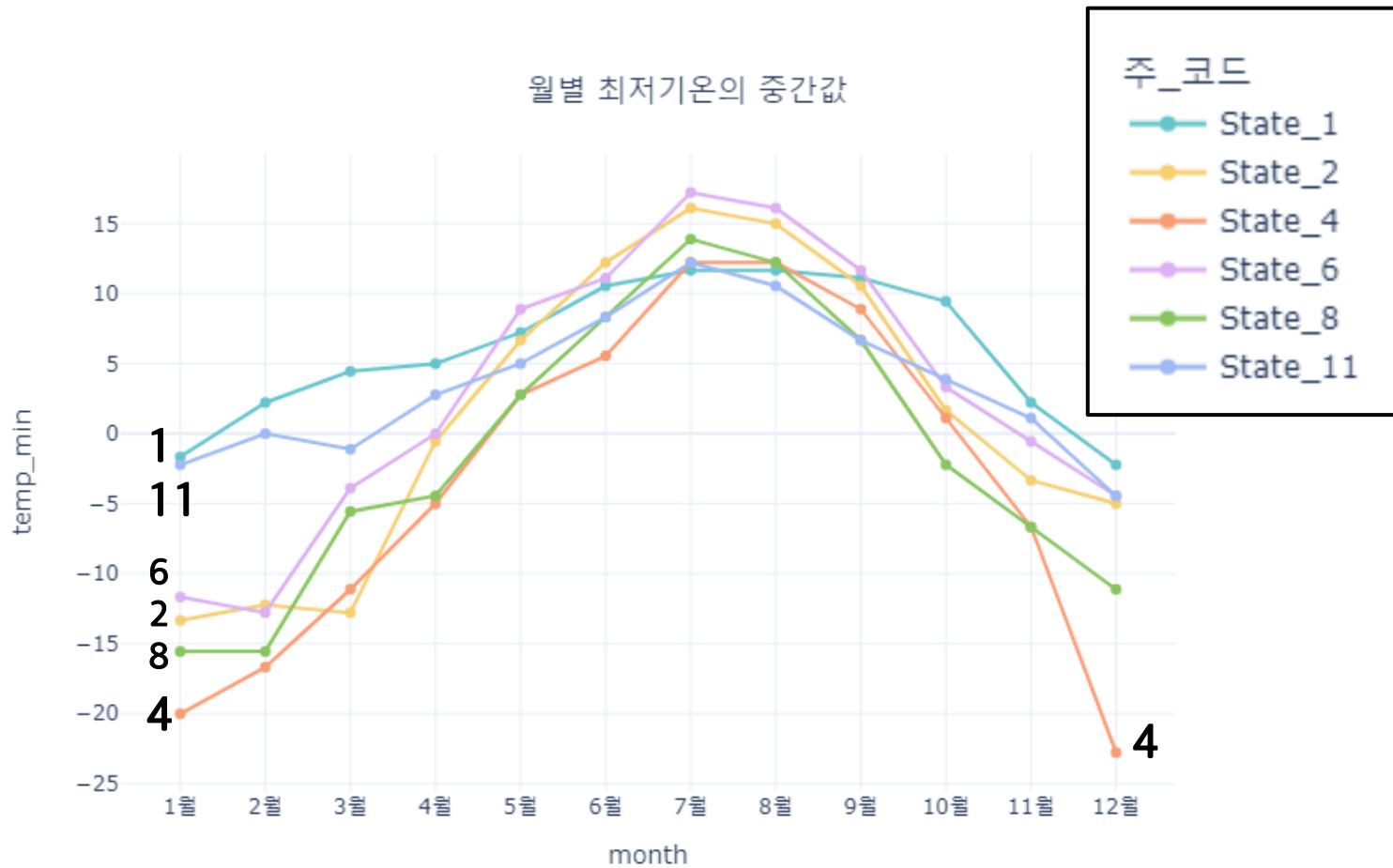


대체로 비슷하나
State 4와 6에서 가장 많이 사용한다.

State 8 : 가장 IQR이 넓다.
State 1, 11 : 가장 적게 사용한다.

3. EDA 및 변수 설명

주 코드 해석해보기



- State 1, 11 : 겨울엔 따뜻하고 여름엔 덥지 않아 연간 온도차가 크지 않은 지역
- State 4 : 가장 추운 지역
- State 6, 8 : 1~3월의 경향이 매우 비슷. 1월보다 2월에 오히려 춥거나, 기온이 올라가지 않는 지역

3. EDA 및 변수 설명

주 코드 해석해보기

	전력량	관측치 개수	주거/상업	고도	건물 나이	바닥 면적	기온
State 1	낮음		상업지구		오래된 건물	작은 건물 많음	따뜻, 연교차 작음
State 2					최신 건물		
State 4	약간 높음			매우 높은 지역		가장 다양한 분포	매우 추운 지역
State 6	약간 높음	다른 지역의 약 10배	주거지 70%				
State 8			상업지구				
State 11	낮음					작은 건물 많음	따뜻, 연교차 작음



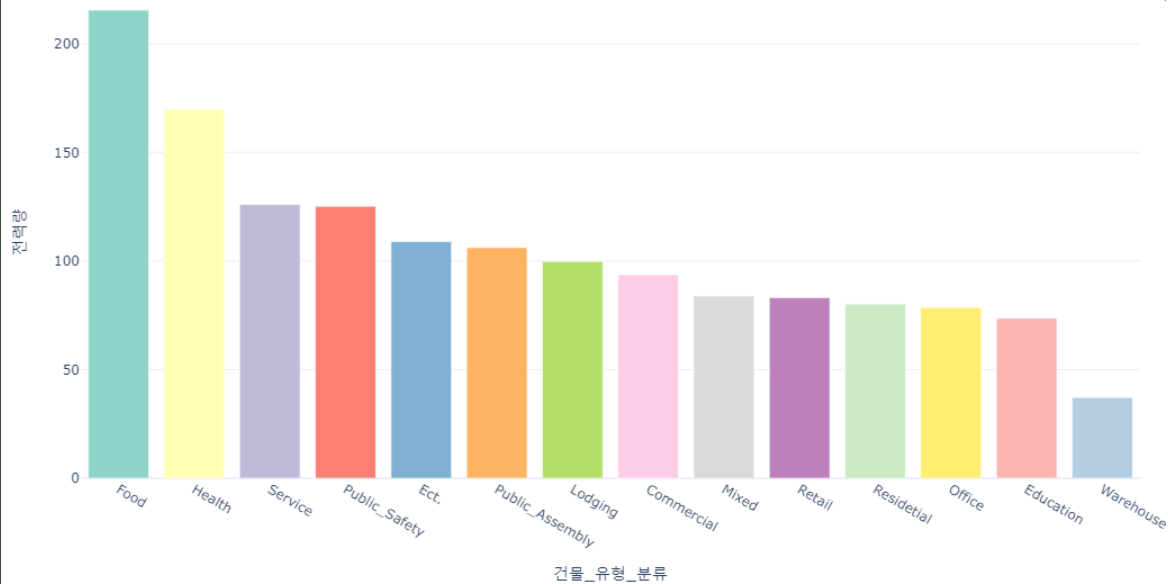
**어떤 용도의 건물이
전력을 많이 쓸까?**

3. EDA 및 변수 설명

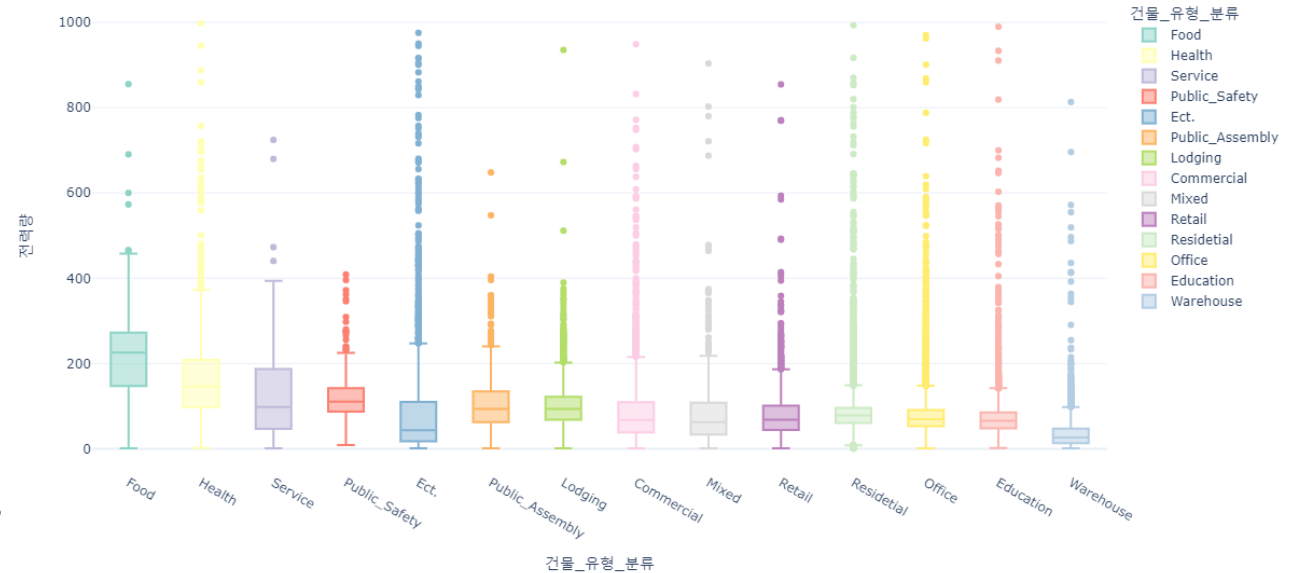
건물 유형 (대분류)

총 60개의 건물 유형 변수 \Rightarrow 14개의 대분류로 묶어, '건물 유형 분류' 변수 생성

건물 유형별 평균 전력량: 건물 대분류

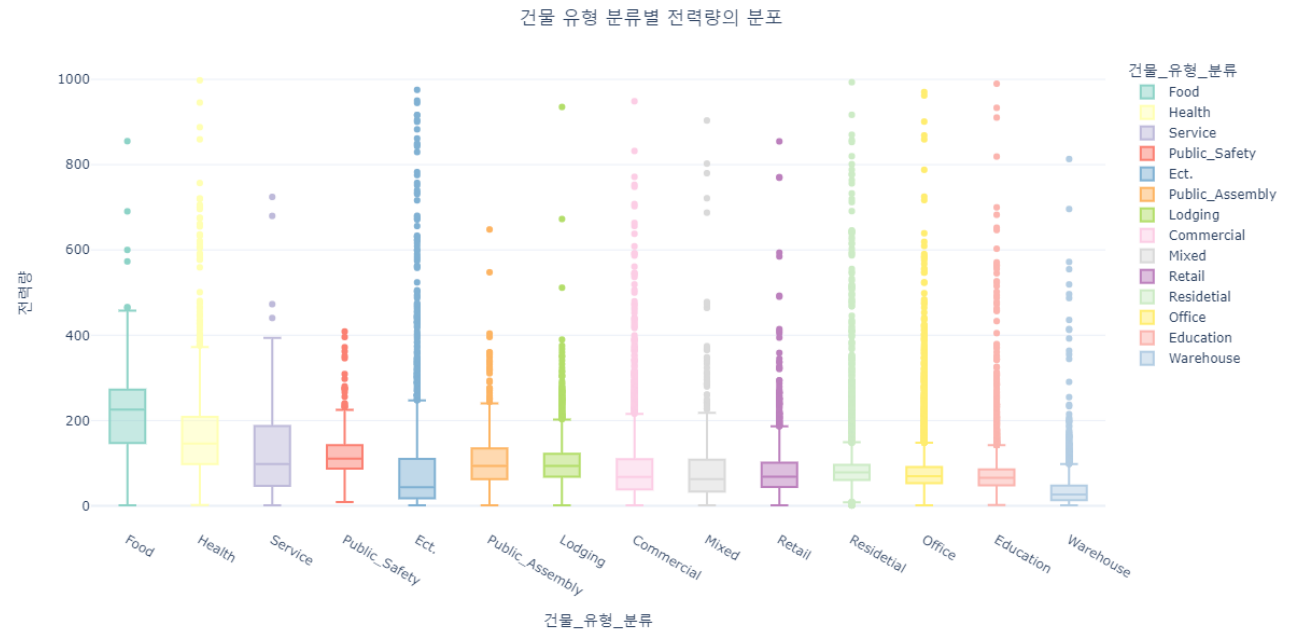
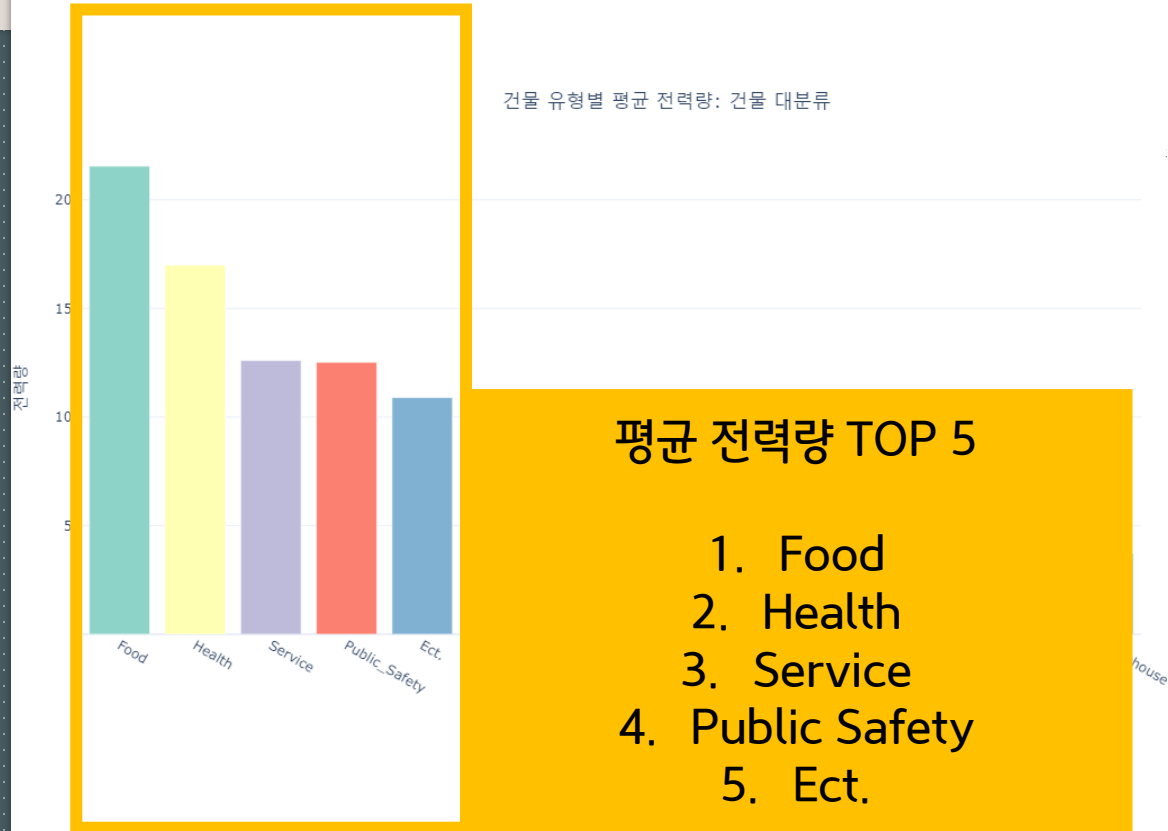


건물 유형 분류별 전력량의 분포



3. EDA 및 변수 설명

건물 유형 (대분류)

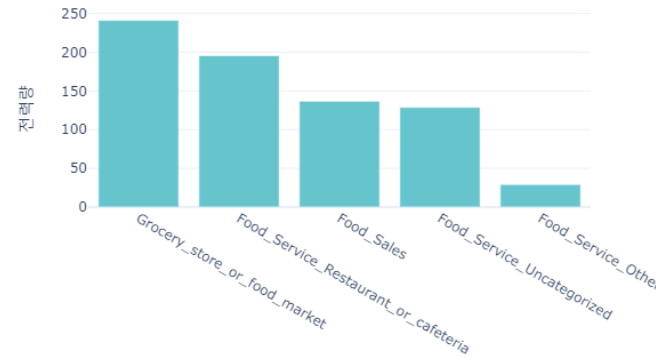


3. EDA 및 변수 설명

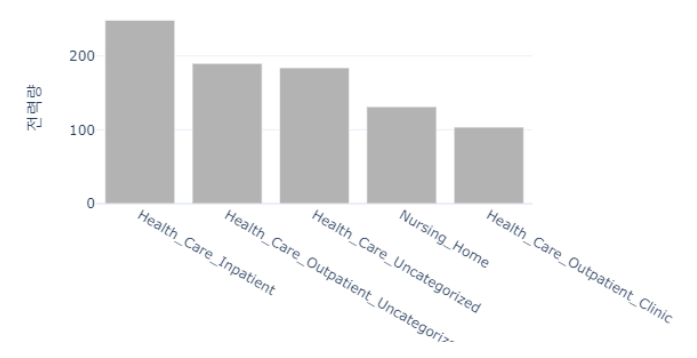
건물 유형 (대분류)

Grocery store or food market	식료품점
Food Service Uncategorized	음식점
Food Service Restaurant or cafeteria	음식점(식당)
Food Service Other	음식점 기타
Food Sales	식료품점
Health Care Uncategorized	기타 헬스케어기관
Health Care Outpatient Uncategorized	외래 미분류
Health Care Outpatient Clinic	외래 클리닉
Health Care Inpatient	입원 병동
Nursing Home	양로원
Service Vehicle service repair shop	카센터
Service Uncategorized	기타 서비스센터
Service Drycleaning or Laundry	세탁소
Public Safety Uncategorized	기타 공공 기관
Public Safety Penitentiary	교도소
Public Safety Fire or police station	소방서 또는 경찰서
Public Safety Courthouse	법원
Data Center	데이터 센터
Laboratory	실험실
Industrial	산업체
Parking Garage	주차공간

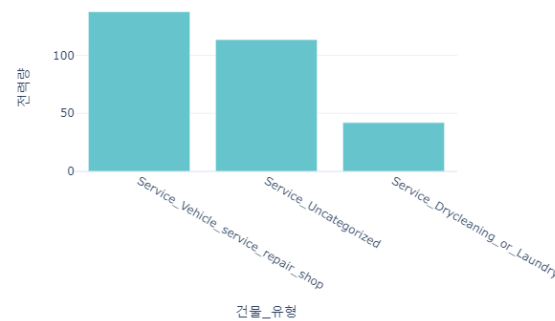
건물유형별 전력량 평균 : Food



건물유형별 전력량 평균 : Health



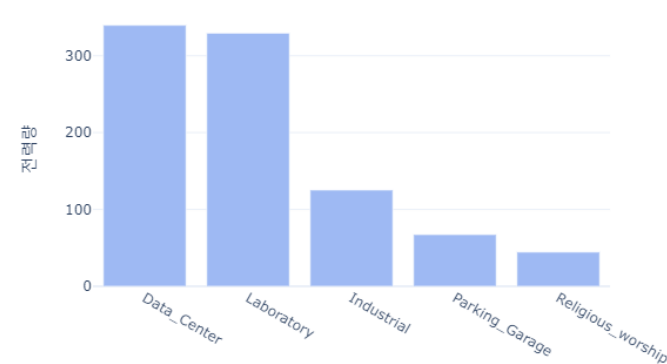
건물유형별 전력량 평균 : Service



건물유형별 전력량 평균 : Public_Safety



건물유형별 전력량 평균 : Ect.



1. Food
2. Health
3. Service
4. Public Safety
5. Ect.

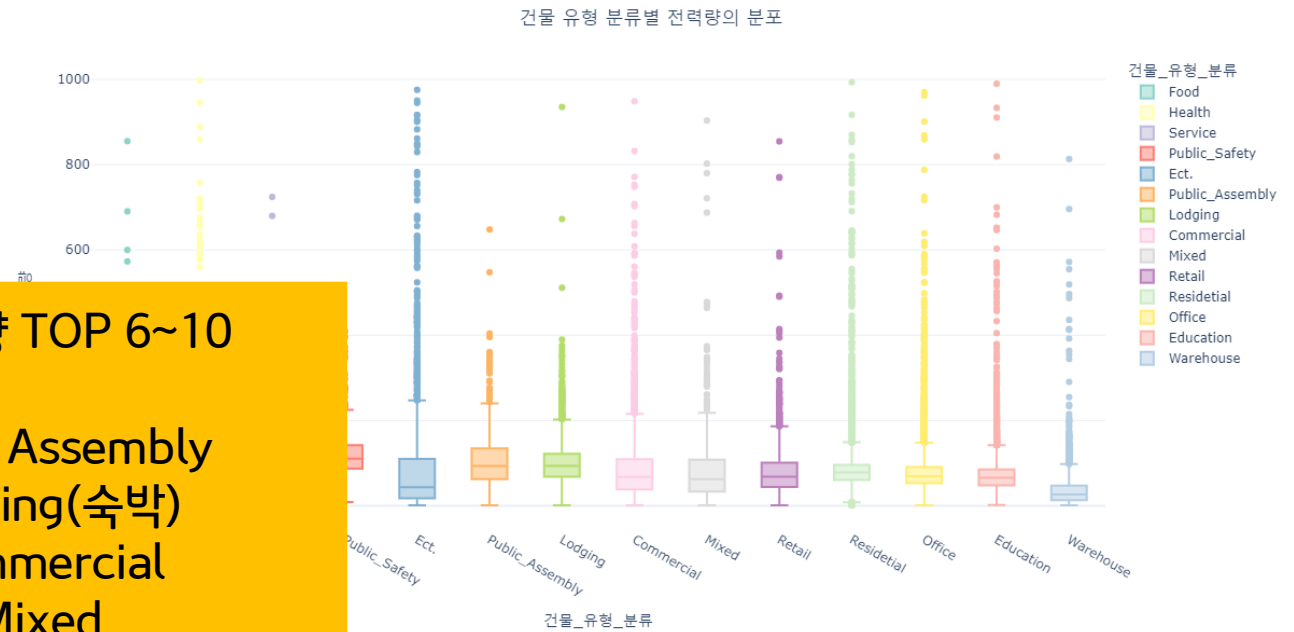
3. EDA 및 변수 설명

건물 유형 (대분류)



평균 전력량 TOP 6~10

6. Public Assembly
7. Lodging(숙박)
8. Commercial
9. Mixed
10. Retail

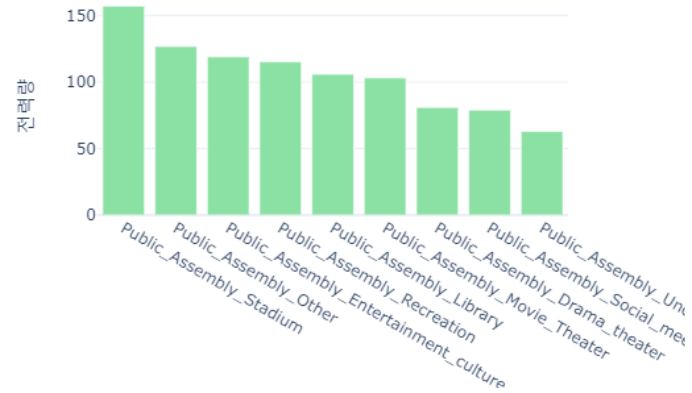


3. EDA 및 변수 설명

건물 유형 (대분류)

Public Assembly Uncategorized	공공 장소 미분류
Public Assembly Stadium	공공 스타디움
Public Assembly Social meeting	공공 모임 장소
Public Assembly Recreation	공공 휴게 장소
Public Assembly Other	기타 공공 장소
Public Assembly Movie Theater	영화관
Public Assembly Library	도서관
Public Assembly Entertainment culture	문화센터
Public Assembly Drama theater	극장
Lodging Uncategorized	숙박 미분류
Lodging Other	숙박 기타
Lodging Hotel	호텔
Lodging Dormitory or fraternity sorority	호스텔
Commercial Unknown	상업용 미분류
Commercial Other	상업용 기타
Mixed Use Predominantly Residential	복합 - 주거위주
Mixed Use Predominantly Commercial	복합 - 상업위주
Mixed Use Commercial and Residential	복합 - 기타
Retail Vehicle dealership showroom	차량(및 기타 운송수단 판매)
Retail Uncategorized	기타 소매상
Retail Strip shopping mall	실외 쇼핑센터
Retail Enclosed mall	실내 복합쇼핑센터

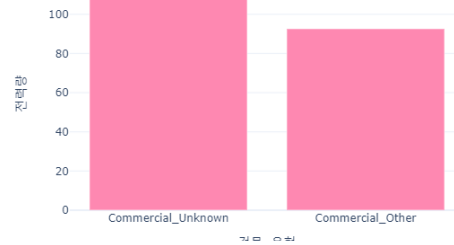
건물유형별 전력량 평균 : Public_Assembly



건물유형별 전력량 평균 : Lodging



건물유형별 전력량 평균 : Commercial



건물유형별 전력량 평균 : Mixed



건물유형별 전력량 평균 : Retail

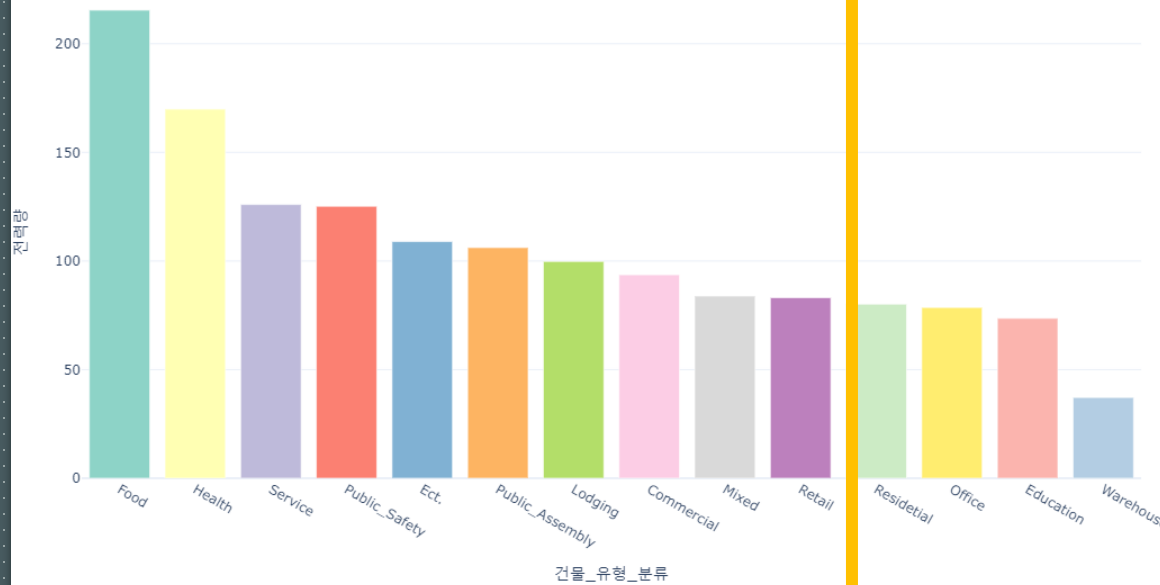


6. Public Assembly
7. Lodging(숙박)
8. Commercial
9. Mixed
10. Retail

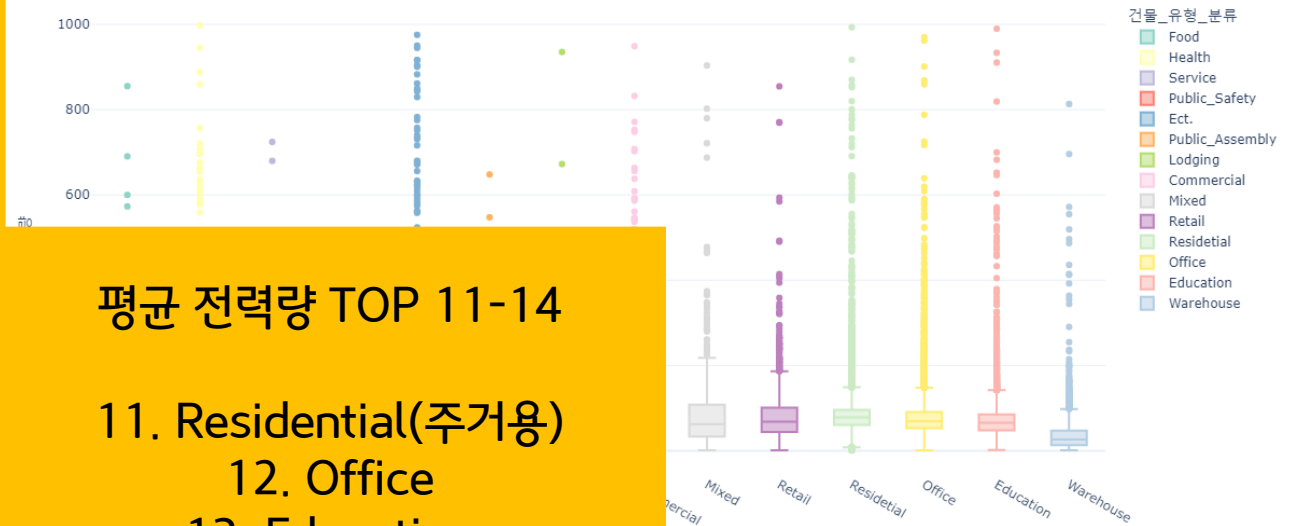
3. EDA 및 변수 설명

건물 유형 (대분류)

건물 유형별 평균 전력량: 건물 대분류



건물 유형별 분류별 전력량의 분포



평균 전력량 TOP 11-14

11. Residential(주거용)

12. Office

13. Education

14. Warehouse

3. EDA 및 변수 설명

건물 유형 (대분류)

Multifamily Uncategorized	주거 공간 미분류
5plus Unit Building	빌라 - 5plus unit
2to4 Unit Building	빌라 - 2~4 unit
Office Uncategorized	기타 사무공간
Office Mixed use	복합 사무실
Office Medical non diagnostic	의료 사무실
Office Bank or other financial	은행 및 금융기관
Education Uncategorized	교육기관 미분류
Education Preschool or daycare	유치원 어린이집
Education Other classroom	교실 교육기관
Education College or university	대학
Warehouse Uncategorized	기타 창고
Warehouse Selfstorage	개인 창고
Warehouse Refrigerated	냉장용 창고
Warehouse Nonrefrigerated	냉장 아닌 창고
Warehouse Distribution or Shipping center	물류 창고

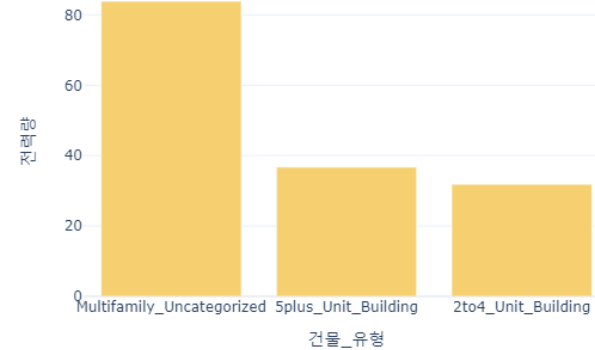
11. Residential(주거용)

12. Office

13. Education

14. Warehouse

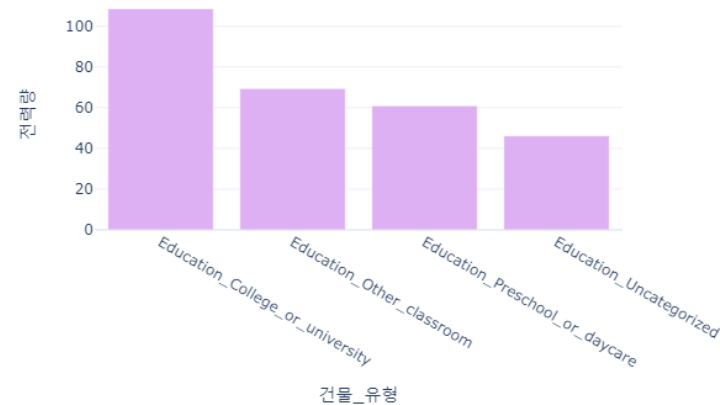
건물유형별 전력량 평균 : Residential



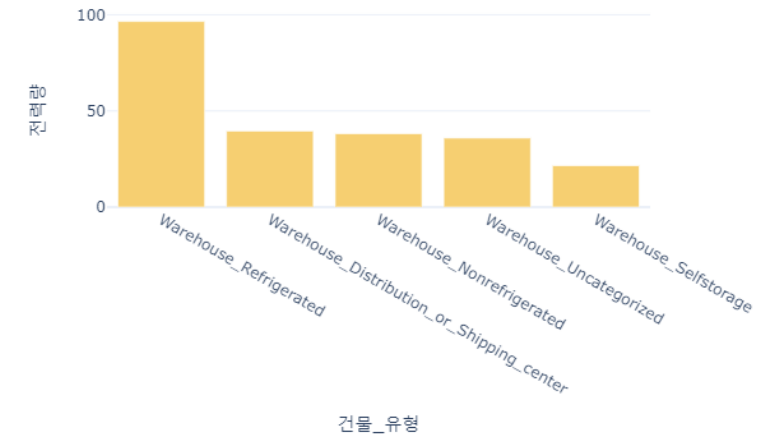
건물유형별 전력량 평균 : Office



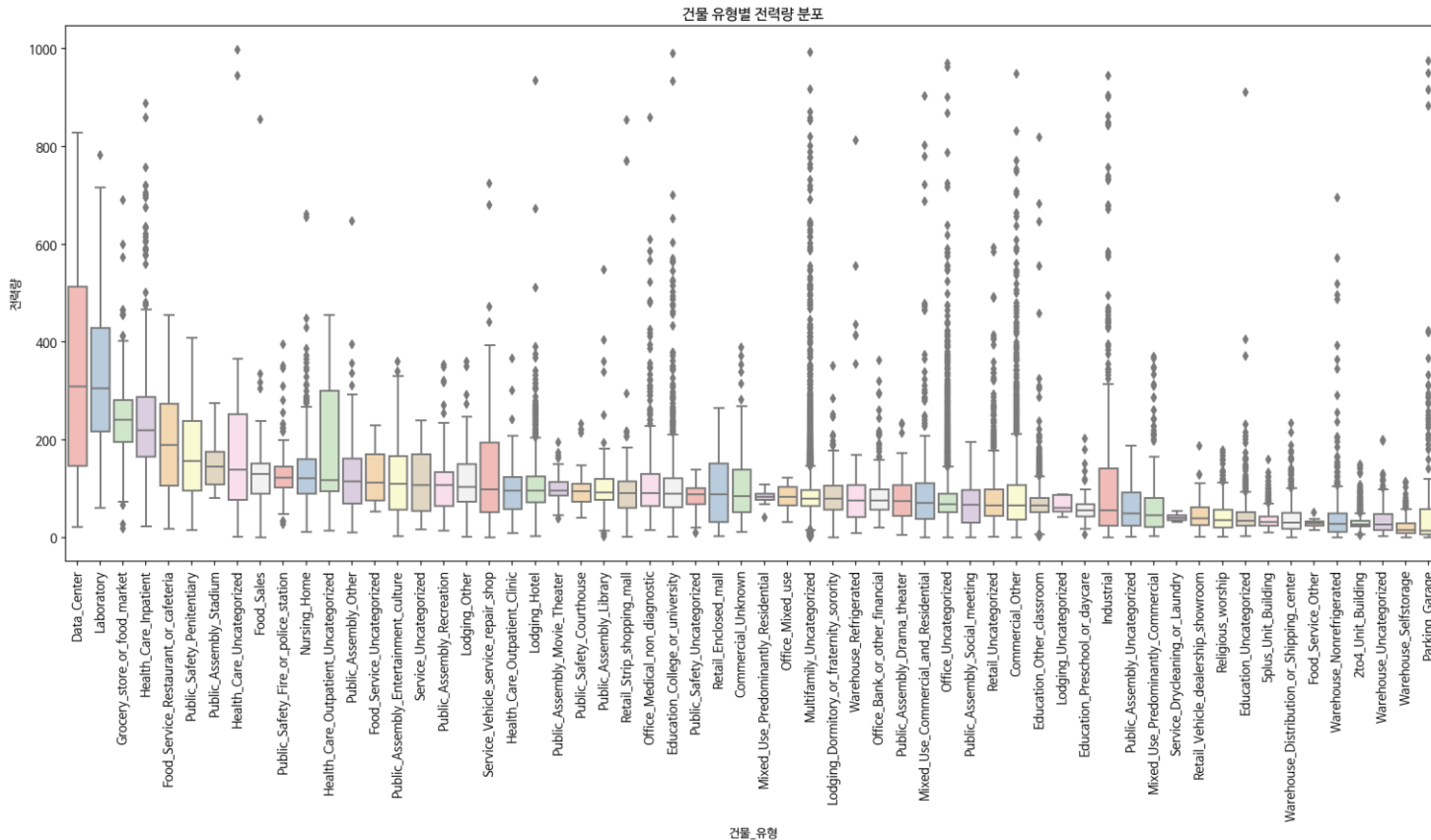
건물유형별 전력량 평균 : Education



건물유형별 전력량 평균 : Warehouse



건물 유형 (소분류)



건물 유형별 전력량의 중간값 기준의 워드클라우드.

3. EDA 및 변수 설명

건물 유형 (24시간)

많은 전력을 24시간동안 써야하는 건물의 경우를 분류하여 **건물_유형_24** 라는 이진 범주형 변수를 만들었다

데이터 센터

실험실/연구실

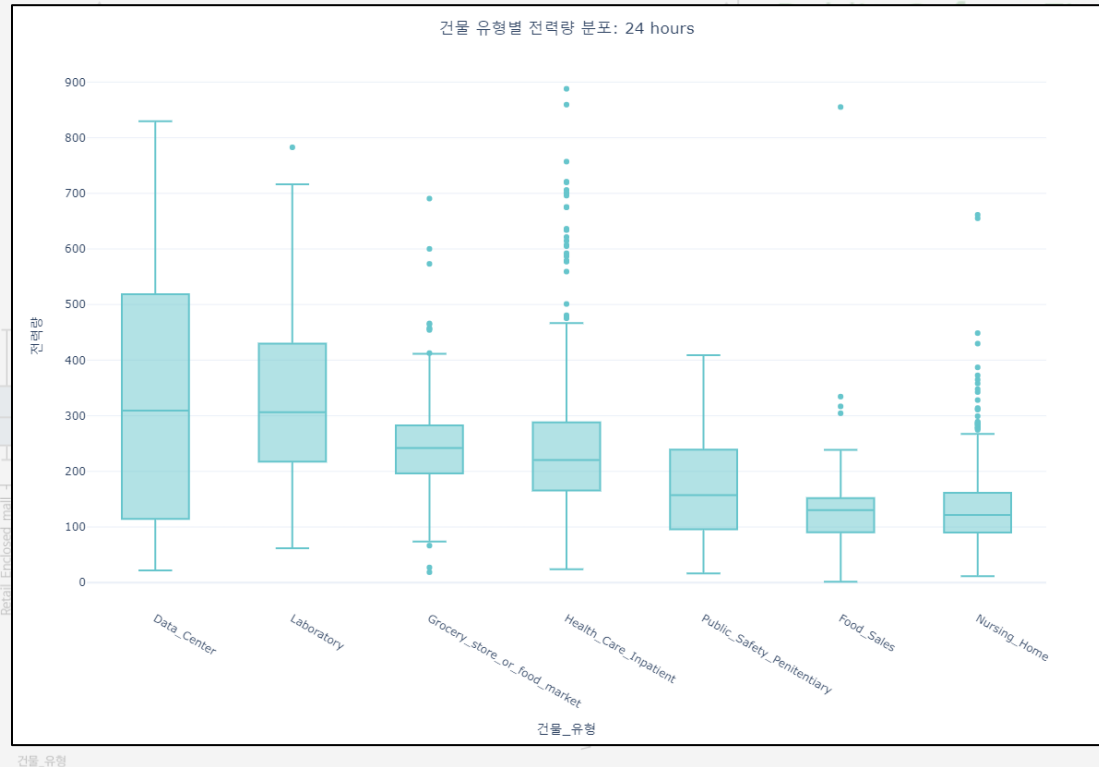
식료품점

기타 식료품점

입원 병동

요양원

교도소



의 중간값 기준의 워드클라우드.

3. EDA 및 변수 설명

건물 유형 (24시간)

많은 전력을 24시간동안 써야하는 건물의 경우를 분류하여 **건물_유형_24** 라는 이진 범주형 변수를 만들었다

데이터 센터

실험실/연구실

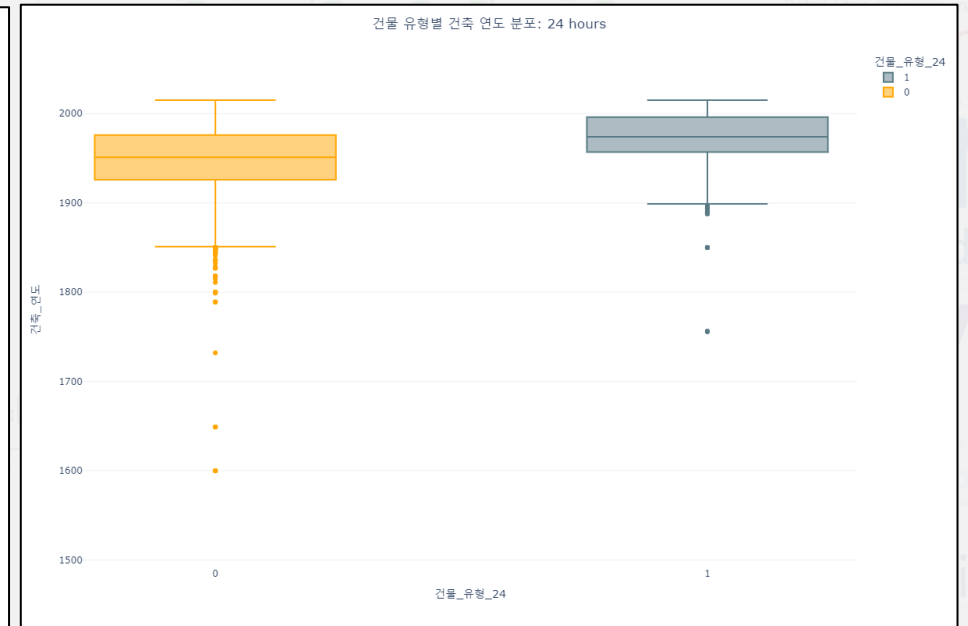
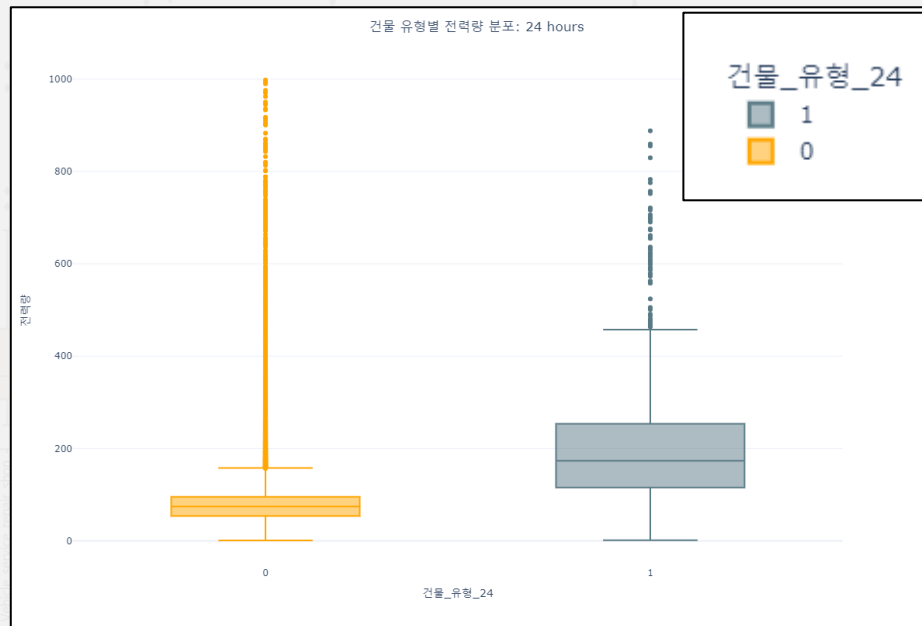
식료품점

기타 식료품점

입원 병동

요양원

교도소



24시간 운영에 해당하는 건물들이 중간값 기준 에너지 소비량이 더 큰 것을 확인

건물 유형별 건축 연도가 비교적 최근임.



기온 변수를 파헤쳐 보자

3. EDA 및 변수 설명

- 기온 변수 pca

1월 최저기온, 평균기온, 최고기온,
2월 최저기온, 평균기온, 최고기온,
...
12월 최저기온, 평균기온, 최고기온

총 36개의 기온 변수

PC 1	PC 2	PC 3
47.54%	16.24%	11.79%

=> 3개의 주성분으로 총 75.56% 설명

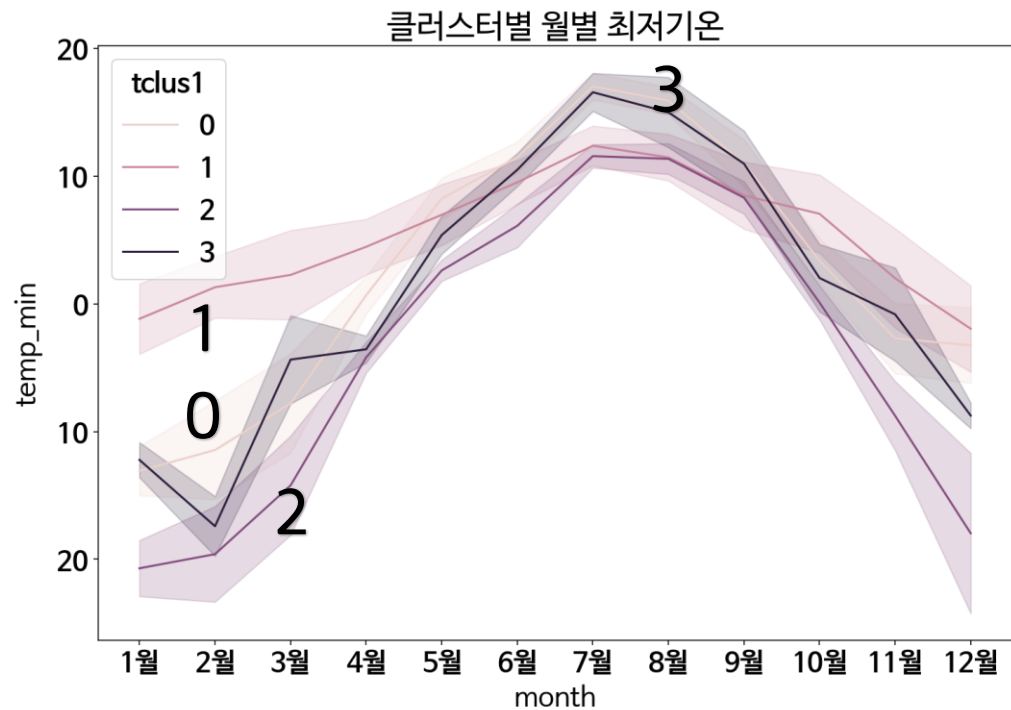
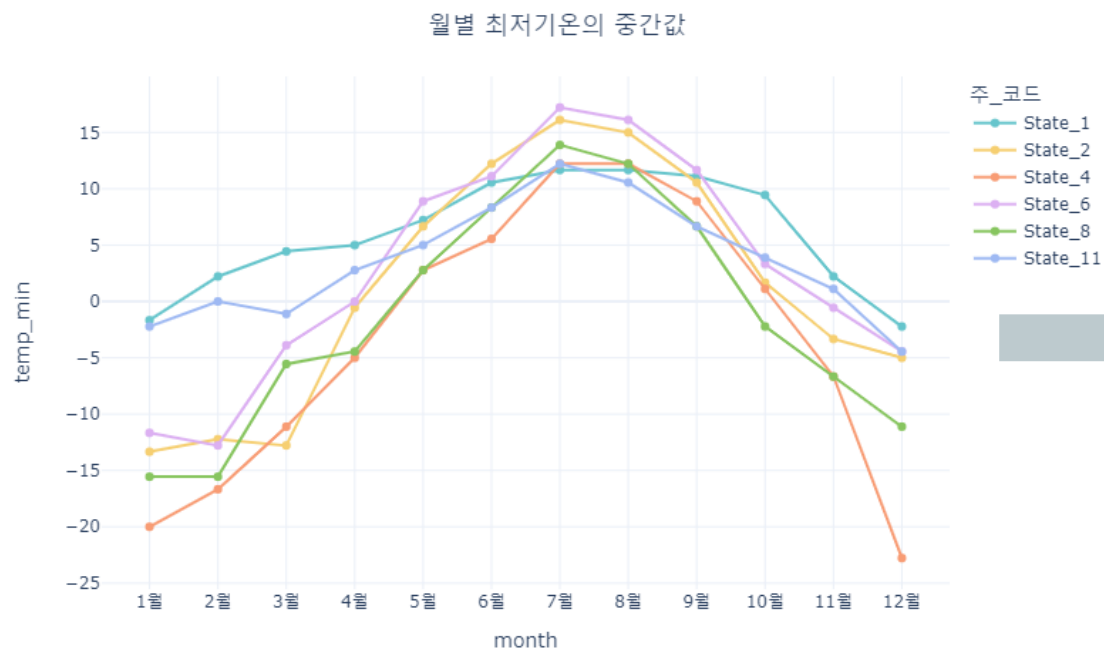
PC1 계수 (절대값 큰 순)

변수명	계수
2월 최저기온	0.457
3월 최저기온	0.369
1월 최저기온	0.362
2월 평균기온	0.348
1월 평균기온	0.277
2월 최고기온	0.248
3월 평균기온	0.228
11월 최저기온	0.179
4월 최저기온	0.169
1월 최고기온	0.157
10월 최저기온	0.143

겨울의 최저기온이 많은 정보를 담고 있다.

3. EDA 및 변수 설명

월별 최저기온을 이용한 시계열 클러스터링



Clus 1은 비교적 따뜻한 지역

Clus 2는 추운 지역

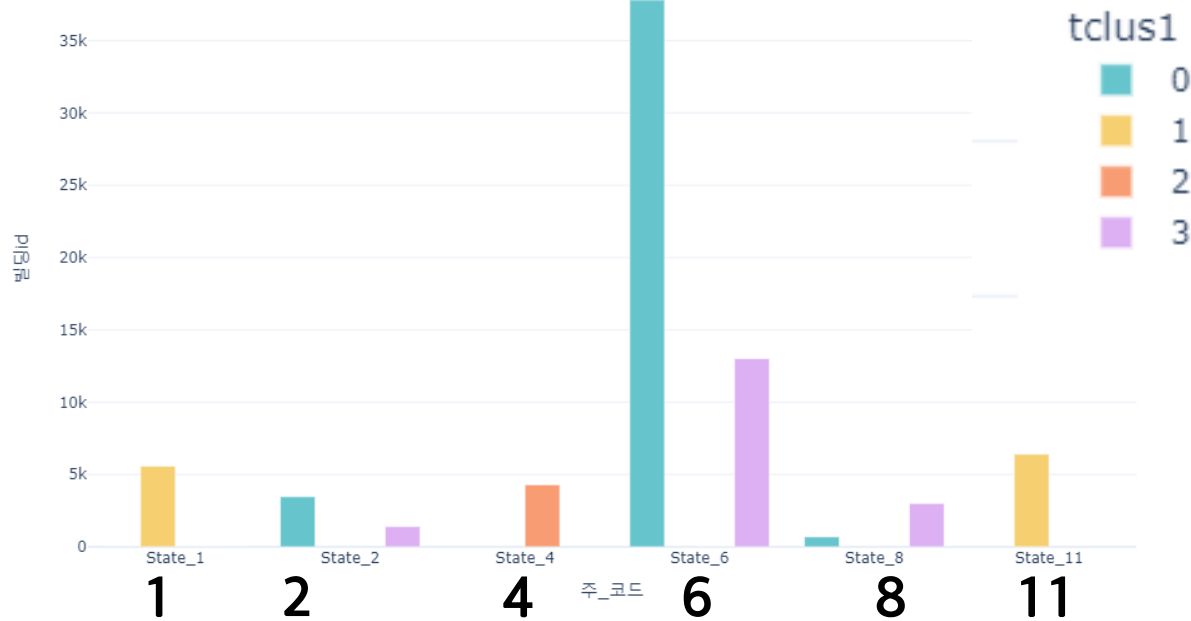
Clus 3은 1월보다 2월에 추워지는 지역, 여름에 가장 더운 지역

Clus 0은 가장 평균에 가까운 일반적인 지역

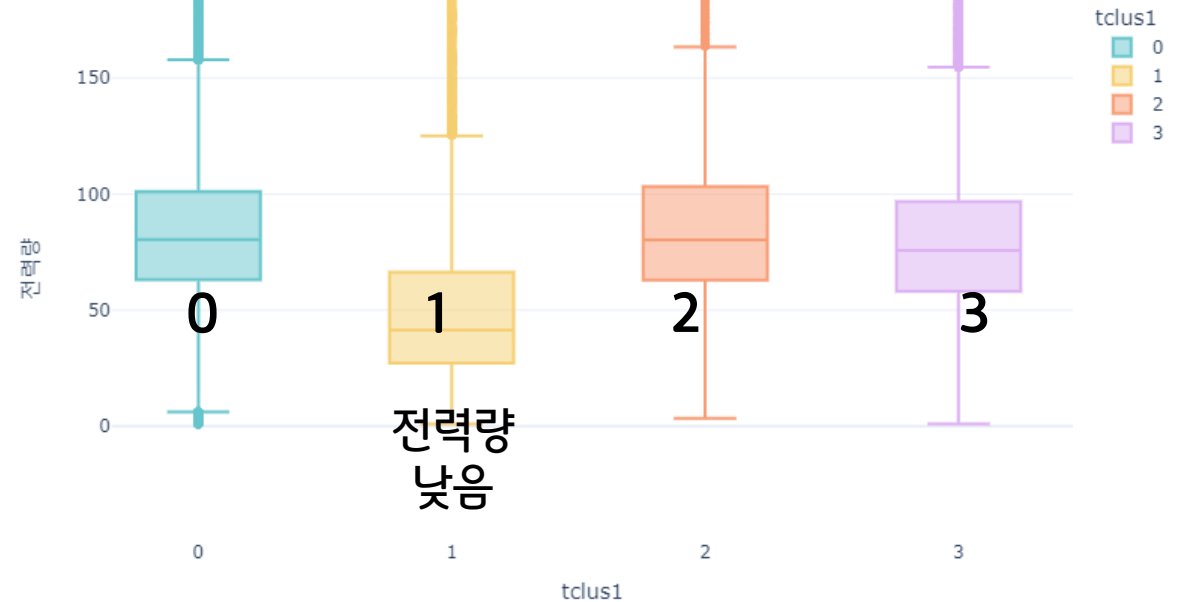
3. EDA 및 변수 설명

기온과 주 코드(State)

클러스터와 주 코드 비교



클러스터별 전력량 분포



Clus 1은 비교적 따뜻한 지역 - state 1과 11이 함께 묶임

Clus 2는 추운 지역 - state 4와 동일

Clus 3은 1월보다 2월에 추워지는 지역, 여름에 가장 더운 지역 - state 2, 6, 8에 존재

Clus 0은 가장 평균에 가까운 일반적인 지역 - state 6에 압도적으로 존재

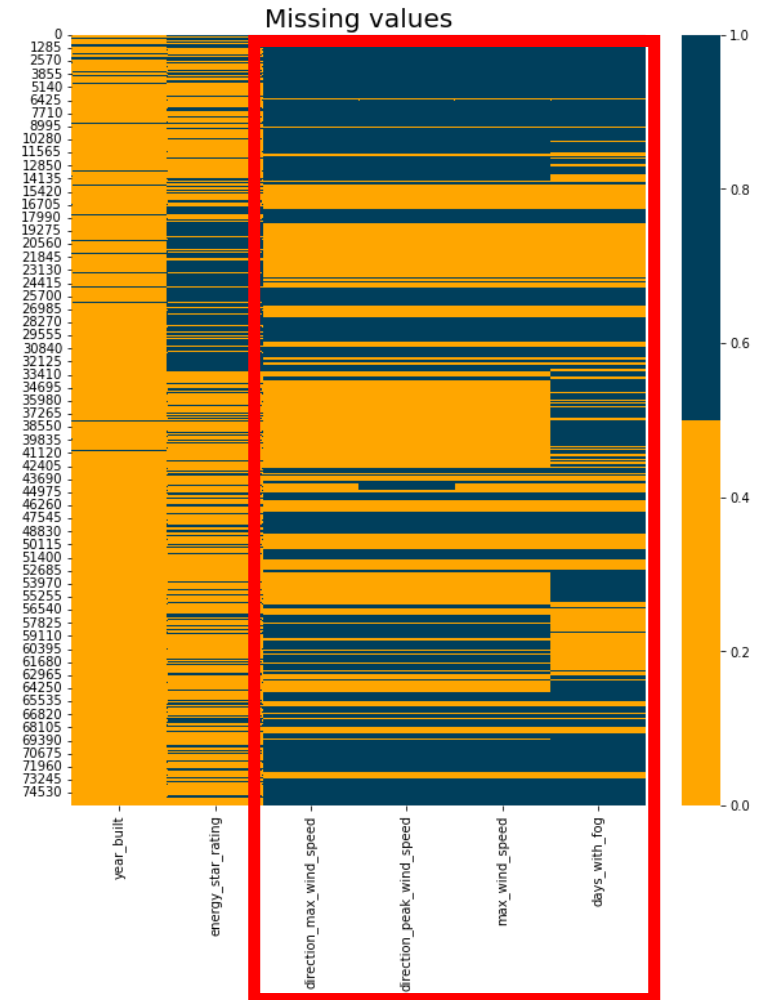


모델링

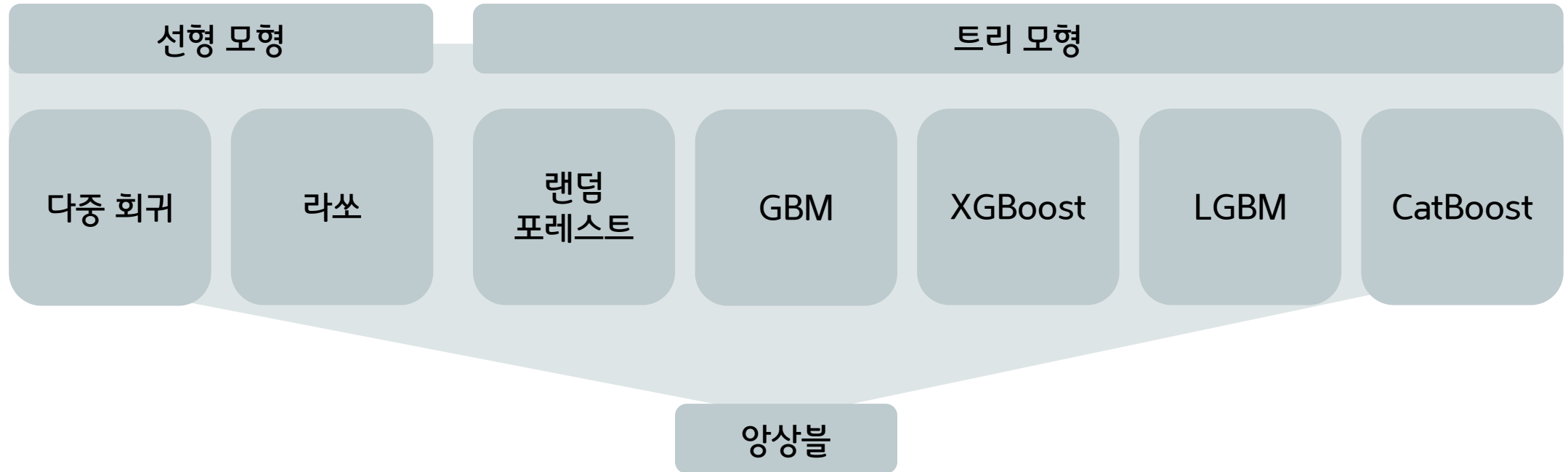
4. 모델링 - 결측치 처리

변수	결측치 개수	결측치 비율
건축 연도	1837	2.424858
에너지 등급	26709	35.25615
최대 풍속 풍향	41082	54.22865
돌풍 풍속 풍향	41811	55.19094
최대 풍속	41082	54.22865
안개 일수	45796	60.45118

- 건축년도 : 주 별 중간값으로 대체
- 에너지 등급 : 전체 중간값으로 대체
- 나머지 : 열 삭제



4. 모델링 - 과정

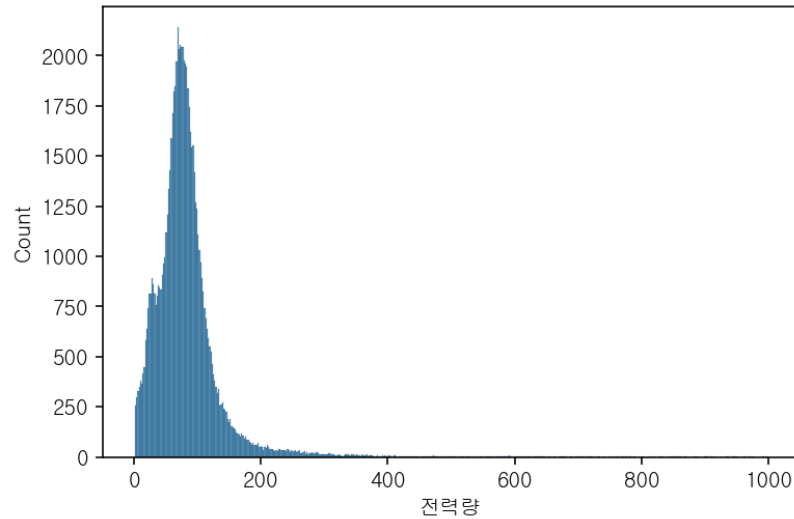


Train set	Test set
80%	20%

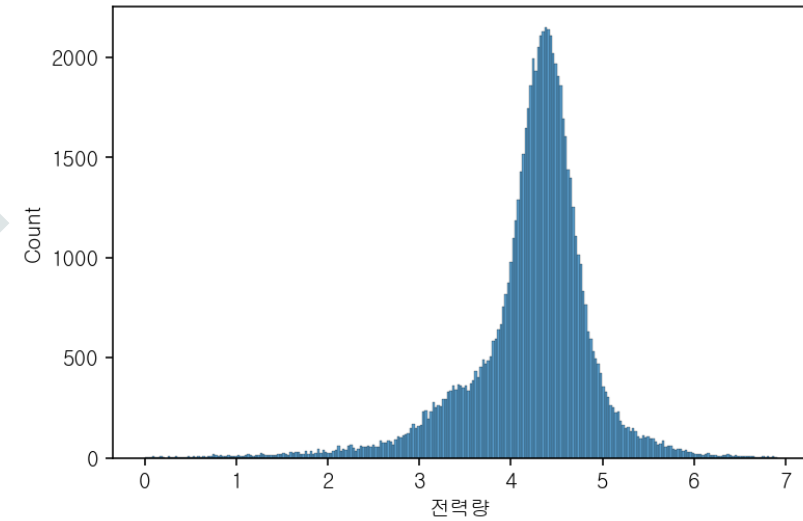
비교 지표 : RMSE (평균 제곱근 오차) $\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$

4. 모델링 - 결과

선형 모델



Y 로그 변환



Y 모델링		Log Y 모델링		PCA 기온 변수 사용	
다중 회귀	라쏘	다중 회귀	라쏘	다중 회귀	라쏘
50.727	50.702	52.399	52.507	52.732	52.728

4. 모델링 - 결과

트리 모형 : 최종

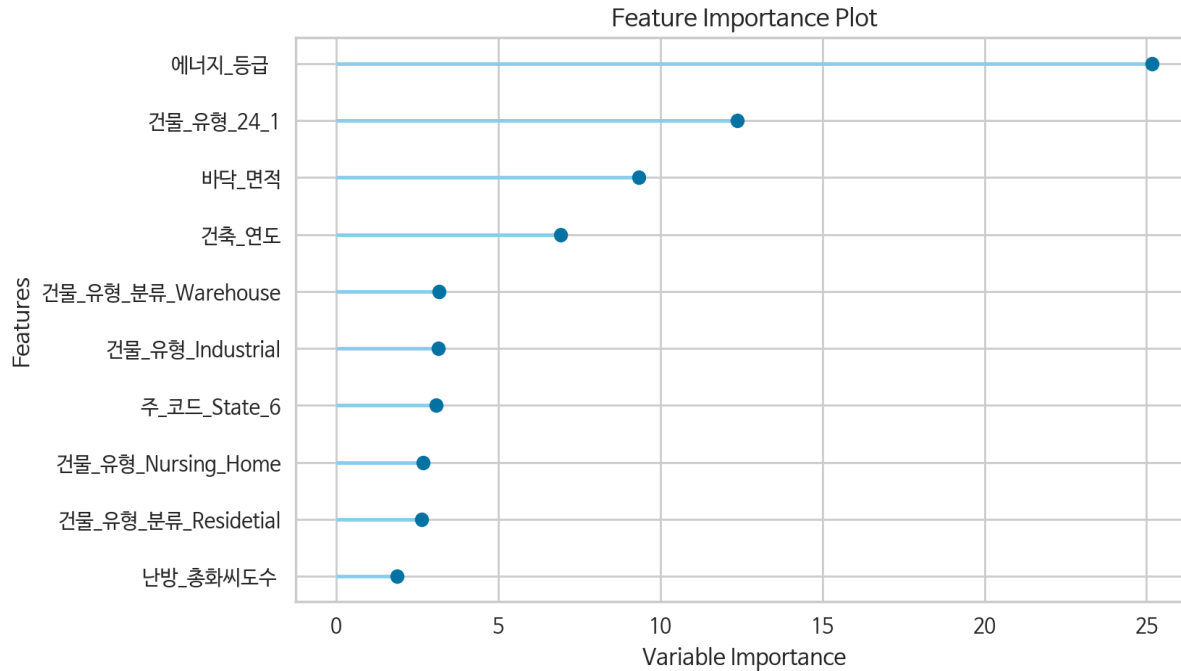
	랜덤포레스트	GBM	XGBoost	LightGBM	CatBoost
그리드서치 결과	max depth 20 min samples leaf 5 n estimators 100	learning rate 0.1 max depth 10 min samples leaf 10 n estimators 200	learning rate 0.2 max depth 20 min child weight 10	learning rate 0.1 max depth 10 min child samples 10 n estimators 200	learning rate 0.1 l2 leaf reg 0.01
test RMSE	41.767	40.505	46.771	40.923	40.372

CatBoost + RandomForest + XGBoost 앙상블

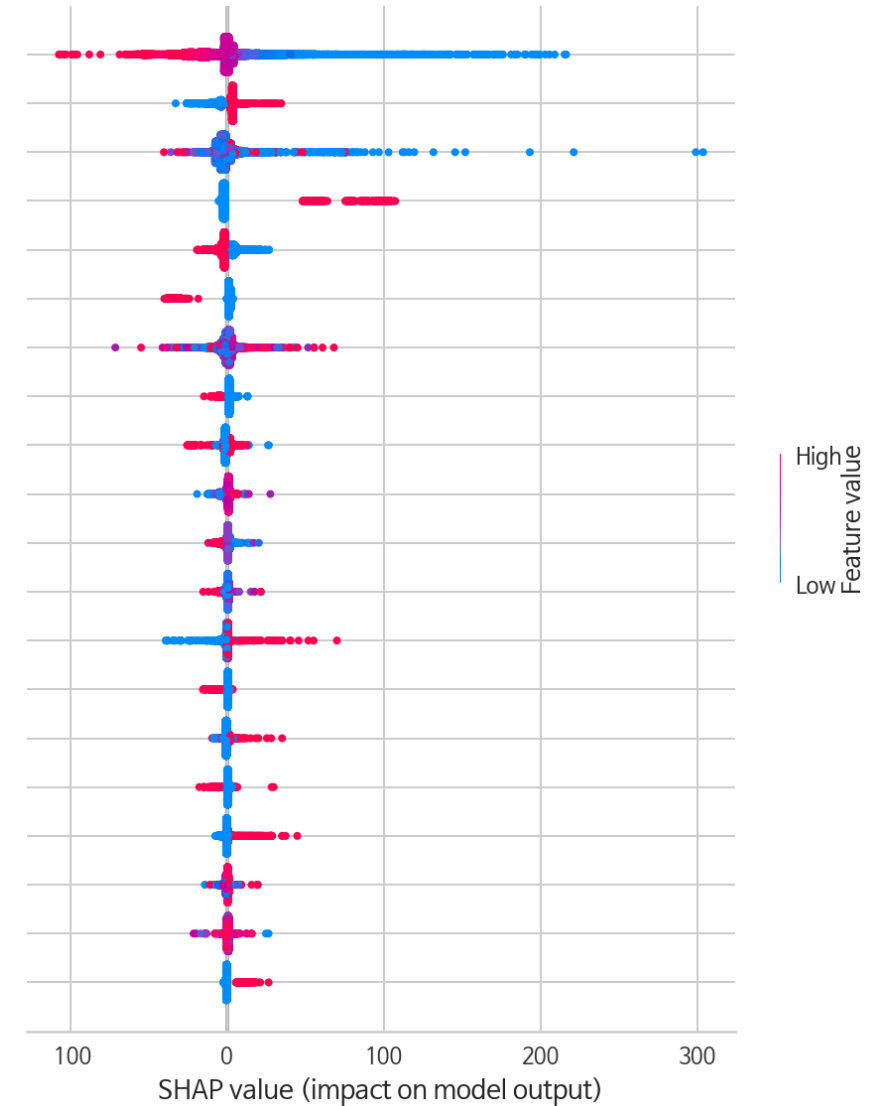
RMSE 37.971

4. 모델링 - 결과 해석

변수 중요도

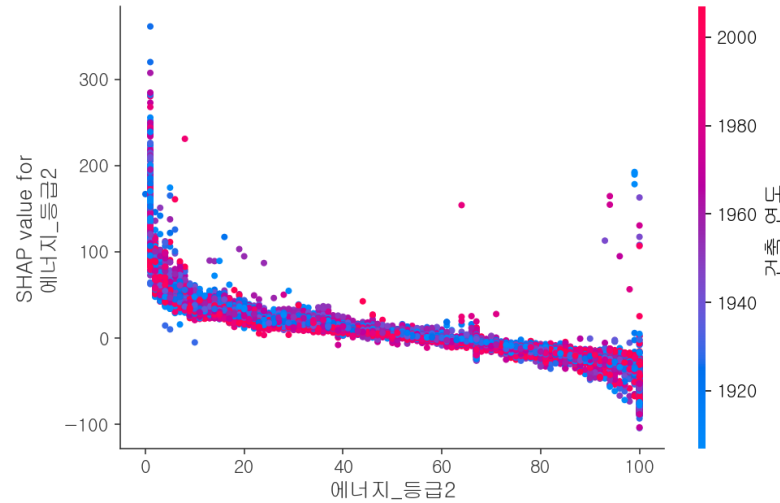
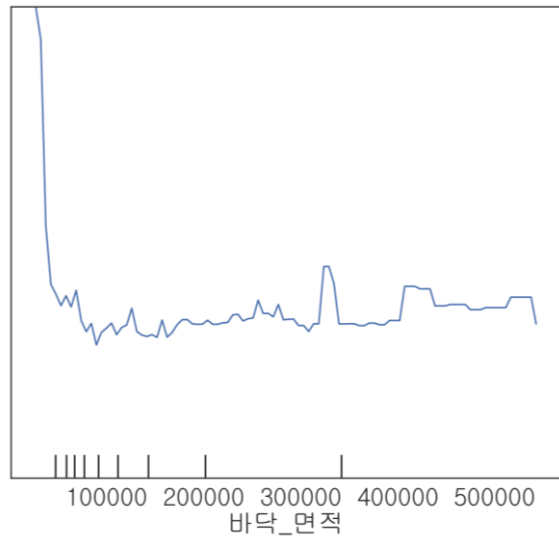
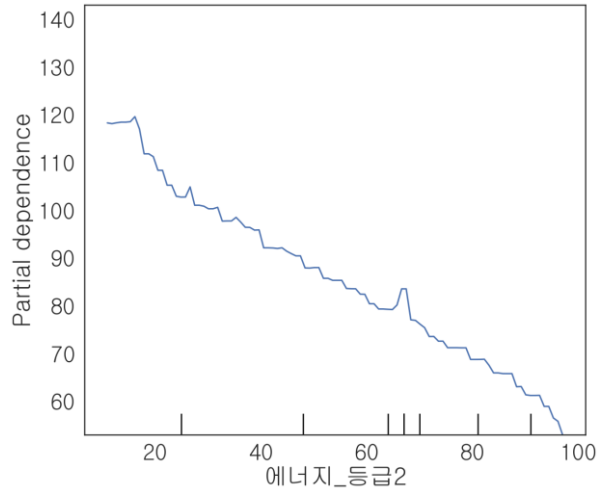


에너지_등급
주_코드_State_6
바닥_면적
건물_유형_24_1
건물_유형_분류_Residential
건물_유형_분류_Warehouse
건축_연도
최저기온_clus_1
주거상업여부_Commercial
난방_총화씨도수
최저_기온_1월
최저_기온_2월
최저기온_clus_0
건물_유형_분류_Education
건축_연도_그룹_1950_1975
주_코드_State_2
건물_유형_분류_Office
최저_기온_7월
냉각_총화씨도수
건물_유형_Lodging_Hotel



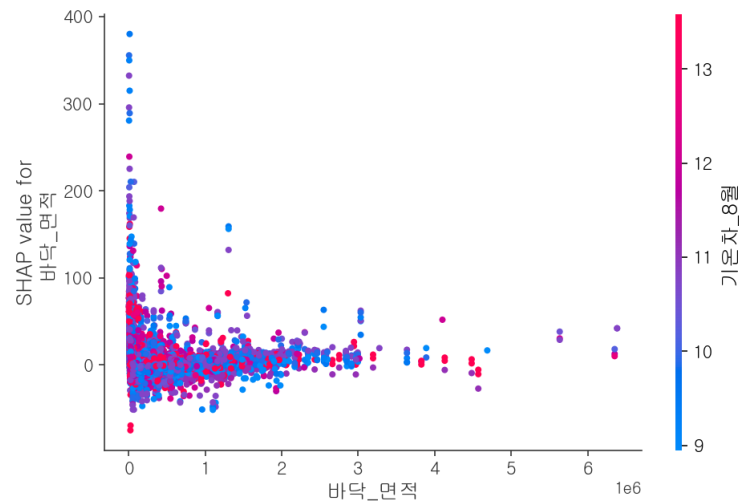
4. 모델링 - 결과 해석

PDP, Shap value



에너지 등급

등급이 높을수록 전력을 적게 사용한다.
같은 에너지 등급 내에서는, 최근에 지어진 건물일수록 전력량이 더 감소한다.

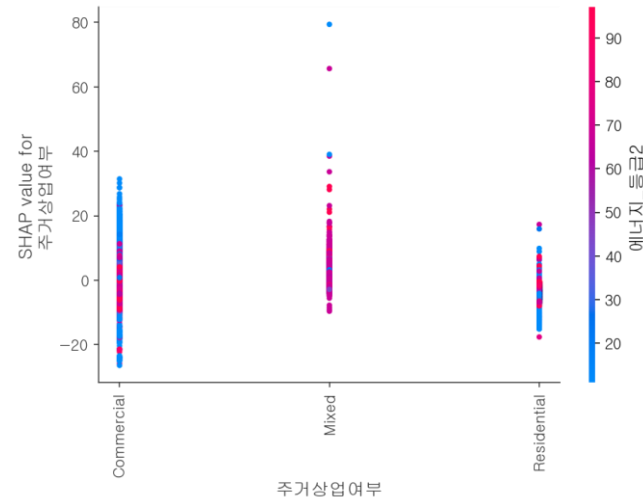
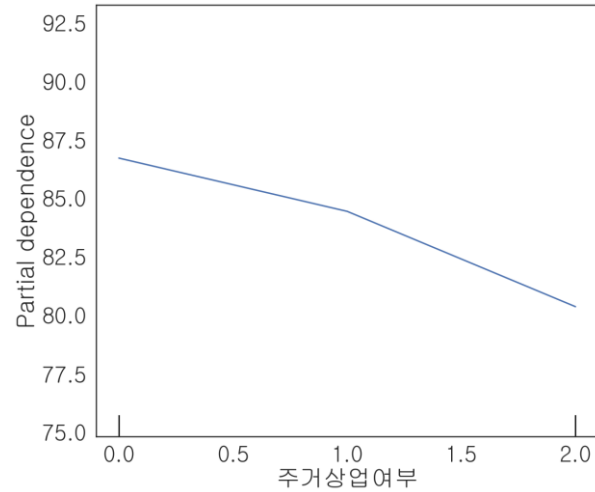


바닥 면적

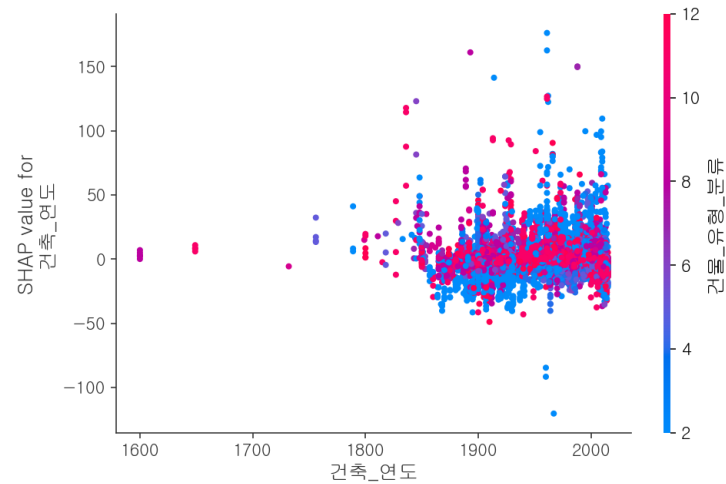
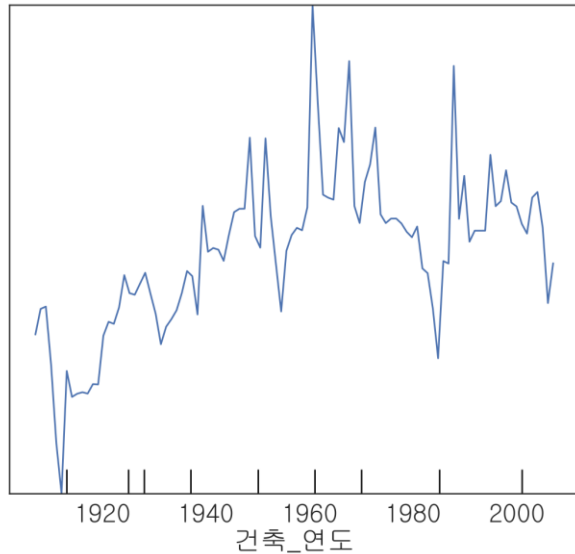
전반적으로 사용 면적이 매우 좁을 때,
전력량이 높음.

4. 모델링 - 결과 해석

PDP, Shap value



주거/상업
상업지구 > 복합 > 주거지구



건축 연도
1950 ~ 1980년 사이의 전력량이 가장 높다.

5. 결론 및 한계점

변수	해석
에너지 등급	에너지 등급이 높을수록 전력량 감소
건물 유형 24	24시간 돌아가는 건물의 전력량이 높음
바닥 면적	바닥 면적이 매우 좁으면 전력량 증가
주 코드	state 6일 때 전력량 증가 (많은 표본의 효과), state 2일 때 전력량 감소 (비교적 최신 건물)
건축년도	1950 ~ 1975년에 지어진 건물에서 전력량 높음
주거상업여부	상업용 건물보다 주거용 건물의 전력량이 낮음
건물 유형	창고, 교육 - 낮음 / 오피스, 숙박, 요양원 - 높음
1, 2월 최저 기온	겨울에 따뜻할수록 전력량 감소

결론

상업용 건물, 오피스, 24시간 돌아가는 건물 유형에서 전력 소모를 줄이기 위해 노력해야 한다.

난방을 위한 에너지 효율화를 통해 많은 에너지 절감 효과를 노릴 수 있을 것이다.

에너지 등급이 전력량을 잘 설명하고 있으므로 이를 평가하고 감독하는 추가적인 제도가 도움이 될 것이다.

5. 결론 및 한계점

한계점

- ✓ 건물의 사용 인원, 난방 시스템, 층 수, 전력기기 개수 등 전력량에 영향을 미치는 충분한 데이터가 있다면 더 좋은 성능과 해석이 가능할 것이다.
- ✓ 특정 주의 건물들은 같은 기온 데이터를 가지고 있었다.
따라서, 개별 건물의 기상 정보를 알았더라면 정확한 해석이 가능할 것이다.
- ✓ 이 건물들의 정확한 위치 정보를 안다면 외부 데이터를 사용하기 쉬울 것이다.
- ✓ 다양한 나라의 건물을 분석한다면 각 정보의 영향을 다채롭게 파악할 수 있을 것이다.

감사합니다!

