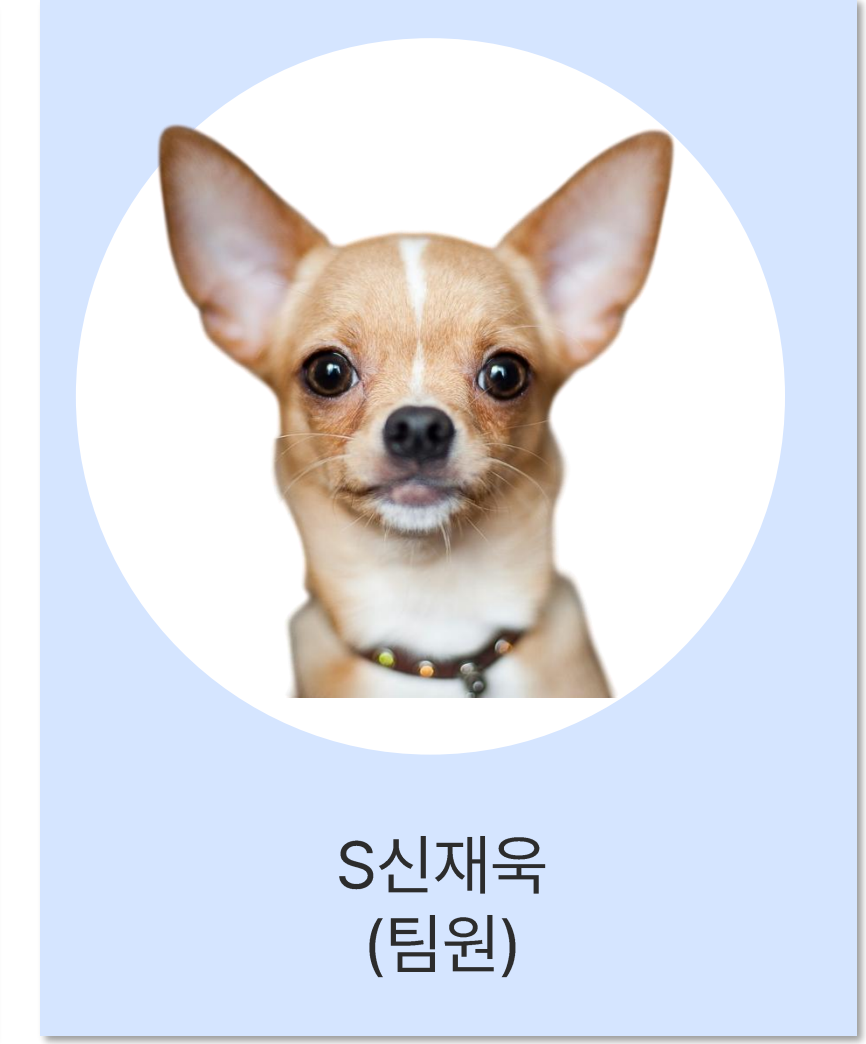
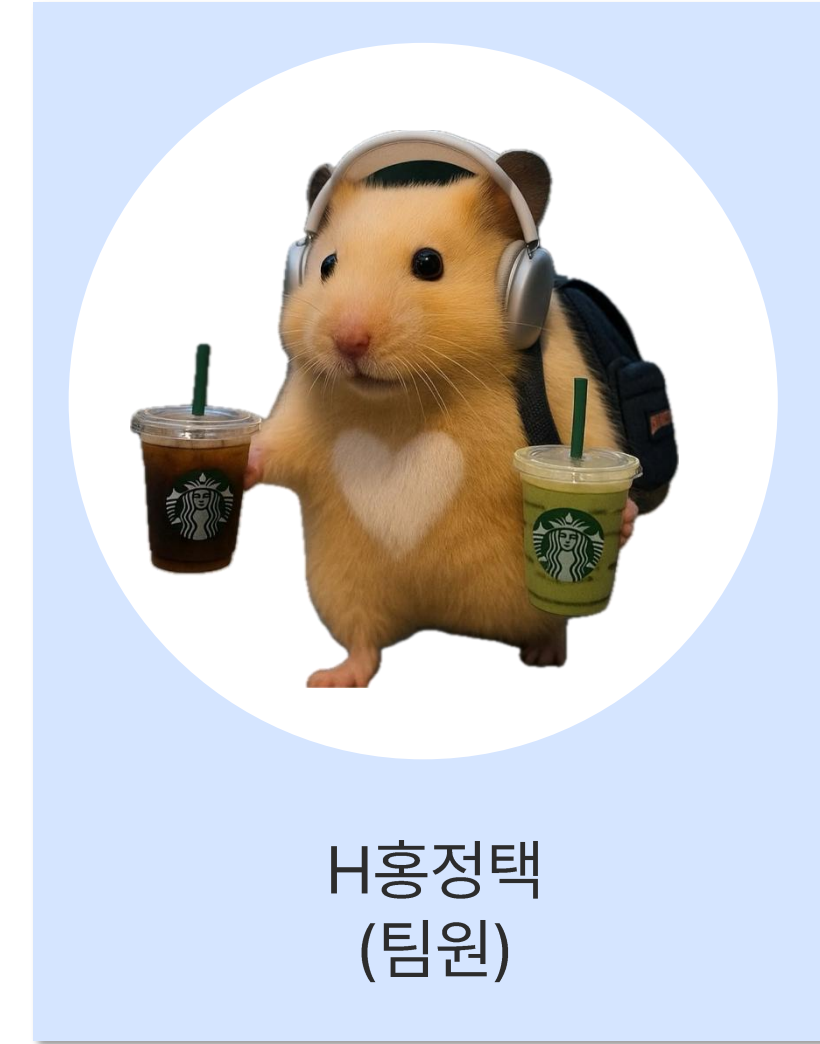
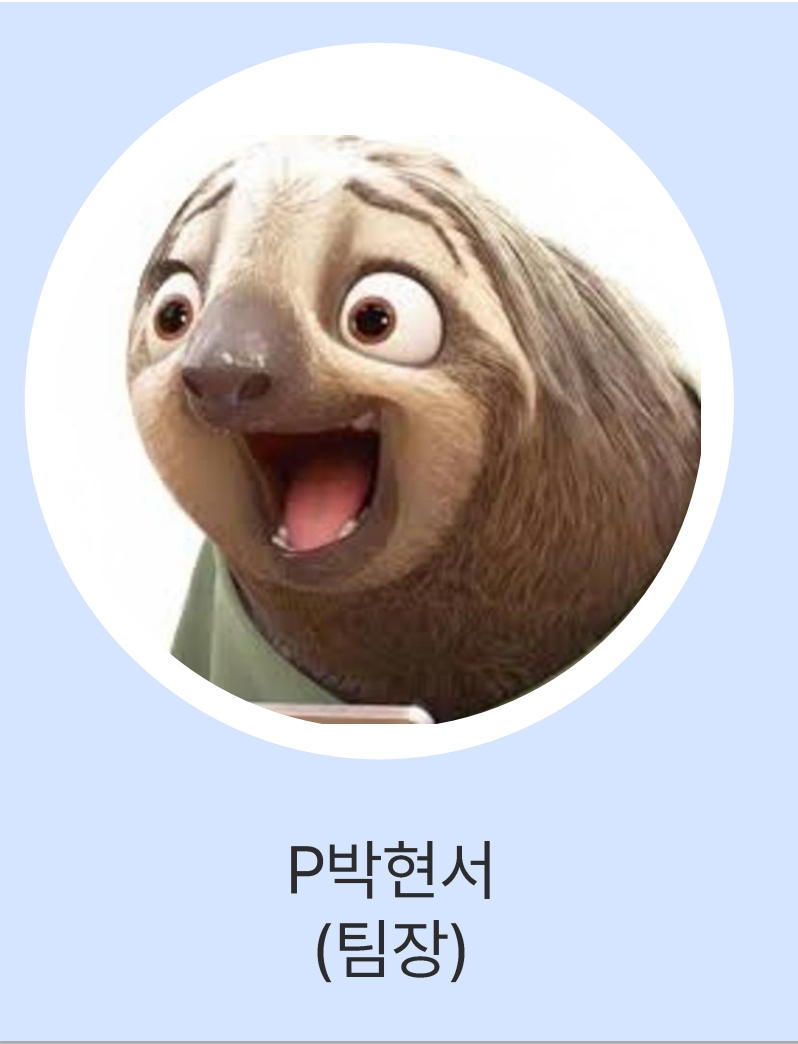


LightGBM 기반 철강 공장 전력 소비량 예측 모델 개발 및 고도화

팀 소개



"누구에게나 예측가능한 솔루션을
제공합니다."



- P** 데이터 기반 예측(Predictive)과
- H** 전력 관련 요소를 통합한 하이브리드 분석(Hybrid)을 통해
- S** 산업 시설의 전력 사용을 사전에 예측하고
최적의 의사결정을 지원하는 솔루션(Solution)을 제공합니다.

CONTENTS

01

프로젝트 배경 및 목표 (KPI)

- 전기료 인상
- 에너지 절감, 비용 절감 설정

02

프로젝트 진행 계획

- 시간대별, 요일별 가동 패턴
- 전력 사용량과 에너지 효율
(역률) 간의 상관관계 분석

03

데이터 분석 & 모델링

- Data Preprocessing
- Exploratory Data
Analysis (EDA)
- Feature Engineering
 - ML Modeling
 - Application

04

결론 및 기대효과

- 에너지 절감, 비용 절감 확인
- 전력 사용량 패턴 정밀 분석

TEAM
PROJECT

PART 1. 프로젝트 배경 및 목표(KPI)

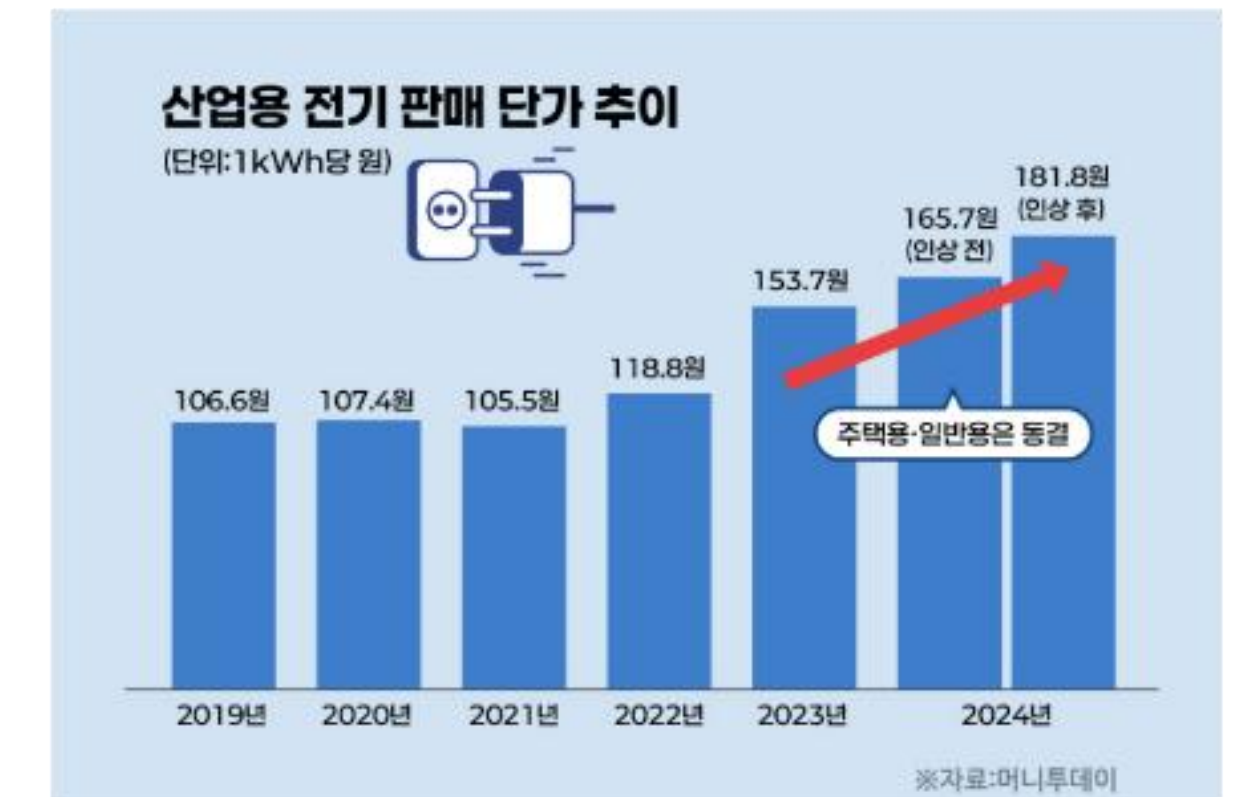
추진 배경

웹크롤링을 활용한 산업 현황 및 문제점



현대제철 '전기료 감면' 빠진 K스틸법에 원가 부담 지속, 서강현 관세 부담 겹쳐 해외 생산체제 가속

신재희 기자 JaeheeShin@businesspost.co.kr | 2025-12-01 16:03:59

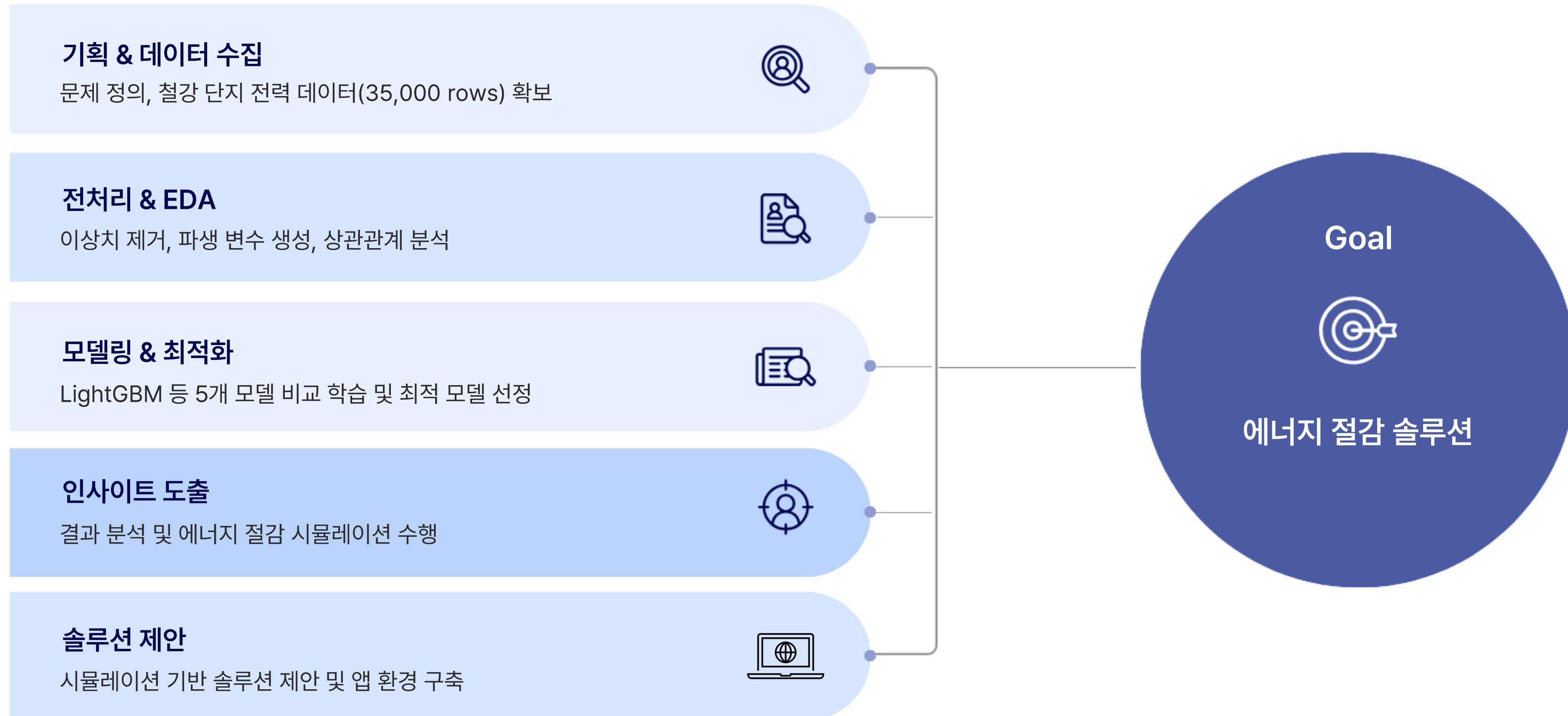


KPI

TEAM
PROJECT

PART 2. 프로젝트 진행 계획

프로젝트 진행 계획



TEAM
PROJECT

PART 3. 데이터 분석 (Data Analysis)

Data Preprocessing

Exploratory Data
Analysis (EDA)

Feature Engineering

ML Modeling

3-1. Data Preprocessing

Dataset 확인

	date	Usage_kWh	Lagging_Current_Reactive.Power_kVarh	Leading_Current_Reactive_Power_kVarh	CO2(tCO2)	Lagging_Current_Power_Factor	Leading_Current_Power_Factor	NSM	WeekStatus	Day_of_week	Load_Type
0	01/01/2018 00:15	3.17	2.95	0.0	0.0	73.21	100.0	900	Weekday	Monday	Light_Load
1	01/01/2018 00:30	4.00	4.46	0.0	0.0	66.77	100.0	1800	Weekday	Monday	Light_Load
2	01/01/2018 00:45	3.24	3.28	0.0	0.0	70.28	100.0	2700	Weekday	Monday	Light_Load
3	01/01/2018 01:00	3.31	3.56	0.0	0.0	68.09	100.0	3600	Weekday	Monday	Light_Load
4	01/01/2018 01:15	3.82	4.50	0.0	0.0	64.72	100.0	4500	Weekday	Monday	Light_Load

'date' 컬럼 dtype 변환

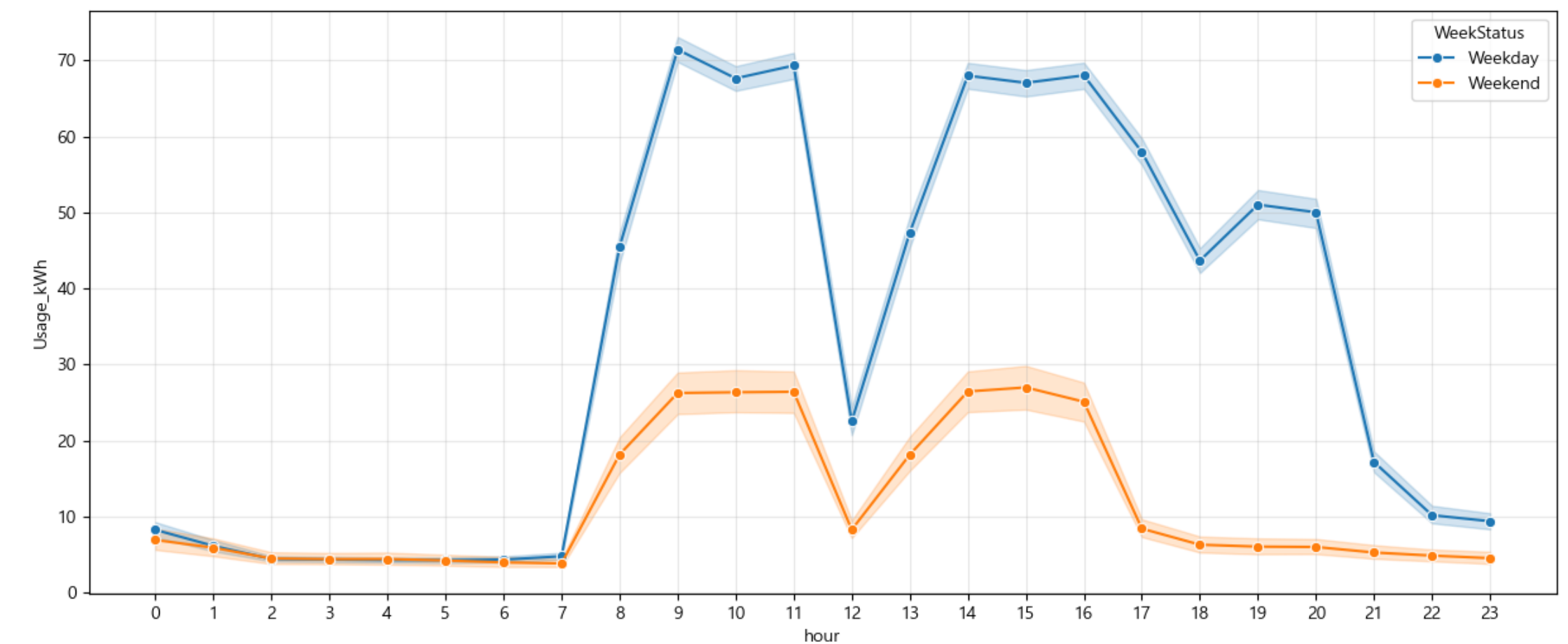
```
# 'date' 컬럼을 문자열에서 날짜/시간(datetime) 타입으로 변환
df['date'] = pd.to_datetime(df['date'], dayfirst=True)
# 시간(0~23시), 월(1~12월), 요일(0~6, 월~일) 정보를 새로운 컬럼으로 만들기
df['hour'] = df['date'].dt.hour
df['month'] = df['date'].dt.month
df['dayofweek'] = df['date'].dt.dayofweek
```

3-2. Exploratory Data Analysis

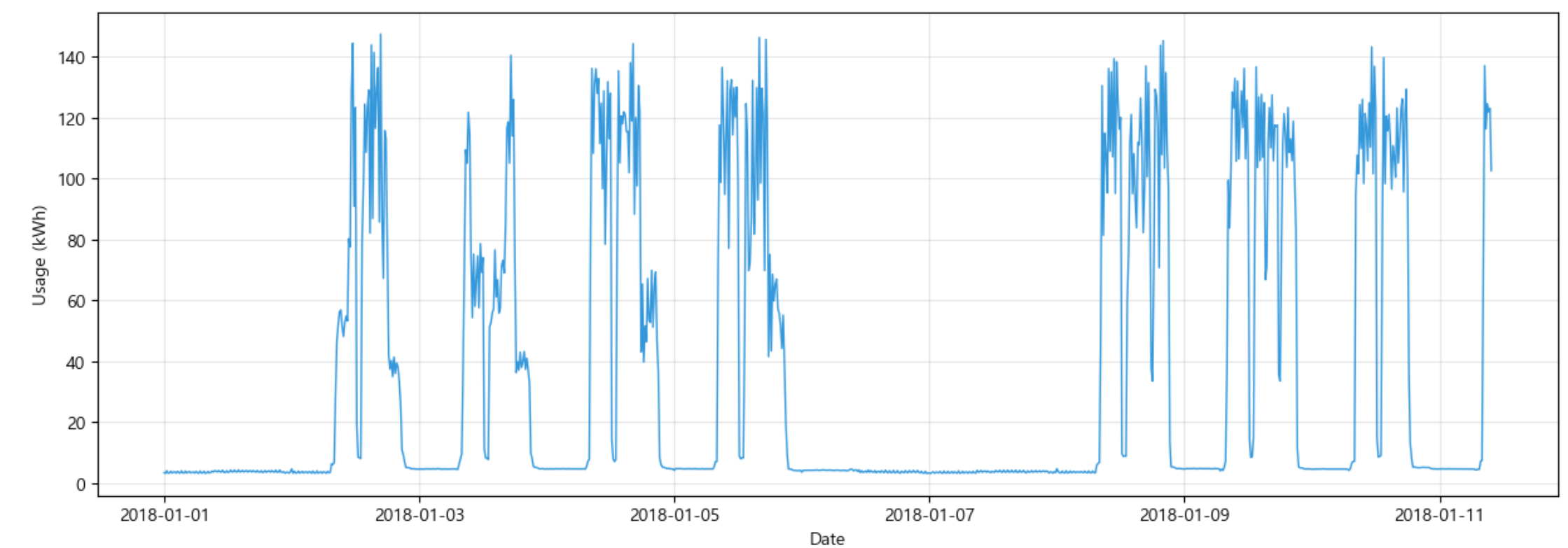
데이터 간 상관관계 지도 (Heatmap)



시간대별 전력 사용 패턴 (평일 vs 주말)

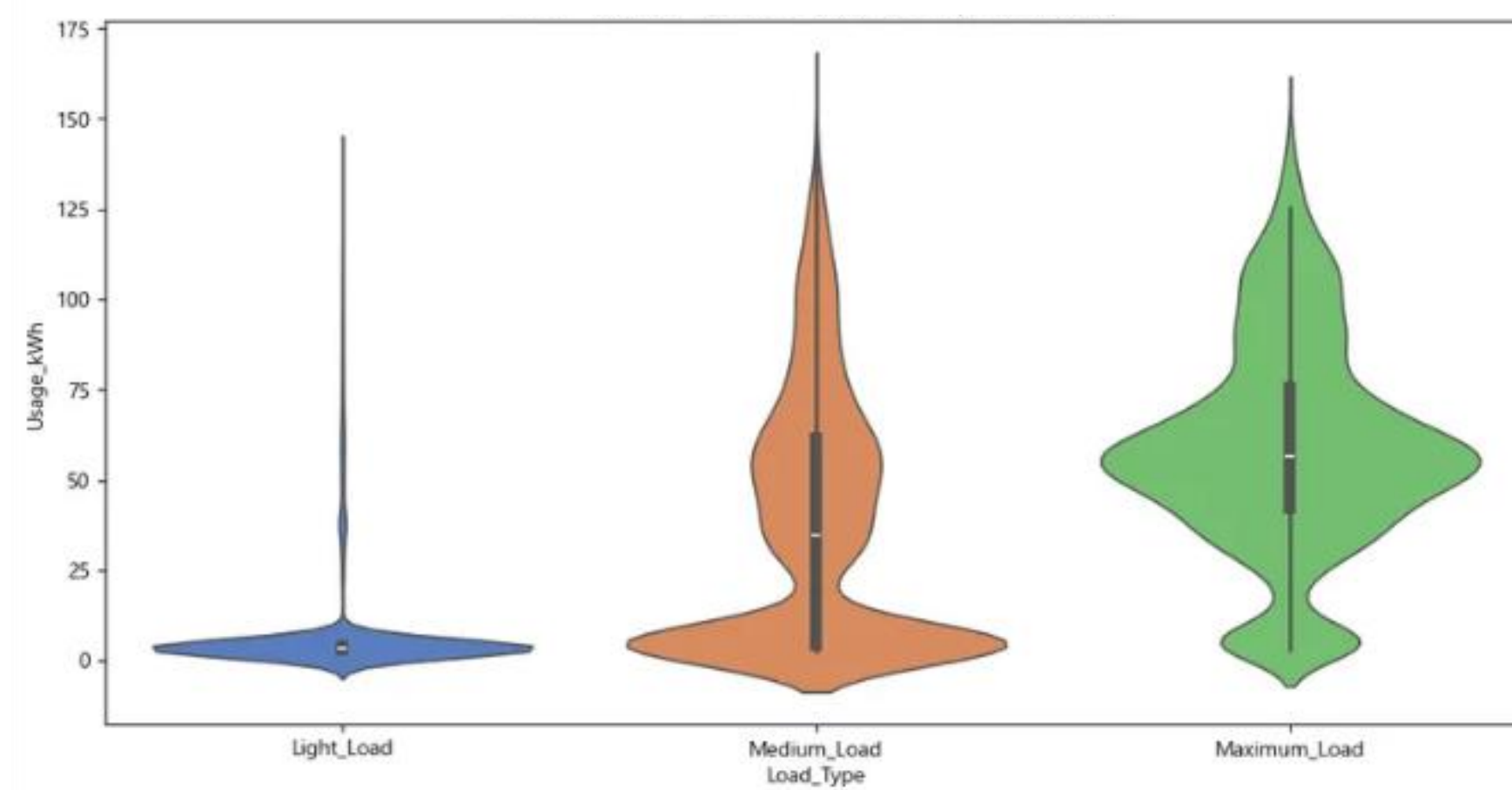


초기 전력량 사용 추이 (Sample : 10 days)

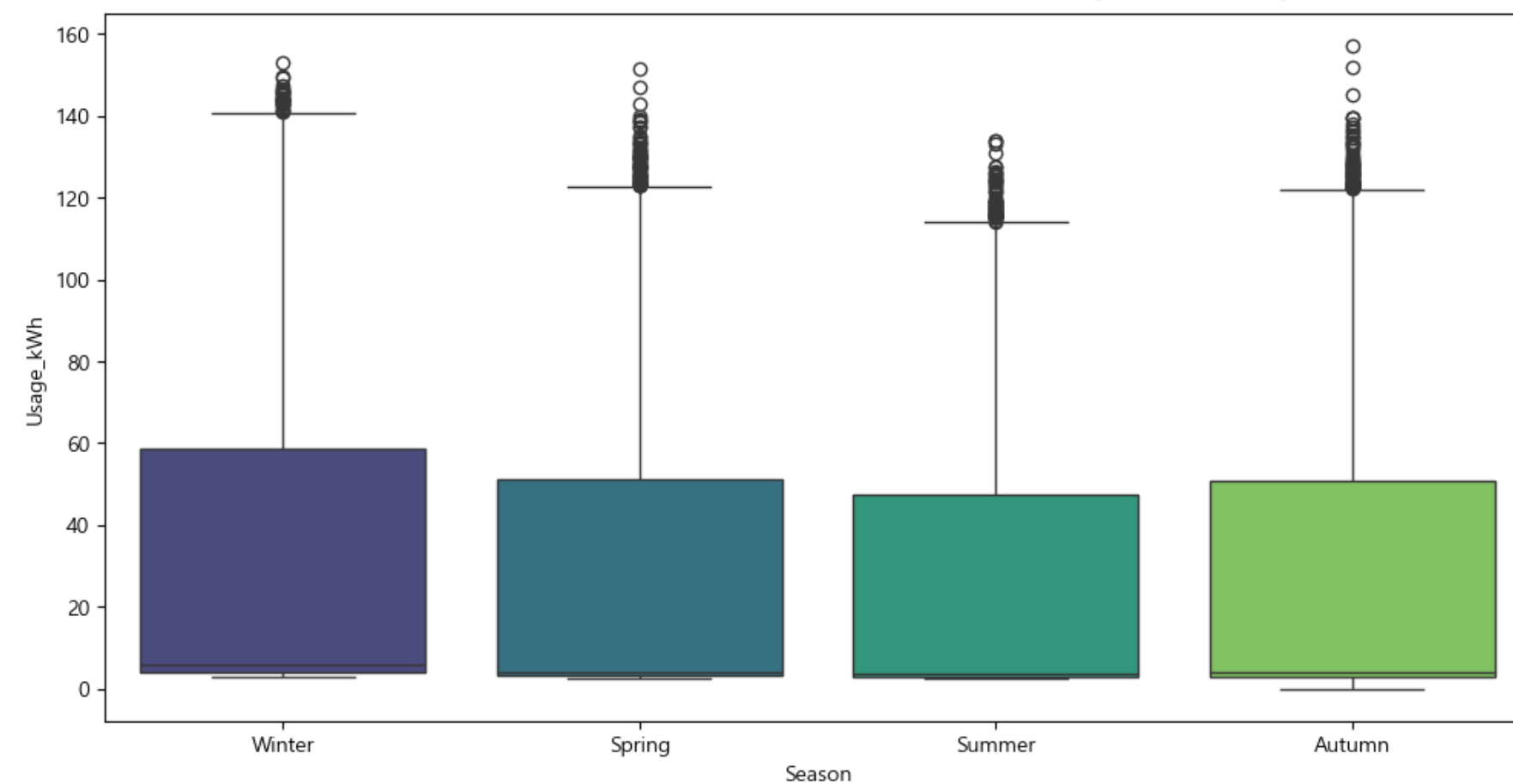


3-2. Exploratory Data Analysis

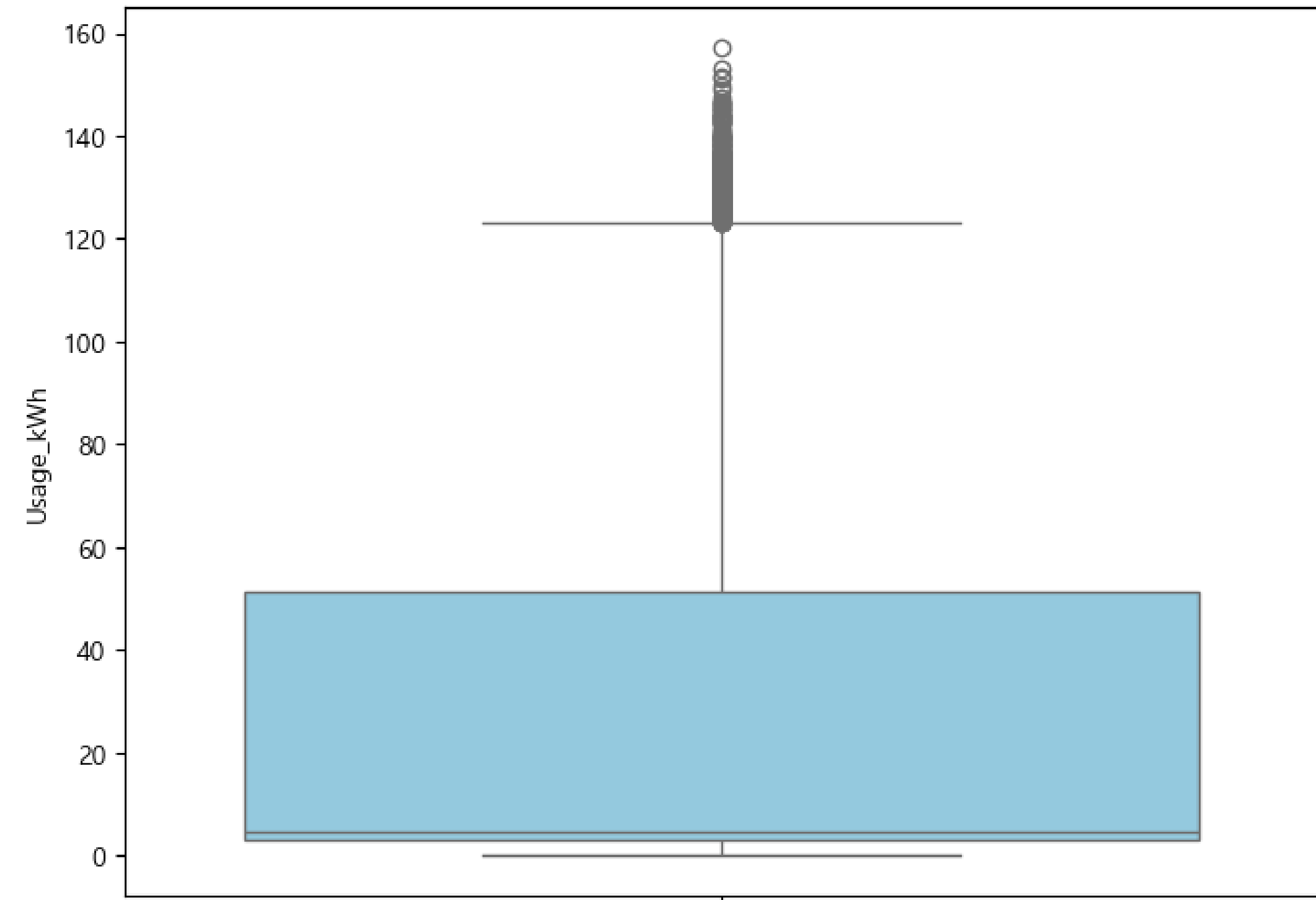
부하 상태별 전력 사용량 분포(Violin Plot)



계절별 전력 사용량 분포 (Seasonality Analysis)



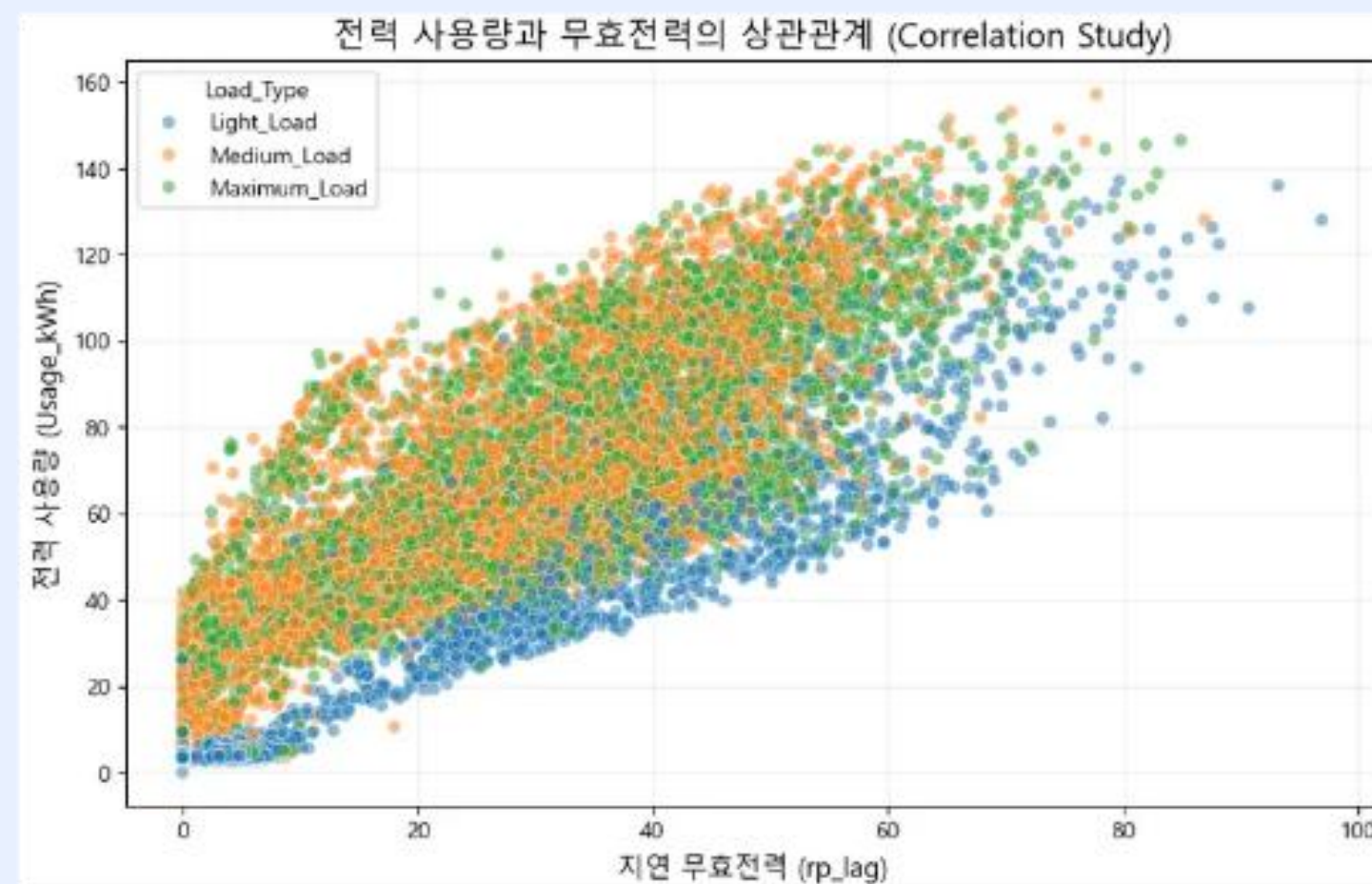
전력 사용량 이상치 확인(Box Plot)



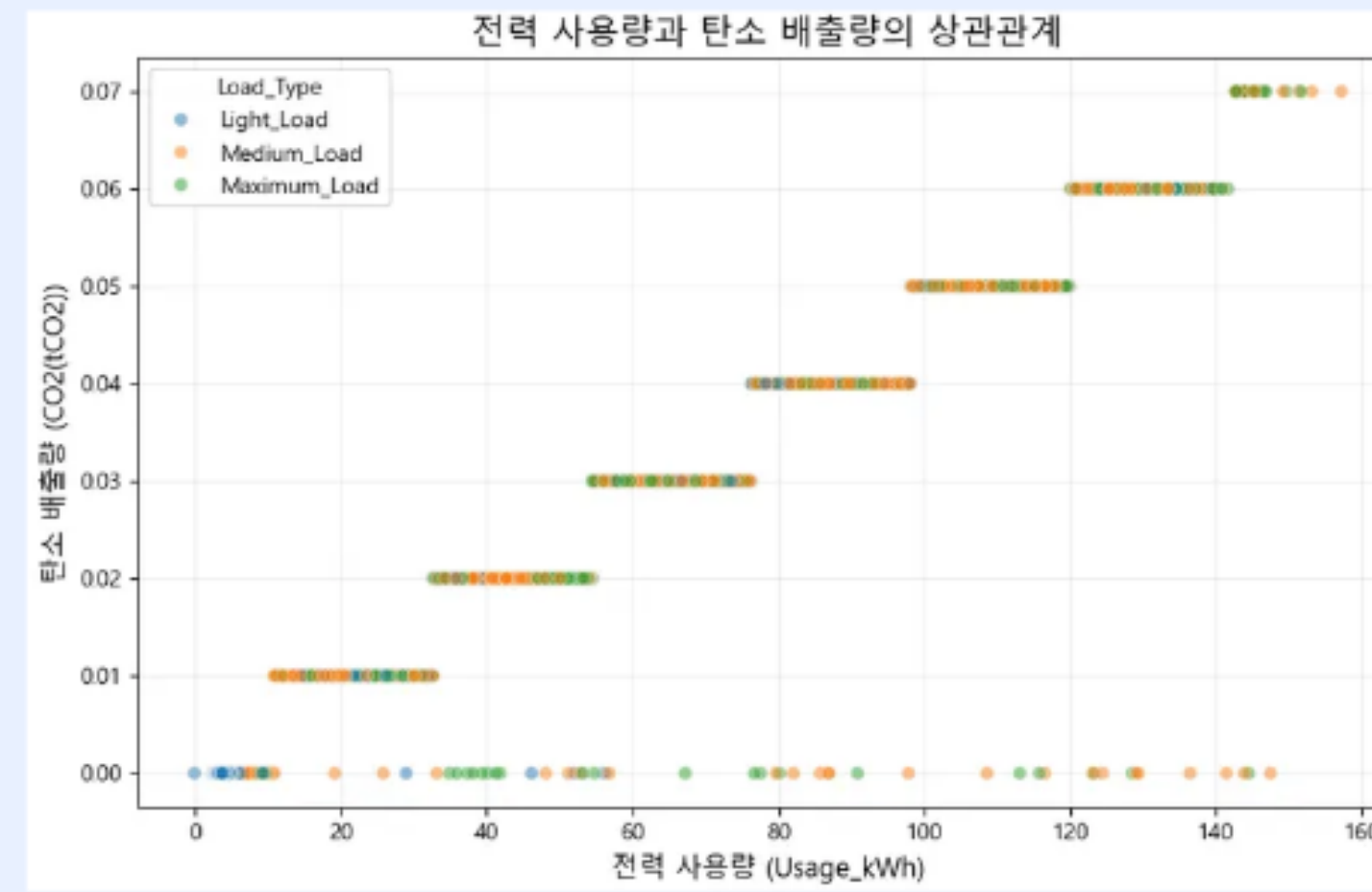
분석 결과 이상치는 측정 오류가 아닌,
실제 철강 공정 가동 중 발생하는 '피크 전력'으로 판단

3-2. Exploratory Data Analysis

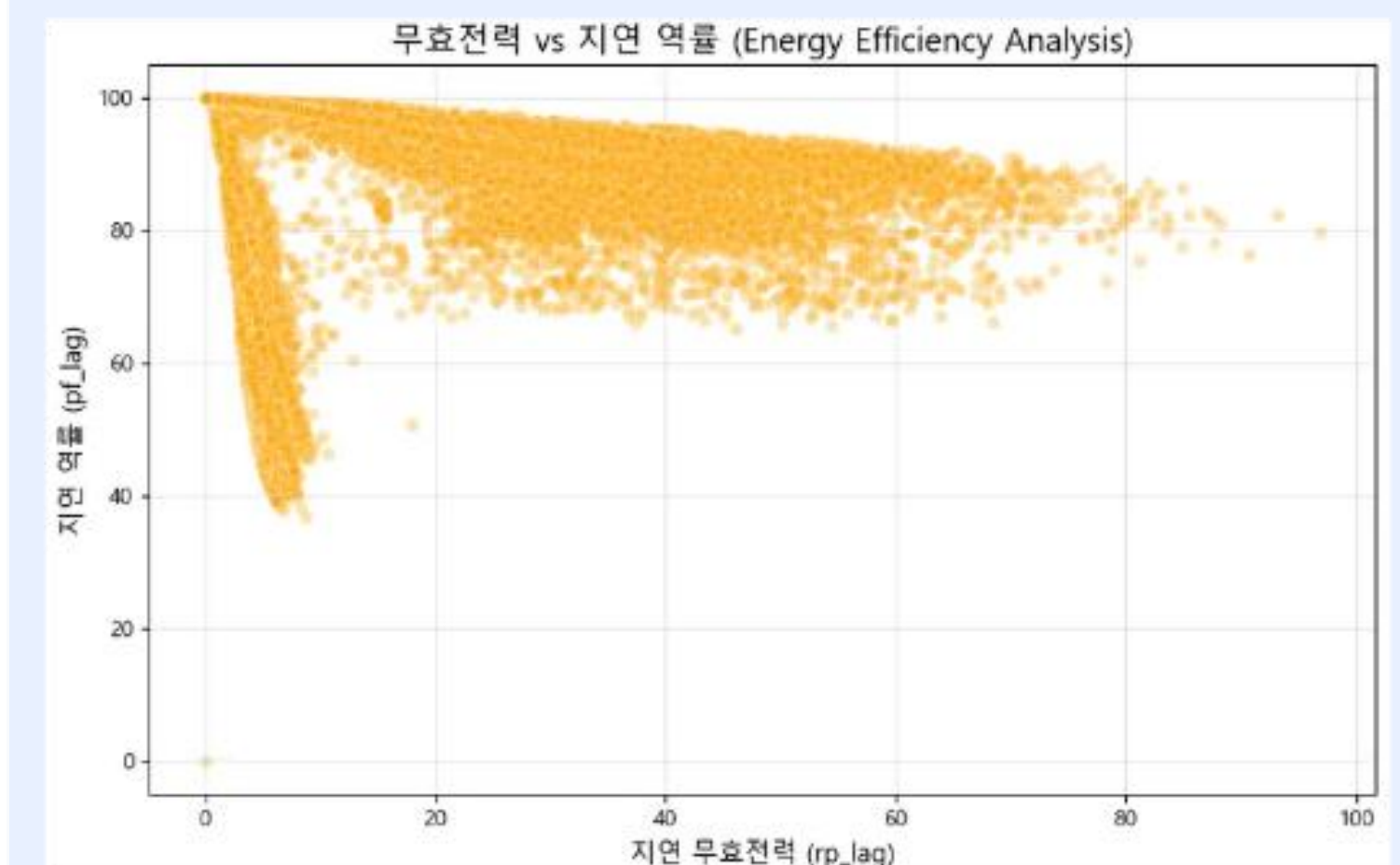
- Correlation Study



전력 사용량과 무효전력의 상관관계



전력 사용량과 탄소 배출량의 상관관계



무효전력 vs 지연 역률
(Energy Efficiency Analysis)

3-3. Feature Engineering

Before

WeekStatus, Day_of_week,
Load_Type, Season, month

After

정수(0, 1, 2...) 형태로 변환

Feature Engineering 코드

```
from sklearn.preprocessing import LabelEncoder

# 1. 글자 데이터를 숫자로 변환 (LabelEncoder)
le = LabelEncoder()

# 분석에 사용할 '운영 관련' 컬럼들만 정의
cols = ['WeekStatus', 'Day_of_week', 'Load_Type', 'Season', 'month']

for col in cols:
    if col in df.columns:
        df[col] = le.fit_transform(df[col])

# 2. 불필요한 변수 제거
# 'Usage_kWh'와 물리적 공식으로 연계있는 컬럼들을 모두 제거. -> 과적합 가능성
drop_cols = [
    'date', 'CO2(tcO2)',
    'rp_lag', 'rp_lead',
    'pf_lag', 'pf_lead',
]

# 실제로 데이터프레임에 존재하는 컬럼만 골라서 제거
existing_drop_cols = [c for c in drop_cols if c in df.columns]
train_df = df.drop(existing_drop_cols, axis=1)

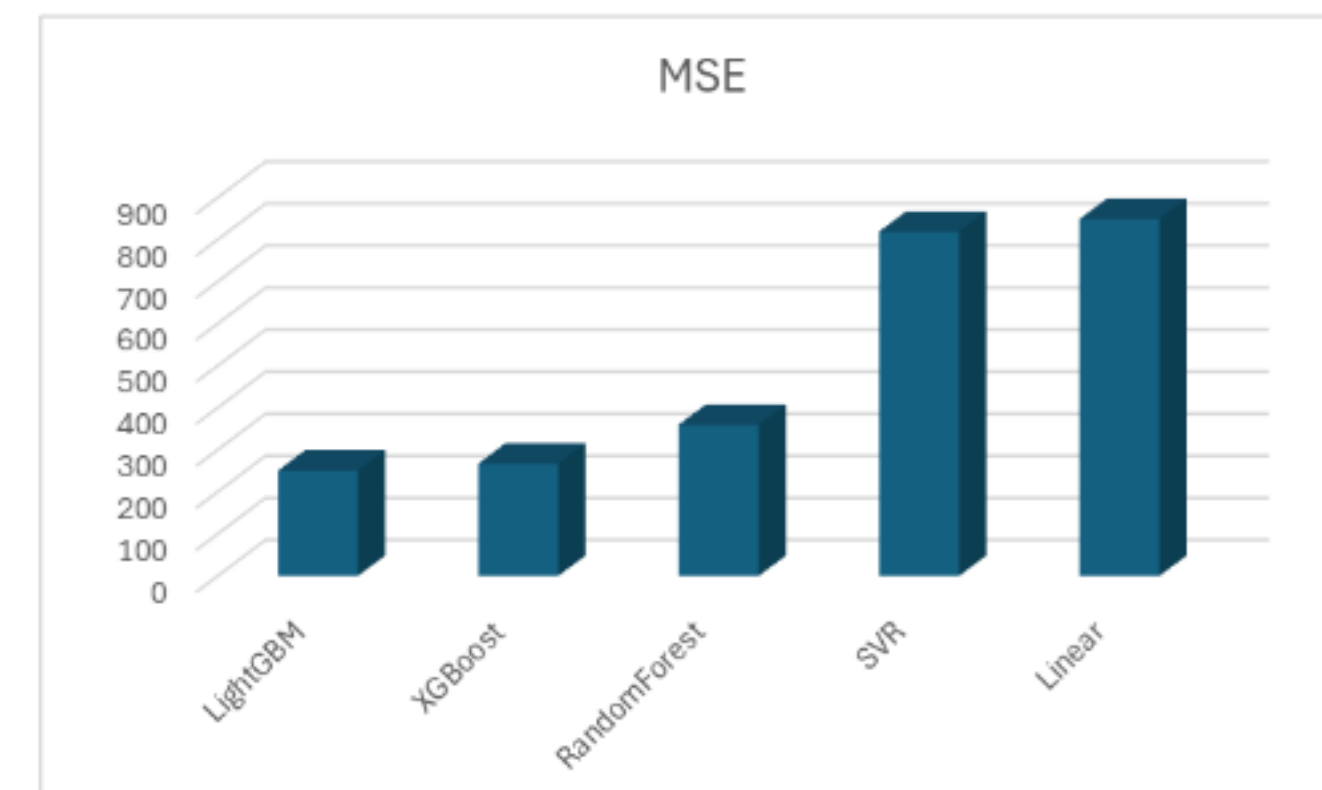
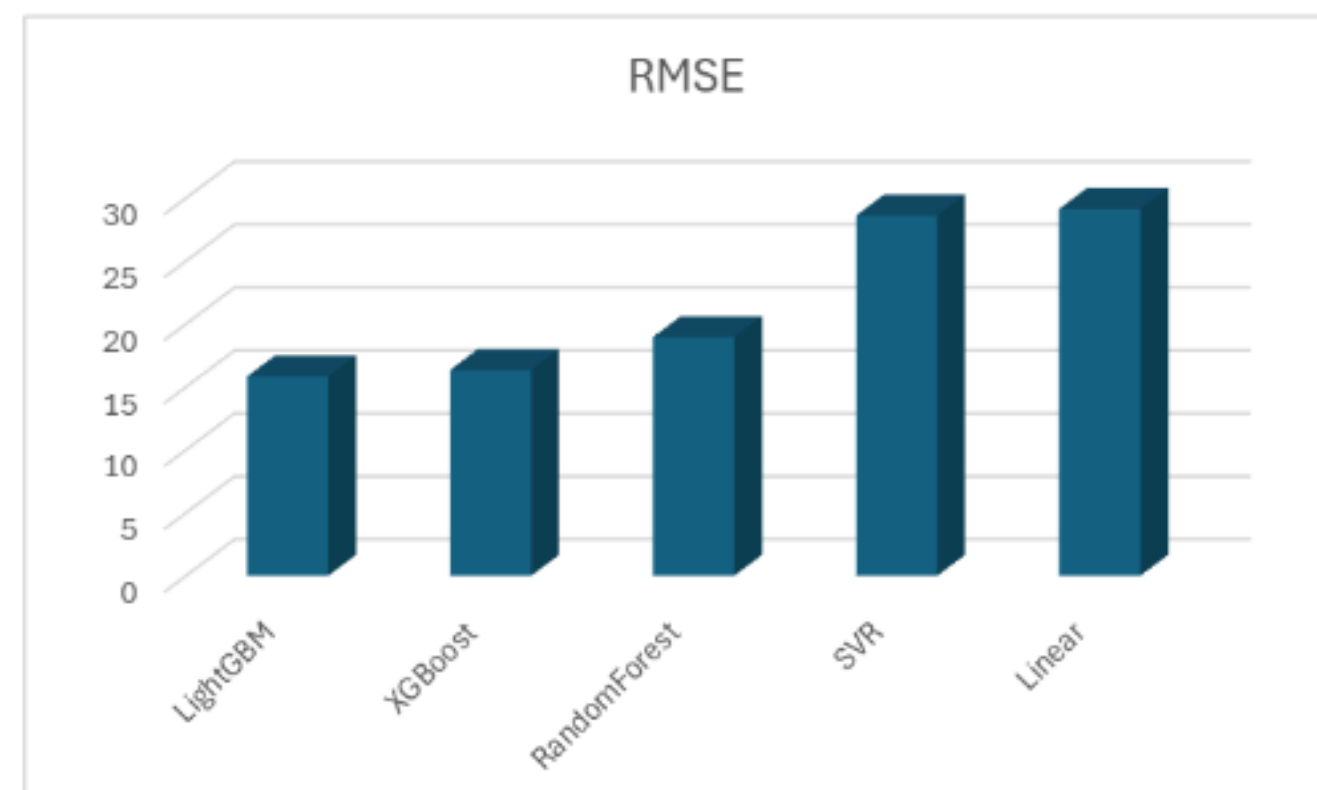
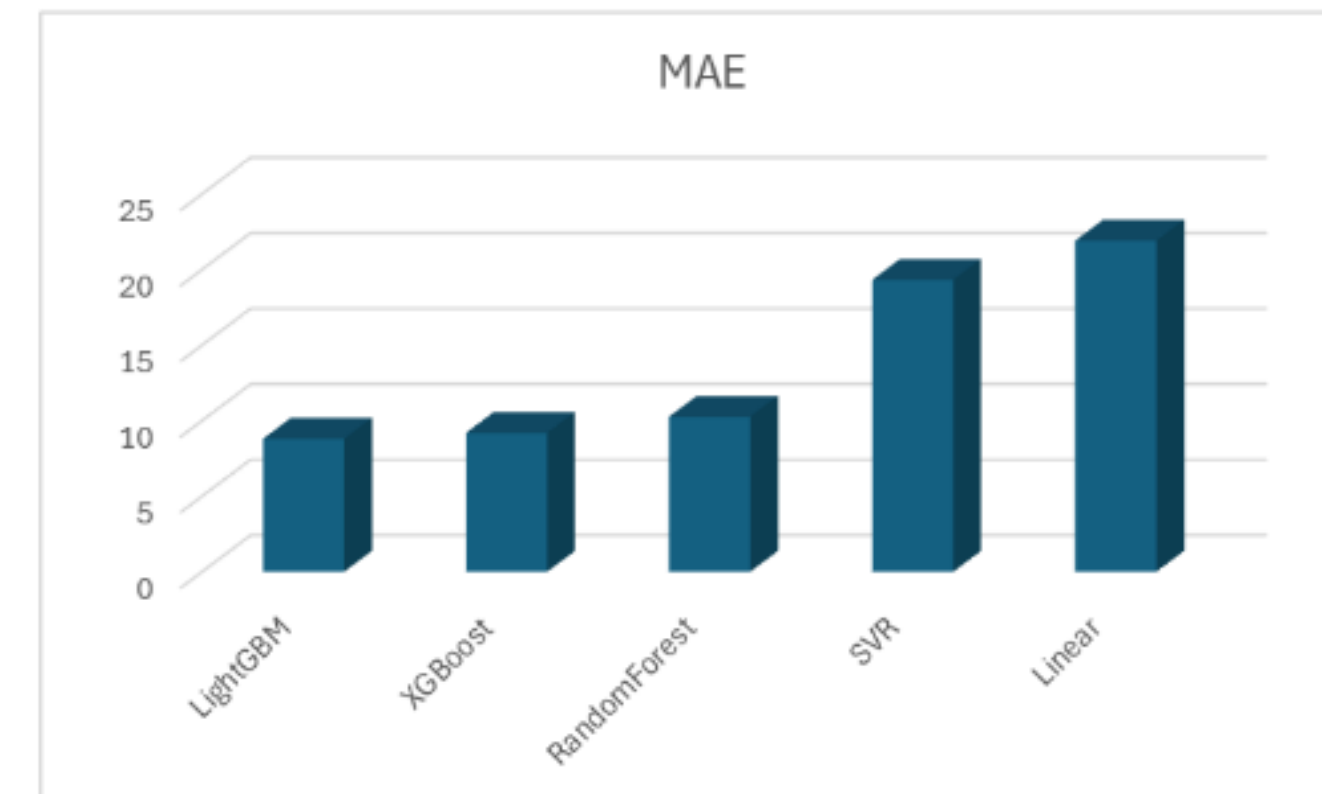
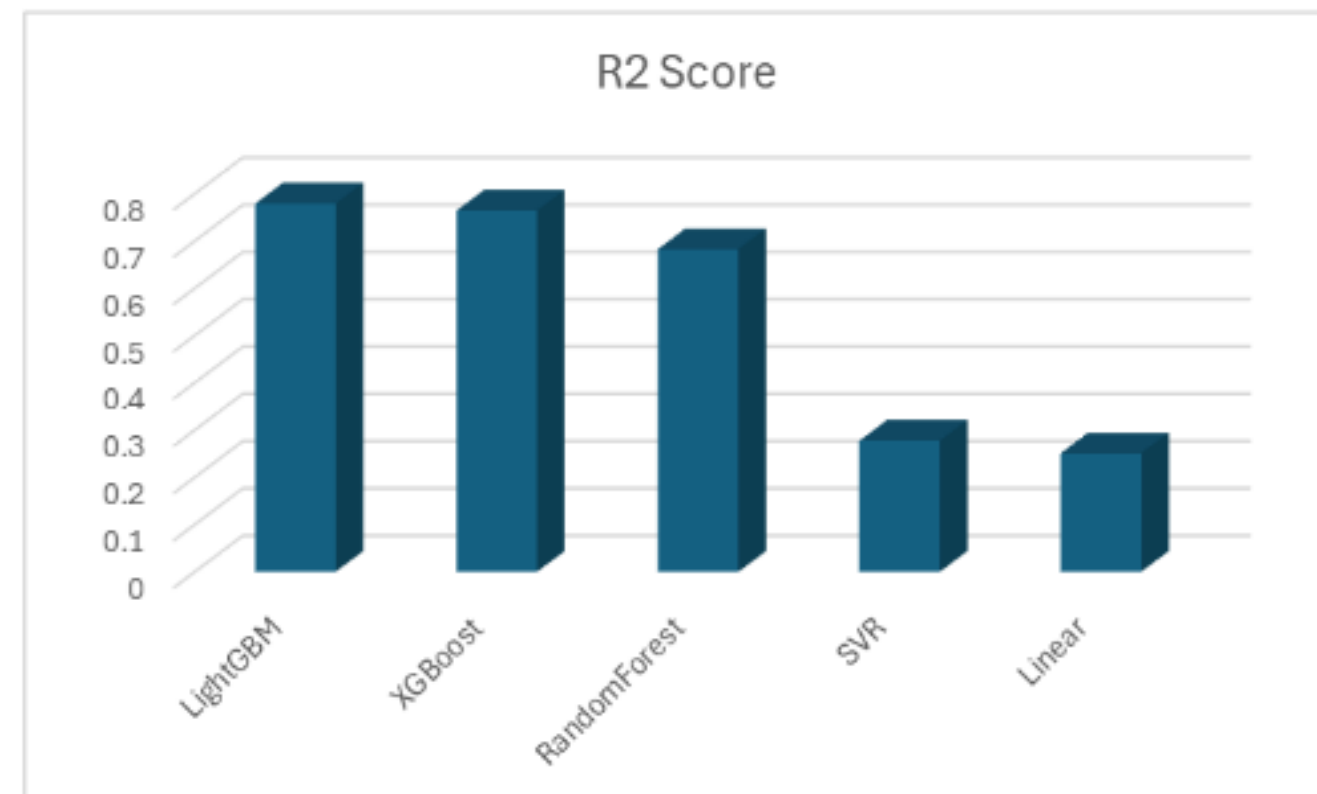
# 3. 최종 확인
print("✅ 전처리 완료 남은 컬럼들:")
print(train_df.columns.tolist())
train_df.head()
```

```
# 2. 불필요한 변수 제거
# 'Usage_kWh'와 물리적 공식으로 연계있는 컬럼들을 모두 제거. -> 과적합 가능성
drop_cols = [
    'date', 'CO2(tcO2)',
    'rp_lag', 'rp_lead',
    'pf_lag', 'pf_lead',
]
```

- 문자열인 'date' 제거
- 과적합 방지를 위해 'CO2', 'Reactive Power (rp)', 'Power Factor (pf)' 제거 (전력 사용량의 파생/상관 변수)

	Usage_kWh	NSM	WeekStatus	Day_of_week	Load_Type	hour	month	dayofweek	Season
0	3.42	0	0	1	0	0	0	0	3
1	3.17	900	0	1	0	0	0	0	3
2	4.00	1800	0	1	0	0	0	0	3


3-4. ML Modeling



※ n_estimators=100, random_state=42

3-5. Solution & Application

<<

 운영 조건 설정

현재 시간 (NSM) ?

36000

0 86400

주중/주말

☒ Weekday

☐ Weekend

요일

Monday ▼

부하 상태

Light_Load ▼

계절

Fall ▼

월 (Month)

1

스마트 팩토리 에너지 관리 AI

Deploy ⋮

🔥 PHS | Predictive Hybrid Solution
누구에게나 예측가능한 솔루션을 제공합니다 ;)

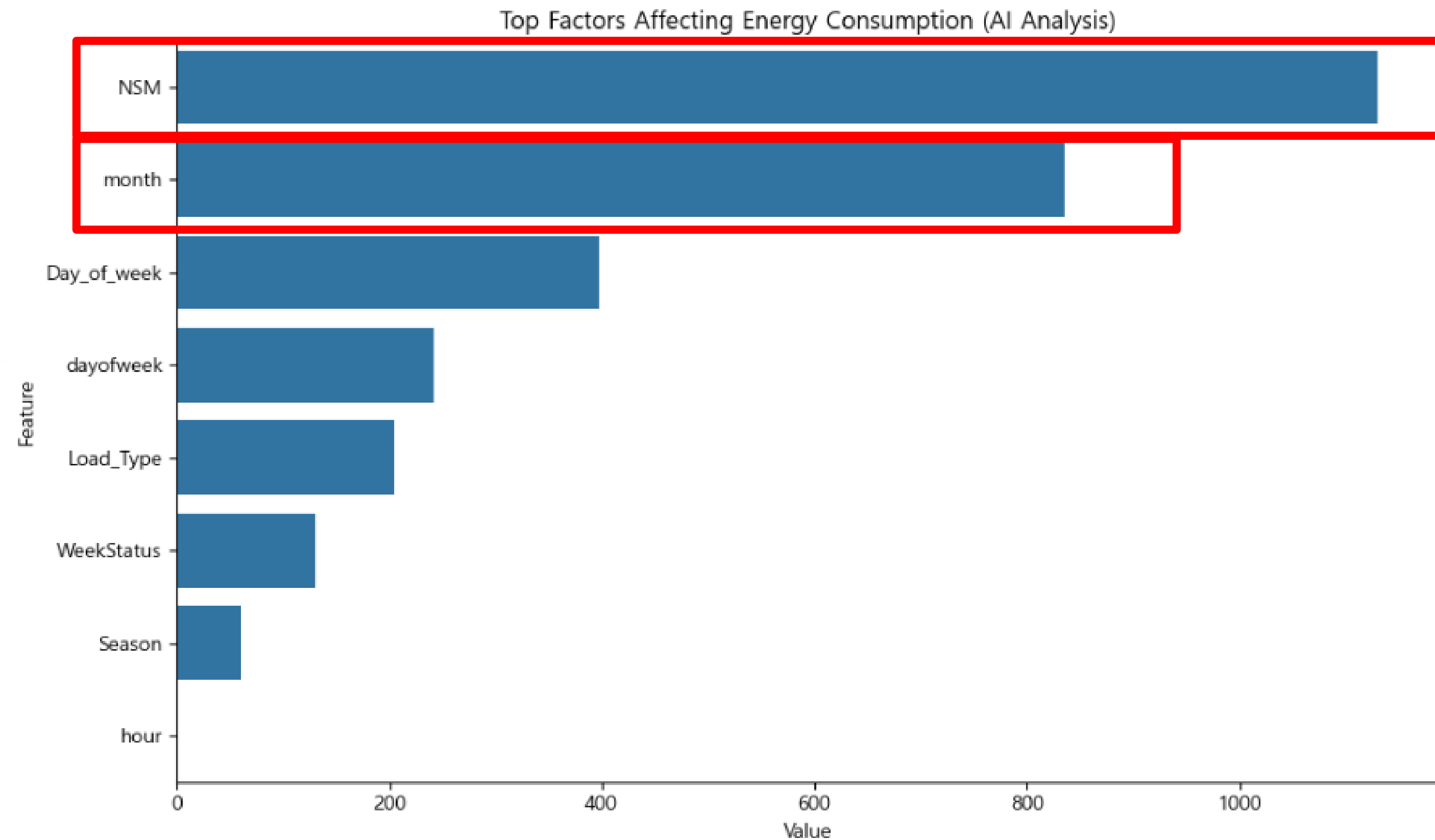
이 대시보드는 운영 변수(시간, 요일, 부하)로 공장의 예상 전력 소비량을 예측합니다.

⚡ 전력 사용량 예측하기

TEAM
PROJECT

PART 4. 결론 및 기대효과

4-1. 결론



Time

피크 시간대
분산 운영
(NSM 기반)

Month

월별 맞춤형
에너지 목표제
(Month 기반)

기대 효과

Summary

예상 절감률:
16.01%



Business Impact

연 **5,895,722원**
정도의 비용절감



Action Plan

전력 사용량 패턴
정밀 분석



Action Plan

데이터 기반 의사결정
체계 수립



감사합니다.

References

김성철, 정호석, 김진해. "국민DR 참여활성화를 위한 머신러닝 기반의 전력수요 예측에 관한 연구." 한국IT정책경영학회 논문지 16.4 (2024): 3711-3715.

성종훈, 조영식. "머신러닝 기법을 활용한 공장 에너지 사용량 데이터 분석." 정보처리학회논문지: 컴퓨터 및 통신시스템(KTCCS), vol. 8, no. 4, Apr. 2019, pp. 87-92. KISS 한국학술정보, kiss.kstudy.com/Detail/Ar?key=3672815. Accessed 20 Jan. 2026.

신재희. "현대제철 '전기료 감면' 빠진 K스틸법에 원가 부담 지속, 서강현 관세 부담 겹쳐 해외 생산체제 가속." Business Post, 1 Dec. 2025, www.businesspost.co.kr/BP?command=article_view&num=422055.

이석주, 다어반권. "머신러닝 기반 공장 HVAC 시스템의 에너지 효율화 운영 시뮬레이션." 한국산업정보학회논문지 29.2 (2024): 47-54, KISTI1.1003/JNL.JAKO202414243237329

"[대담한 대담] 3편. 철강산업 미래 경쟁력, 에너지 전략에 달렸다!" 포스코그룹 뉴스룸, 7 July 2025, newsroom.posco.com/kr/대담한-대담-3편-철강산업-미래-경쟁력-에너지-전략/.

한국전력공사. "용도업종별 전력사용량." 빅데이터 한전 전력데이터 개방포털, bigdata.kepco.co.kr/cmsmain.do?scode=S01&pcode=000166&pstate=L&redirect=Y. Accessed Day Month Year."

팀 구성 및 역할

이름	역할	담당 업무
박현서	팀장	데이터 전처리, EDA, 웹크롤링, 발표 자료 제작
홍정택	팀원	EDA, 머신러닝 모델링 및 튜닝, 어플리케이션 구현
신재욱	팀원	EDA, 발표 자료 제작, 최종 발표

기술 스택 & 환경

	이름	선정 이유
언어	Python 3.x	Python은 데이터 분석 및 머신러닝 분야에서 가장 널리 활용되는 언어로, 풍부한 라이브러리와 높은 가독성을 갖추고 있기에 본 프로젝트에 적합
라이브러리	Pandas	시계열 기반 전력 사용 데이터를 정제·가공하고, 결측치 처리 및 feature engineering을 수행하는 데 적합한 라이브러리
	Matplotlib, Seaborn	전력 사용량의 시간대별 변화, 분포 특성, 변수 간 상관관계를 시각적으로 표현하는 데 사용, 분석 결과를 효과적으로 전달
	Scikit-learn	데이터 분할, 전처리 파이프라인 구성, 회귀 및 예측 모델 구현 등 머신러닝 분석의 전반적인 과정에서 안정적이고 표준화된 기능을 제공하기에 선정
	XGBoost, LightGBM	비선형 관계를 효과적으로 학습할 수 있는 모델로, 전력 사용량과 같이 복잡한 요인에 의해 변동하는 시계열 데이터를 예측하는 데 높은 성능을 보임

프로젝트 세부 일정

마일스톤	기간	세부 설명
사전 기획	1/16(금)~1/17(토)	프로젝트 기획 및 주제 선정, 기획안 작성
데이터 전처리 및 EDA	1/16(금)~1/17(토)	필요 데이터 수집, 데이터 정제 및 정규화, 데이터 분석
웹크롤링 및 분석	1/17(금)~1/19(토)	키워드 웹크롤링 진행, 동향 분석
머신러닝 모델 성능 비교	1/19(금)~1/20(토)	모델 선정, 모델링 및 튜닝
최종 보고서	1/17(금)~1/21(토)	최종 보고서 작성
발표 자료	1/17(금)~1/21(토)	발표 자료 제작
프로젝트 총 기간	1/16(금)~1/22(토) (총 7일)	--