

# 전력 사용량 예측 공모전

---

인공지능 기반 건물 유형별 전력소비량 예측

# CONTENTS

## I

### 프로젝트 개요

1. 프로젝트 필요성
2. 프로젝트 설명
3. 프로젝트 개념도

## II

### 데이터 이해

1. 개발 환경
2. 데이터 설명

## III

### 데이터 전처리

1. building 전처리
2. train 전처리
3. test 전처리

## IV

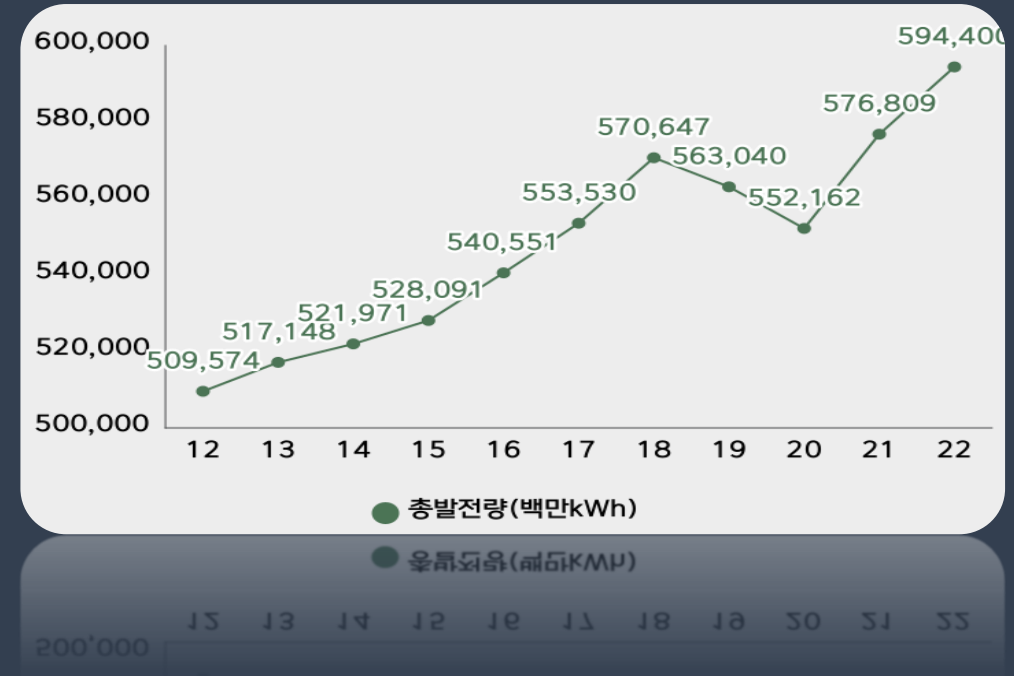
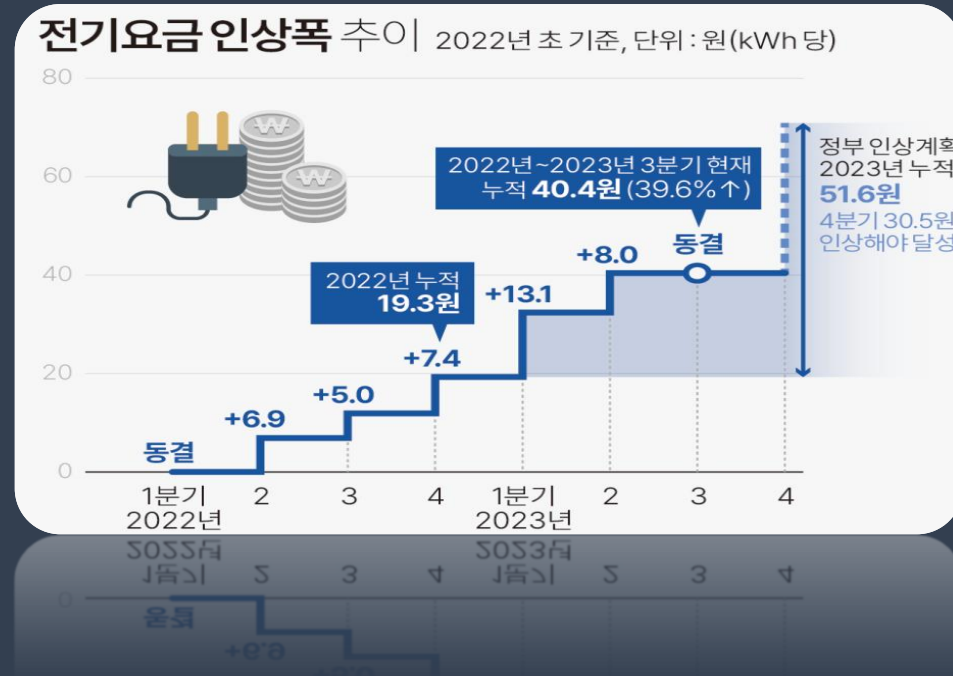
### 모델 설계 및 개발

1. 모델 선정
2. 모델 개발 및 평가
3. 최종 모델 선정

# Part 1, 프로젝트 개요



# 1. 프로젝트 개요



꾸준한 전력 소비량과 전기세 상승에 따른 효율적인 에너지 공급의 필요성 증가  
 따라서 전력 사용량 예측 시뮬레이션을 통한 효율적인 인공지능 알고리즘 필요

# 1. 프로젝트 개요



건물정보 데이터와 시공간 날씨 데이터를 이용해 특정 시점의 전력 사용량 예측  
SMAPE 값을 평가 기준으로 가장 낮은 점수를 받는 사람이 우승하는 방식으로 진행

# 1. 프로젝트 개요

## 01. 데이터 수집

**Building**  
(건물 정보 데이터)

**Train**  
(학습 데이터)

**Test**  
(예측 데이터)

**Submission**  
(예측 값 제출 파일)

## 02. 분석 및 전처리

데이터 이해

데이터 분석 및 시각화

데이터 전처리

## 03. 모델 생성 및 개발

모델 설계

모델 개발

모델 결과

모델 평가

# Part 2, 데이터 이해



## 2. 데이터 이해

### Language



Python  
3.9.5

### Tool

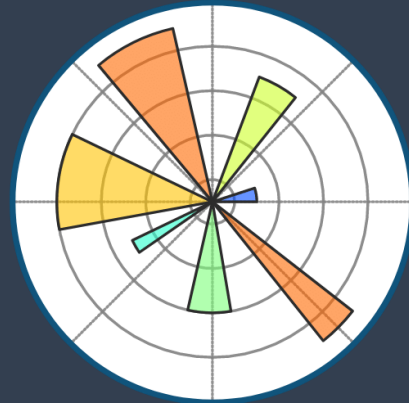


Visual Studio Code



Github & Git

### Visualization



Matplotlib



## 2. 데이터 이해

### Building - 건물정보 데이터

| 변수명      | 데이터 Shape           | 데이터 자료형 | 변수 설명                |            |
|----------|---------------------|---------|----------------------|------------|
| 건물유형     | 100행 7열<br>(100, 7) | Int64   | 1. 건물기타              | 7. 상용      |
|          |                     |         | 2. 공공                | 8. 아파트     |
|          |                     |         | 3. 대학교               | 9. 연구소     |
|          |                     |         | 4. 데이터센터             | 10. 지식산업센터 |
|          |                     |         | 5. 백화점및아울렛           | 11. 할인마트   |
|          |                     |         | 6. 병원                | 12. 호텔및리조트 |
| 건물번호     |                     | Object  | 건물1 ~ 건물100 (총 100개) |            |
| 연면적      |                     | Float64 | 모든 층의 바닥 면적 합        |            |
| 냉방면적     |                     | Float64 | 냉방이 필요한 공간의 면적       |            |
| 태양광용량    |                     | Object  | 태양광이 생성 가능한 전력 총량    |            |
| ESS 저장용량 |                     | Object  | 저장할 수 있는 최대 에너지 양    |            |
| PCS 저장용량 |                     | Object  | 시스템이 처리 가능한 최대 전력량   |            |

## 2. 데이터 이해

**train** - 시공간/날씨 데이터 (2022.06.01 ~ 2022.08.24)

| 변수명           | 데이터 Shape                   | 데이터 자료형 | 변수 설명                 |
|---------------|-----------------------------|---------|-----------------------|
| num_date_time | 204000행 10열<br>(204000, 10) | Object  | 건물번호 + 일시             |
| 건물번호          |                             | Int64   | 건물1 ~ 건물100 (총 100개)  |
| 일시            |                             | Object  | 2022.06.01~2022.08.24 |
| 기온            |                             | Float64 | 온도(C)                 |
| 강수량           |                             | Float64 | 강수량(mm)               |
| 풍속            |                             | Float64 | 바람의 속도                |
| 습도            |                             | Float64 | 대기 중 수증기 량            |
| 일조            |                             | Float64 | 땅 위에 비치는 태양광 시간       |
| 일사            |                             | Float64 | 태양으로부터 방출되는 에너지       |
| <b>전력소비량</b>  |                             | Float64 | <b>Target Data</b>    |

## 2. 데이터 이해

### Test - 시공간/날씨 데이터 (2022.08.25 ~ 2022.08.31)

| 변수명           | 데이터 Shape              | 데이터 자료형 | 변수 설명                 |
|---------------|------------------------|---------|-----------------------|
| num_date_time | 16800행 7열<br>(16800,7) | Object  | 건물번호, 기록일시(1시간)       |
| 건물번호          |                        | Int64   | 1~100                 |
| 일시            |                        | Object  | 2022.08.25~2022.08.31 |
| 기온            |                        | Float64 | 온도(C)                 |
| 강수량           |                        | Float64 | 강수량(mm)               |
| 풍속            |                        | Float64 | 바람의 속도                |
| 습도            |                        | Int64   | 대기 중 수증기 량            |

Test 데이터에는 전력사용량 (Target)과 일조, 일사 데이터가 없음

## 2. 데이터 이해

### Submission – 예측 값 제출 파일

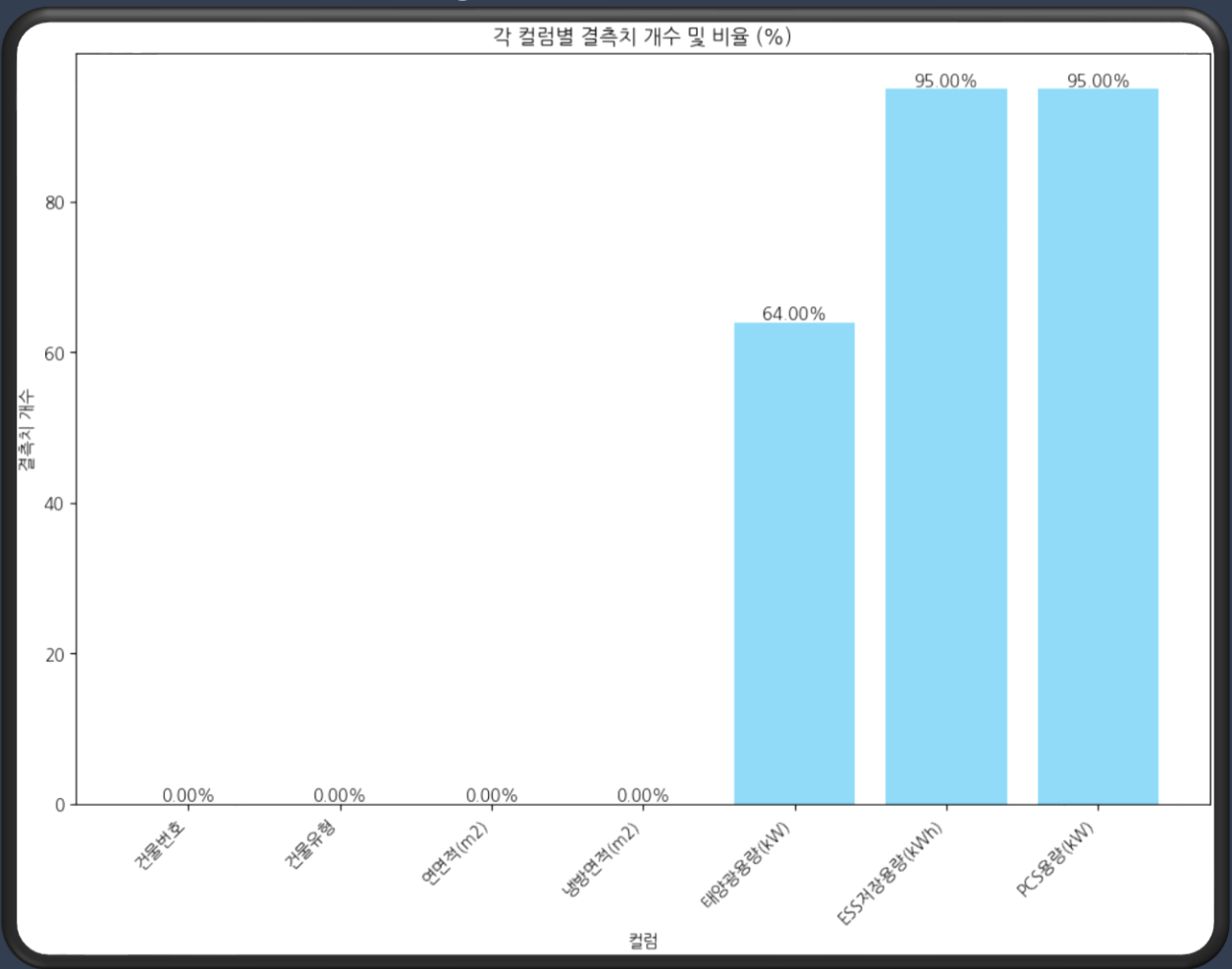
| 변수명           | 데이터 Shape              | 데이터 자료형 | 변수 설명           |
|---------------|------------------------|---------|-----------------|
| num_date_time | 16800행 7열<br>(16800,2) | Object  | 건물번호, 기록일시(1시간) |
| answer        |                        | Int64   | 전력사용량 예측 값 작성   |

# Part 3, 데이터 전처리



### 3. 데이터 전처리

Building 데이터 결측치 비율



| 변수      | 처리 전 | 처리 후 | 근거  |
|---------|------|------|---|
| 태양광용량   | '-'  | 0    | 해당 값이 비어 있는 이유<br>가 태양광 시설이 없기 때<br>문에 관련 컬럼값을 모두 0<br>값으로 처리 |
| ESS저장용량 | '-'  | 0    |   |
| PCS용량   | '-'  | 0    |   |

세 변수의 데이터 타입을 **FLOAT64**로 변경

### 3. 데이터 전처리

냉방면적이 0인 행

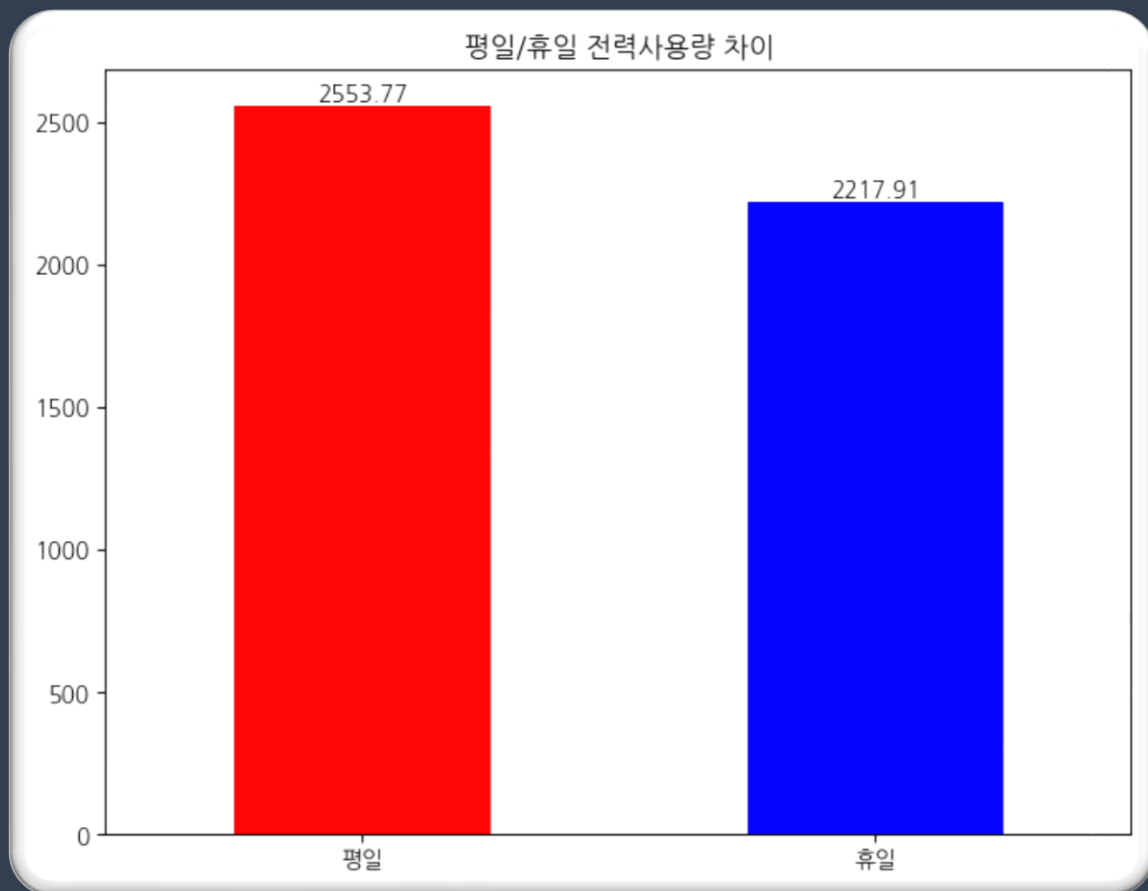
| 건물번호 | 건물유형 | 연면적        | 냉방면적(전) | 냉방면적(후) |
|------|------|------------|---------|---------|
| 64   | 아파트  | 183839.000 | 0       | 14475   |
| 65   | 아파트  | 105073.000 | 0       | 82735   |
| 67   | 아파트  | 389395.928 | 0       | 306611  |

추정된 냉방면적(m²)=연면적(m²)×평균 냉방면적 비율

데이터 상의 일관성, 기존 데이터의 패턴, 추정의 편의성, 실용적 접근을 근거로 해당 공식을 사용

### 3. 데이터 전처리

#### 파생변수 생성



#### 1 휴일

평일과 주말의 유의미한 전력사용량 차이가 존재

#### 2 년/월/일/시

기존의 'num\_date\_time' 변수로 시간을 제대로 설명하기 어려워 해당 변수 생성

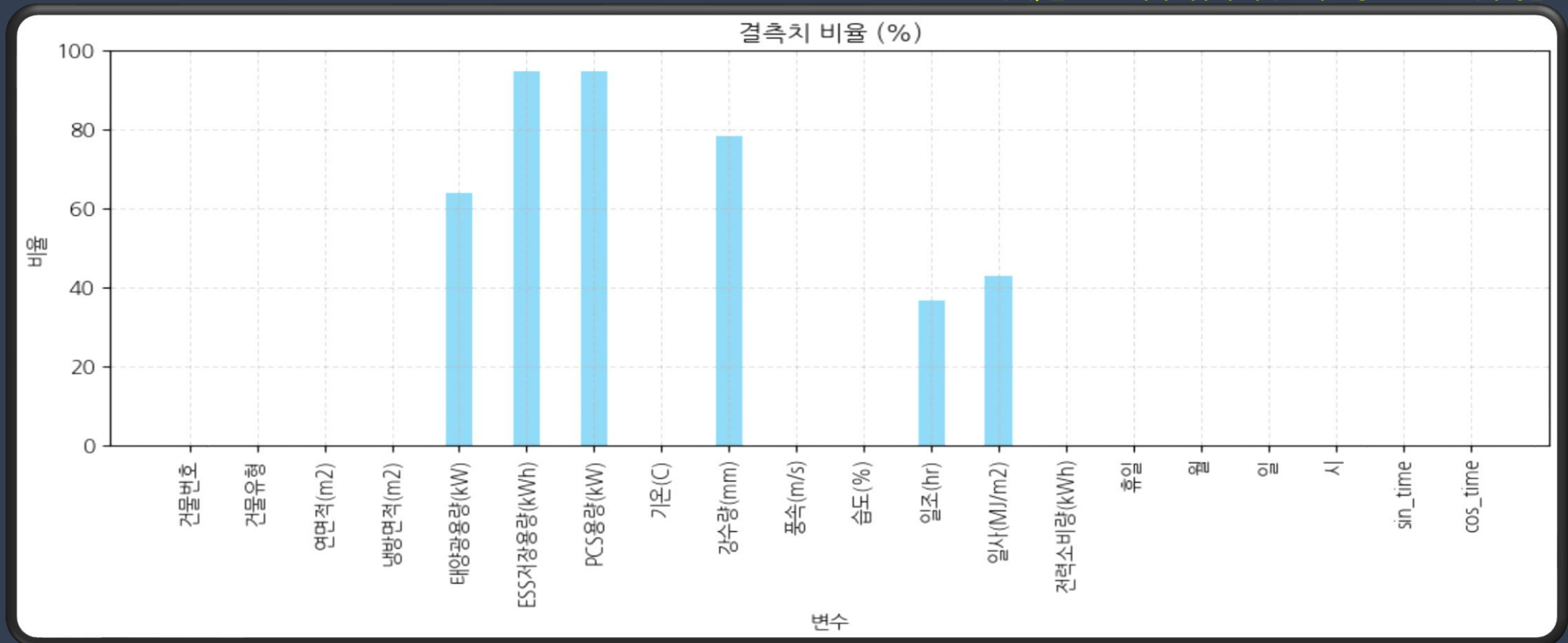
#### 3 cos\_time, sin\_time

시간의 주기성을 표현하기 위해 생성



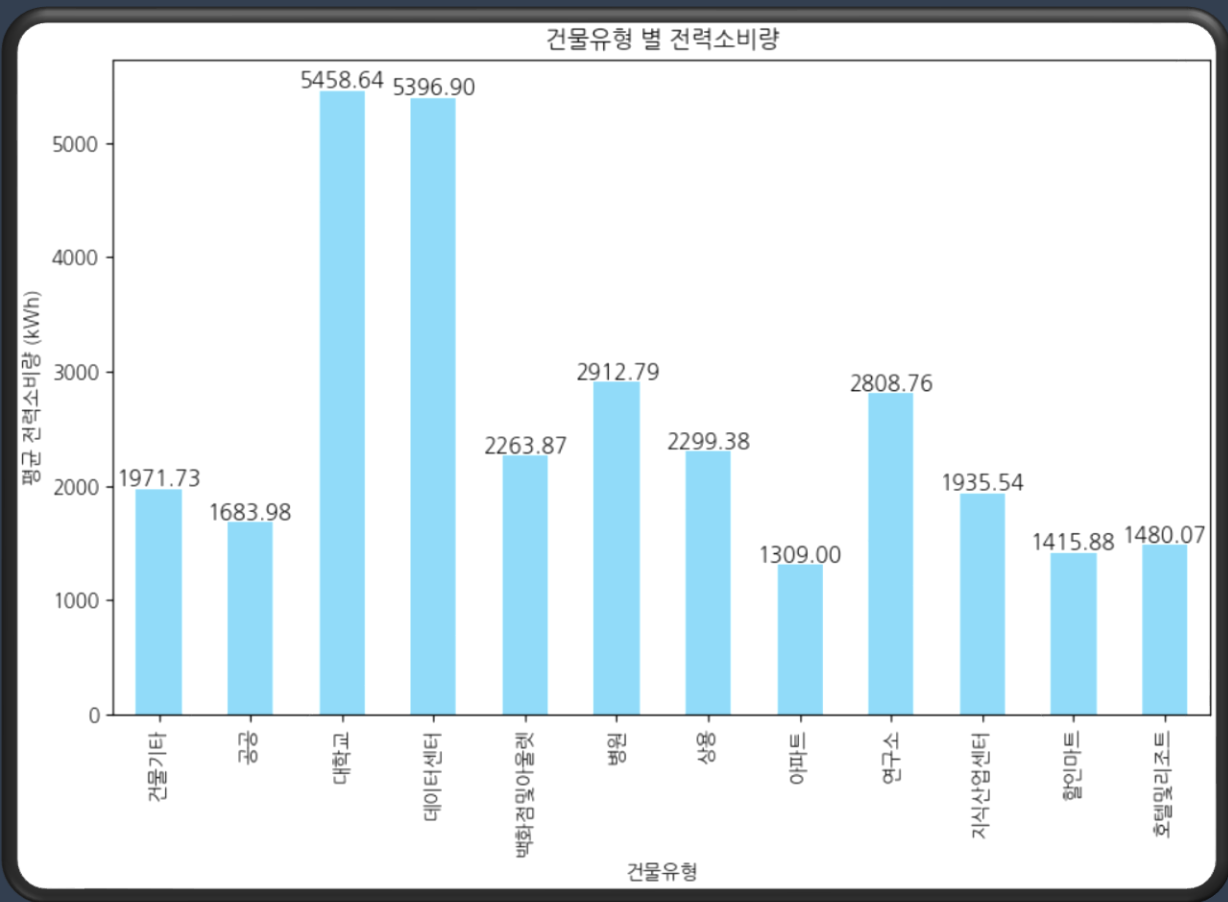
# 3. 데이터 전처리

\* 풍속, 습도는 결측치 개수가 적어 그래프 상으로 잘 보이지 않음



태양광용량, ESS저장용량, PCS용량, 강수량, 풍속, 습도, 일조, 일사에 결측치가 있음을 확인가능

### 3. 데이터 전처리



#### 건물 유형별 전력소비량 차이가 유의미함

- 따라서 건물 유형별로 모델링이 필요함
- Building / train을 건물번호를 기준으로 병합
- 편의성을 위해 건물유형별 숫자 할당

### 3. 데이터 전처리



기상 데이터  
(강수량, 풍속, 습도, 일조, 일사)

#### 선형보간법 (Linear Interpolation)

기상청에서 결측치 처리를 위해 사용하는 방식

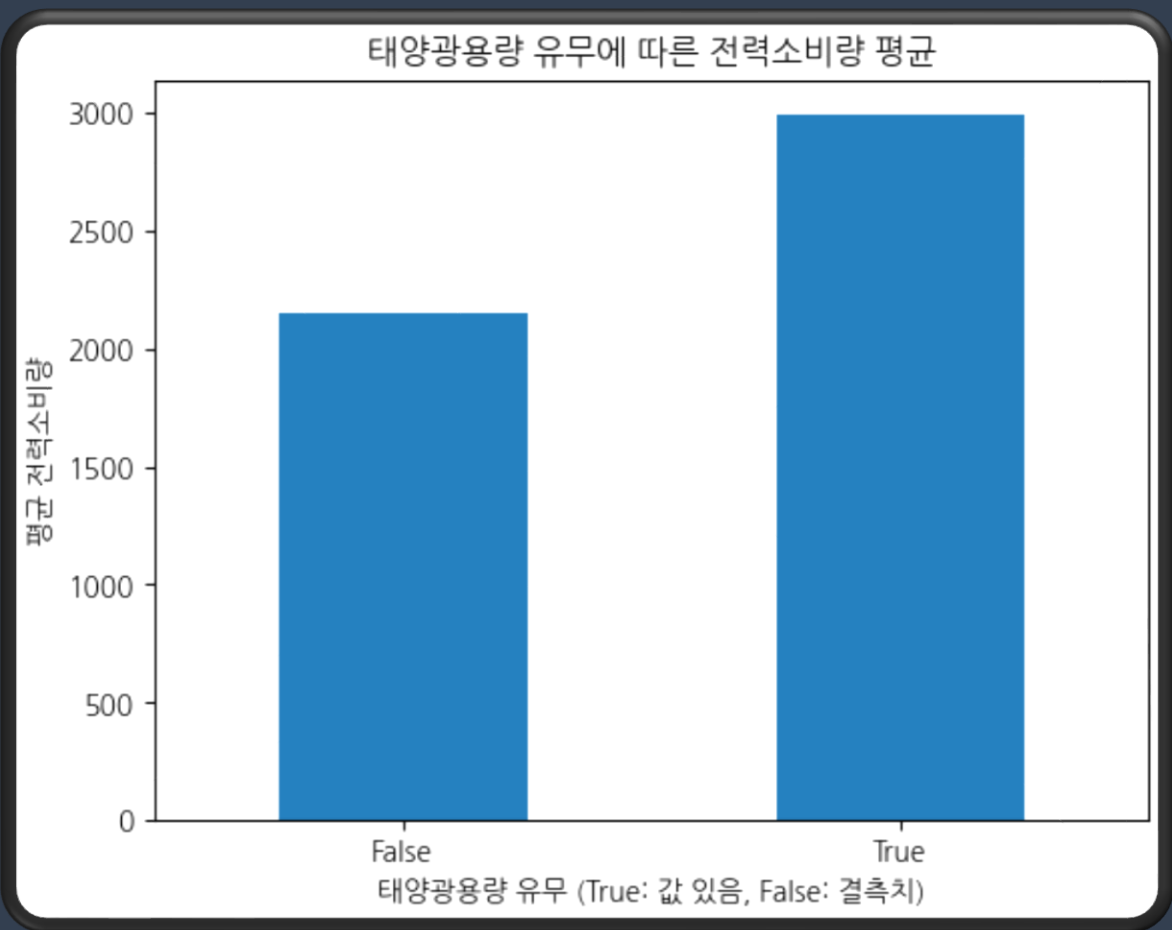


WHY?

- 1. 연속성 유지
- 2. 시간적 또는 공간적 경향 보존
- 3. 단순하고 빠른 계산

### 3. 데이터 전처리

#### 태양광용량 결측치 처리 방식



태양광용량 有: 2150.252 kWh

태양광용량 無: 2985.764 kWh

➔ 유의미한 차이가 있다고 판단하여  
해당 컬럼값을 유(1), 무(0)로 변경

### 3. 데이터 전처리



삭제 피처: ESS저장용량 / PCS용량 / 강수량 / 일조 / 일사

-> 해당 피처의 결측치 비율이 너무 높고 유의미한 활용방안이 없어 삭제

### 3. 데이터 전처리

| 파생변수 | 수식  | 사용근거   |
|------|---|--|
| 불쾌지수 | $1.8 * \text{기온} - 0.55 * (1 - \text{습도} / 100) * (1.8 * \text{습도} - 26) + 32$                  | 1. 소비자 행동의 예측<br>2. 에너지 소비와의 관련성<br>3. 기상정보를 반영하는 추가정보 |
| 체감온도 | $13.12 + 0.6125 * \text{기온} - 11.37 * \text{풍속}^{0.16} + 0.3965 * \text{풍속}^{0.16} * \text{기온}$ |  |
| 냉방효율 | 냉방면적 / 연면적  | 1. 냉방 시스템의 효율성 반영<br>2. 건물 에너지의 효율성 평가                 |

인간의 체감정보와 에너지 효율과 관련된 파생변수 추가

# Part 4, 모델 설계 및 개발



## 4. 모델 설계 및 개발

### XGBOOST

- 분류와 회귀 문제를 다루는데 주로 사용
- 과적합 방지와 대용량 데이터 처리에 적합

### RANDOM FOREST

- 분류와 회귀, 이상치 탐지에 주로 사용
- 과적합 방지 및 복잡한 데이터 패턴에서 높은 예측 성능 제공

건물 유형별로 각 모델과 파라미터를 모두 적용해보고 더 성능이 좋은 모델을 선택하는 방식으로 진행



# 4. 모델 설계 및 개발

## 모델 개발 Flow

### 건물 유형 선택

- 01. 건물기타
- 02. 공공
- 03. 대학교
- 04. 데이터센터
- 05. 백화점및아울렛
- 06. 병원
- 07. 상용
- 08. 아파트
- 09. 연구소
- 10. 지식산업센터
- 11. 할인마트
- 12. 호텔및리조트

### 독립변수 선정

- 1. 건물 유형별 특징 파악
- 2. 전력사용량과의 상관계수 파악
- 3. SMAPE값에 따른 변수 조정

### 모델 선정

- 1. XGBOOST
- 2. RandomForest

### 하이퍼 파라미터 튜닝

- 1. 기본 하이퍼파라미터로 설정
- 2. RandomSearchCV 사용
- 3. 직접 하나씩 파라미터 조절

### SMAPE 값 확인

- SMAPE 값 결과 확인 후 더 낮은 SMAPE 값을 만들기 위해 해당 과정을 반복
- 1. 독립변수 변경해보기
  - 2. 하이퍼파라미터 수정하기

# 4. 모델 설계 및 개발

최종 선정 모델 및 SMAPE 값

| 건물번호 | 건물 종류   | 사용 모델 | SMAPE 값            |
|------|---------|-------|--------------------|
| 1    | 건물기타    | XGB   | 5.845610239860386  |
| 2    | 공공      | XGB   | 3.477446474690761  |
| 3    | 대학교     | XGB   | 2.727848316531297  |
| 4    | 데이터센터   | RF    | 0.3978170909762858 |
| 5    | 백화점및아울렛 | XGB   | 4.982441398697709  |
| 6    | 병원      | XGB   | 2.394588832652312  |
| 7    | 상용      | XGB   | 4.57162870010022   |
| 8    | 아파트     | XGB   | 3.1930089134368074 |
| 9    | 연구소     | XGB   | 3.011617611164601  |
| 10   | 지식산업센터  | XGB   | 2.8901377332781206 |
| 11   | 할인마트    | XGB   | 4.953595982613571  |
| 12   | 호텔및리조트  | XGB   | 5.281170631370898  |



최종 점수: 7.85

최종 등수: 113등 / 1200명