

# 제 주 지 역 전 력 사 용 량 예 측 모 델 구 축

고려대학교  
민윤기 노연수 임채명





# | 전력사용량 예측의 필요성

제9차  
전력수급기본계획  
(2020~2034)

에너지전환 패러다임 변화: 환경성, 안전성을 종합적으로 고려, **재생에너지 확대**  
'40년 발전비중 30~35%에 맞춰 '34년까지 지속 확충

**기온 변동성 확대**가 전력수요에 미치는 영향 고려:  
전력수요 전망에 있어서 최근 30년간 기온실적을 활용하여 분석하고, 기온민감도 상향 조정

에너지전환 과정에서 **안정적 전력공급** 유지 필요

산업통상자원부, 2020. 12. 28.

제5차  
신 · 재생에너지  
기술개발 및  
이용·보급  
기본계획

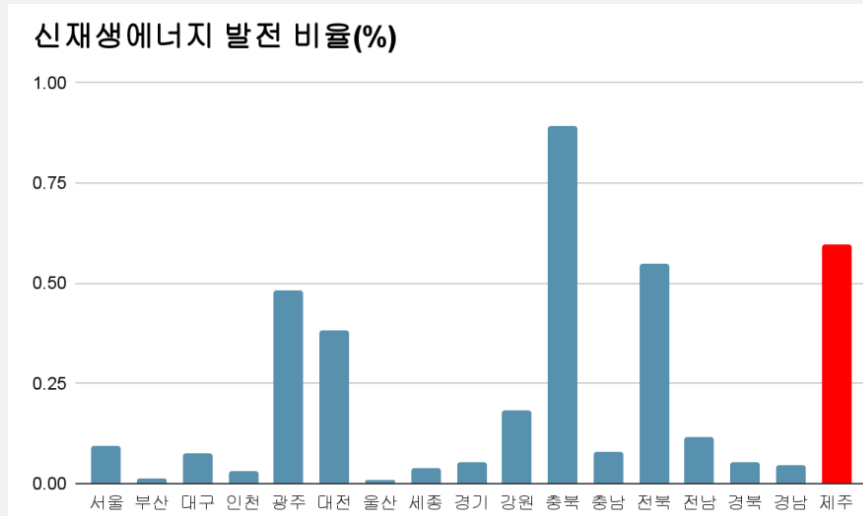
기본방향: 국민의 삶의 질을 높이는 ‘참여형 에너지체제’로 전환하고,  
신재생에너지 확산을 에너지신산업 육성 기회로 적극 활용

주요국은 탄소중립 등 기후변화 대응과 경기부양을 동시에 달성할수 있는 핵심수단으로 재생에너지를 적극 육성중

산업통상자원부, 2020. 12. 29.

## | 제주도에서의 독립적 전력 예측의 필요성

### 높은 신재생 에너지 보급률과 이에 따른 불균형 현상



2022년 제주 지역의 신재생에너지 발전량은 총 발전량의 60% 수준, 충북을 제외한 전국 17개 시도 중에서 높은 신재생에너지 발전 비율을 보임

-> 신재생에너지 보급 확대에 따른 전력 불균형 현상

### 육지와는 다른 특성

신재생에너지 발전: 기후의 영향을 크게 받음

-> 제주 지역에서 전체적인 전력 수급의 불확실성이 높음

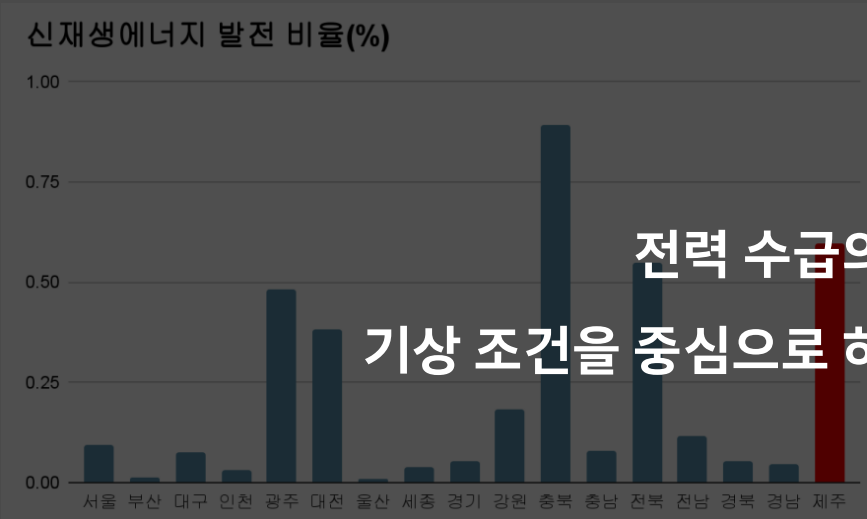
육지와 달리 산업단지나 대규모 기업 등이 없음

제주 지역의 전력수요 피크시간대: 연중 18~21시

육지의 전력수요 피크시간대: 전국의 경우, 여름과 겨울 각  
각 10~11시와 17시 전후

# 제주도에서의 독립적 전력 예측의 필요성

## 높은 신재생 에너지 보급률과 이에 따른 불균형 현상



전력 수급의 안정성을 유지하기 위해

기상 조건을 중심으로 하는 전력 사용량 예측 알고리즘을 제안

2022년 제주 지역의 신재생에너지 발전량은 총 발전량의 60% 수준, 충북을 제외한 전국 17개 시도 중에서 높은 신재생에너지 발전 비율을 보임  
-> 신재생에너지 보급 확대에 따른 전력 불균형 현상

## 육지와는 다른 특성

신재생에너지 발전: 기후의 영향을 크게 받음  
제주 지역에서 전체적인 전력 수급의 불확실성이 높음  
육지와 달리 산업단지나 대규모 기업 등이 없음  
제주 지역의 전력수요 피크시간대: 연중 18~21시  
육지의 전력수요 피크시간대: 전국의 경우, 여름과 겨울 각  
각 10~11시와 17시 전후



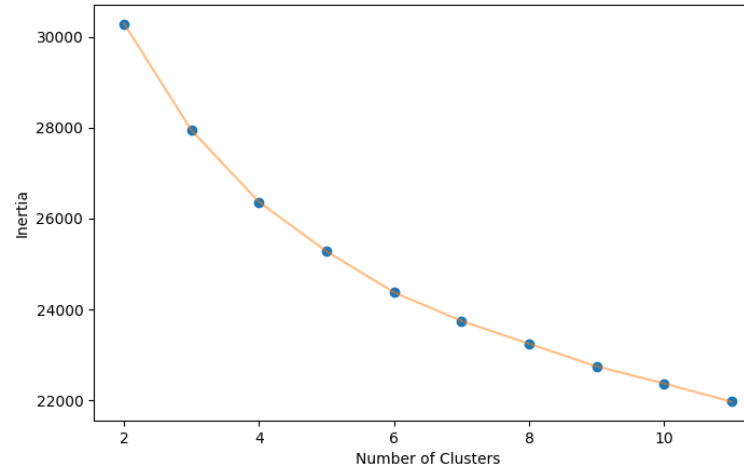
## I 날짜 변수와 클러스터링 변수

### 1. 클러스터링 변수 추가

고객의 요일별, 시간대별, 월별 전력사용량 추이를 반영하여 Kmeans Clustering을 수행

군집개수: elbow method를 통해 4개로 결정

### 2. 고객번호별 평균 전력사용량 변수 추가



고객번호	날짜	시간	전력사용량
1~1500	20210101~20221231	1~24	

1. 날짜 변수 세분화: 날짜를 연, 월, 일, 요일로 구분하여 변수 생성
2. 주말/공휴일 관련 변수 추가: 평일/주말/공휴일 상이한 전력 소비패턴 보임 -> 평일, 주말 구분, 평일 공휴일 및 대체공휴일 여부를 나타내는 변수 추가

## | 파생 변수 추가

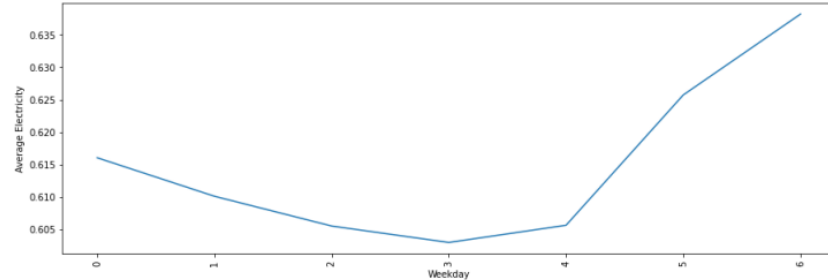
**기상 관련 파생변수 추가:** 기온, 풍속, 습도에 더불어 불쾌지수, 체감온도를 계산하여 활용

**코로나19 관련 파생변수 추가:** 사회적 거리두기 단계를 변수로 추가

**경제적 상황을 반영하는 파생변수 추가:** 전력거래소 제공 공공데이터인 ‘하루전 발전계획용 수요예측량’ 변수 추가.  
해당 데이터는 고용 상황, 소비자 동향 등의 경제 지표를 감안하여 계산됨 -> 경제적 상황을 반영하고자 함

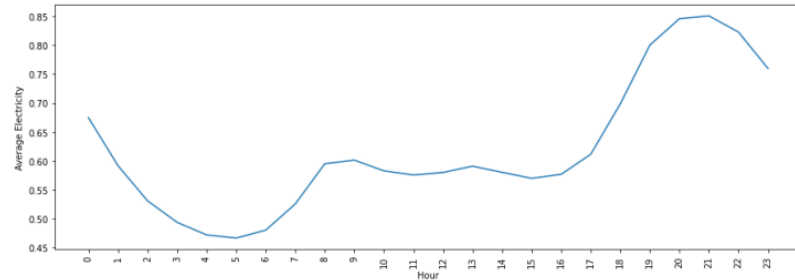
## EDA

[요일별 전력사용량]



평일 전력사용량 < 주말 전력사용량

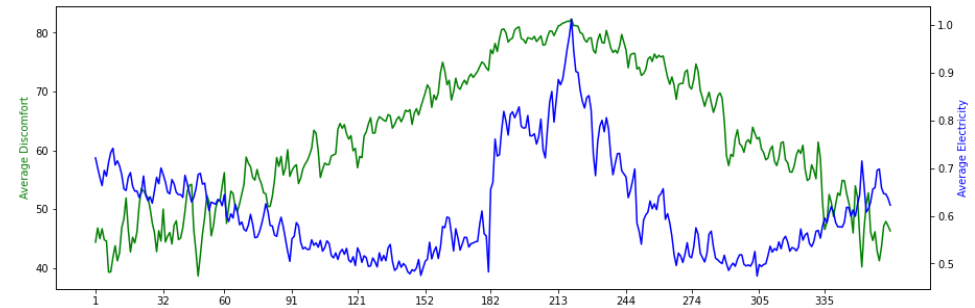
[시간별 전력사용량]



전력 사용 피크시간대: 19~21시 전후

전력 사용 최소시간대: 새벽(취침시간대)

[일평균 불쾌지수와 일평균 전력사용량]



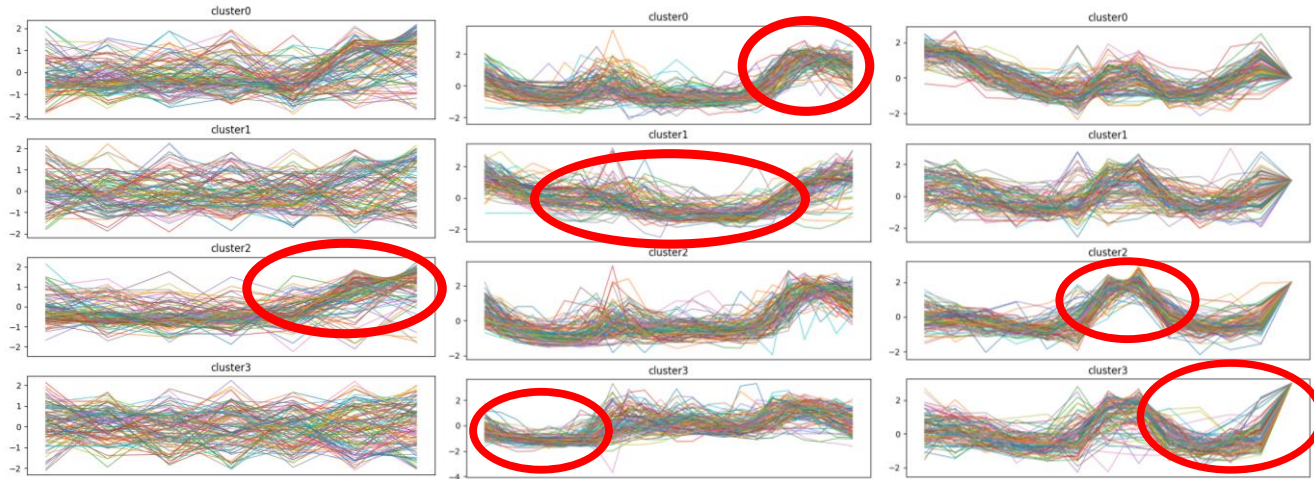
불쾌지수와 전력사용량 간 상관관계를 확인할 수 있음

기온과 체감온도도 비슷한 양상을 보임

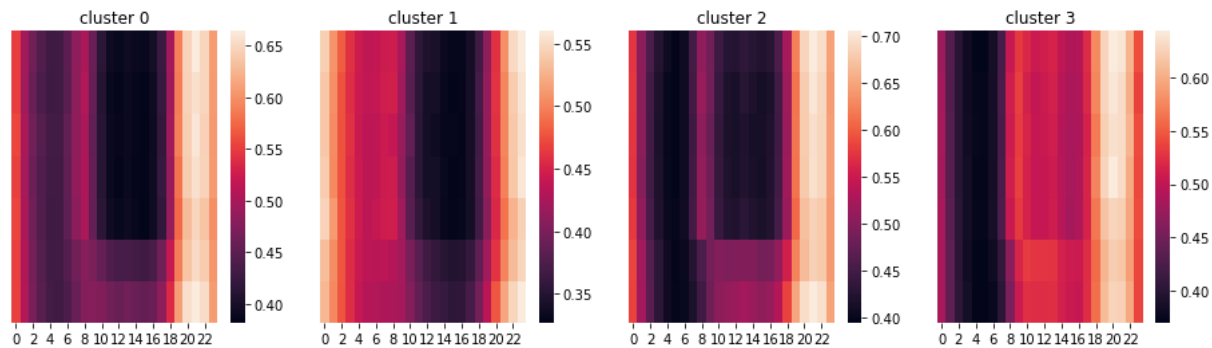


## EDA

모든 군집의 요일별, 시간대별, 월별 전력사용량 추이



요일에 따른 시간대별 전력사용량 추이 heatmap



각자 다른 전력소비패턴을 보이는 것을 확인할 수 있다.

군집 0: 저녁 시간대에 전력사용량이 크게 증가. 비교적 계절과 무관하게 전력사용량이 높은 편

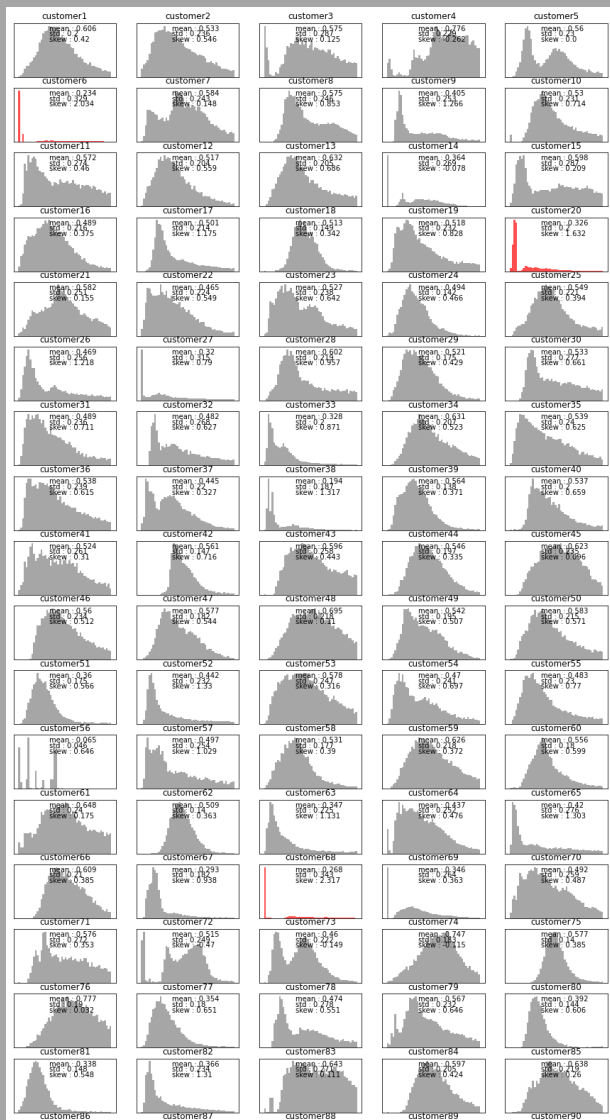
군집 1: 아침 9시부터 저녁 9시까지 전력사용량이 상대적으로 적다가 밤 10시가 되어서야 증가하는 경향을 보인다

군집 2: 저녁 시간대에 전력사용량이 크게 증가한다. 여름에 해당하는 6~8월에 전력사용량이 타군집 대비 두드러진다

군집 3: 새벽(3~6시)에는 전력사용량이 크게 줄어들고 타군집에 비해 낮시간에도 전력사용량이 높은 편. 겨울에 해당하는 11월~12월에 전력사용량이 타군집 대비 두드러진다

군집별 고객의 수는 군집 2(559 명), 군집 0(379 명), 군집 3(323 명), 군집 1(239 명) 순으로 많다.

# 데이터 전처리



고객번호별 전력사용량 편향도

2022년 6월이 2022년 7월 데이터로 중복 제공  
2022년 6월에 해당하는 데이터 제거

전력사용량 결측치 제거  
Target 결측치 5282개 발견, 전체 데이터셋의 약  
0.02% 정도 -> 제거

사회적 거리두기 변수  
one-hot encoding

전력사용량 outlier 제거  
일부 고객 전력사용량 5000 초과  
-> 상위 10%에 해당하는 데이터 제거하여 분석 진행

로그 변환 고려  
고객번호별 전력사용량의 편향도를 확인  
-> 전반적으로 고른 형태

standard\_prediction 표준화  
다른 변수들과 달리 4~8 만 범위의 값  
-> StandardScaler를 사용하여 표준화 진행

## | 최종 데이터셋 특징 및 사용 모델

GBM(Gradient Boosting Machine): CatBoost, LightGBM, XGBoost

### | 1. 시간에 대한 정보가 매우 중요한 변수로 작용하는 시계열 데이터

-> GBM은 이전 시간의 예측 오차를 최소화하도록 모델 학습 -> 시간적 순서에 따른 패턴을 상대적으로 잘 학습하고 비선형성을 효과적으로 처리하여 복잡한 관계를 모델링할 수 있다.

### | 2. 변수 27개, 범주형 변수 여러 개 포함

-> 트리 기반의 GBM 모델은 범주형 변수 처리에 적합하며, 모델을 학습시키는 과정에서 모델의 예측력이 얼마나 감소하는지를 계산하여 변수의 중요도(feature importance)를 파악할 수 있음

### | 3. 2000만 행 이상의 대용량 데이터셋

-> CatBoost는 대용량 데이터셋에 적합한 자체적인 오버피팅 감소 전략을 통해 모델이 과적합되는 것을 방지하고 일반화 성능을 향상. LGBM은 효율적인 리프레벨 학습 방식과 데이터 분할 기법을 사용하여 메모리 사용을 최적화하고 처리 속도를 높임.

# | 모델링 결과

8:2 비율로 TrainTestSplit

하이퍼파라미터 튜닝: GridSearch

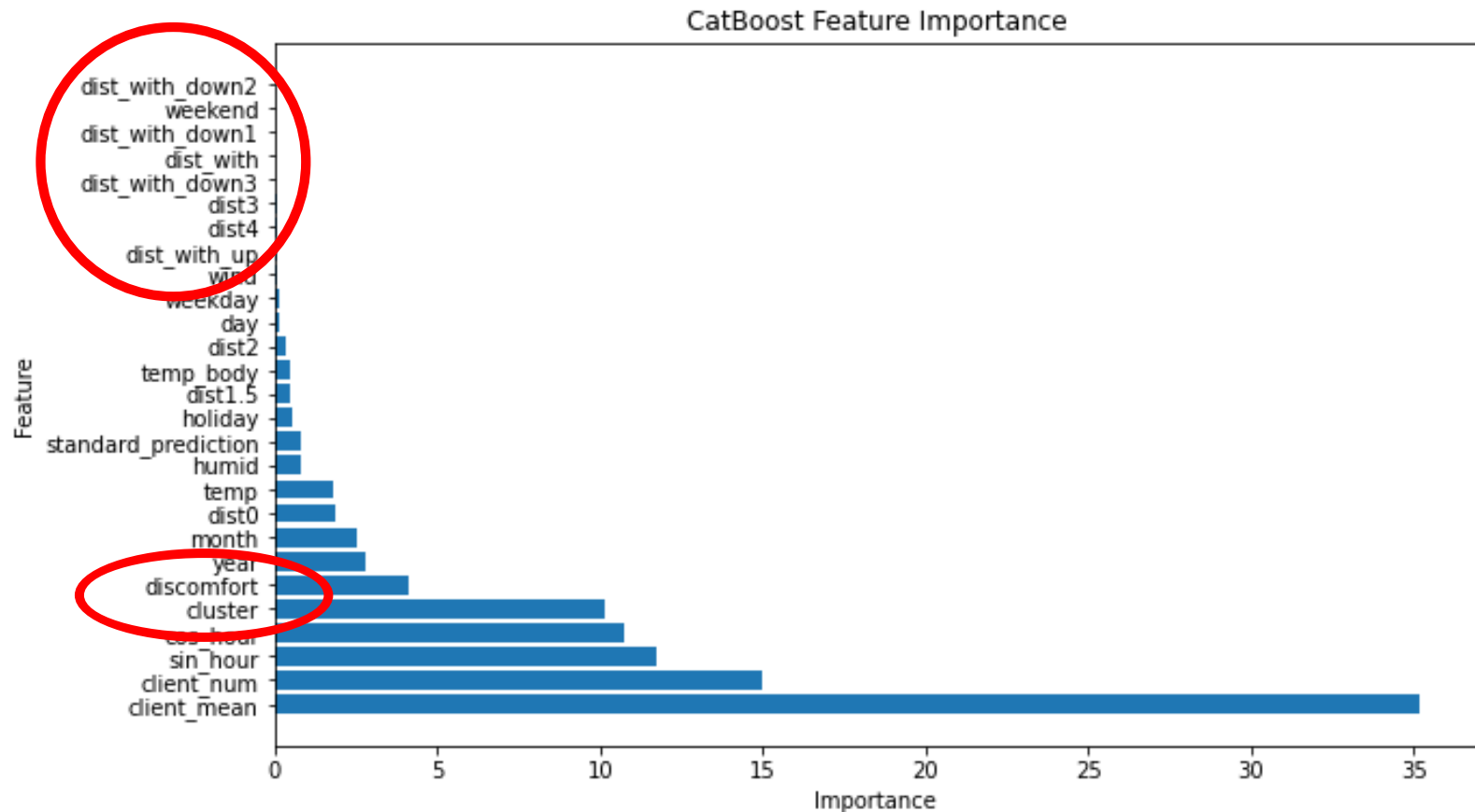
	LightGBM	XGBoost	CatBoost
R M S E	0.21	0.205	<b>0.198</b>
R R M S E	38.4	37.48	<b>36.2</b>

최종 모델: CatBoost



## I 변수 중요도

도출 성능이 가장 높게 나왔던 CatBoost를 사용하여 feature importance 계산 → 인사이트 도출에 활용

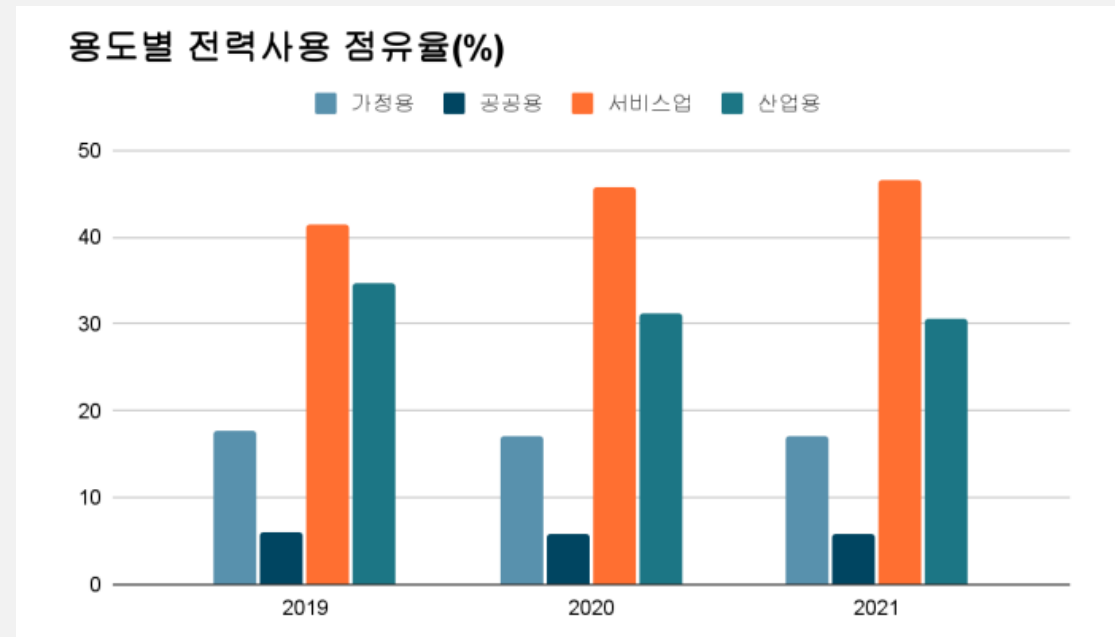


## | 모델링 시도



## | 사업화 방안

### ‘관광 및 서비스업 분야의 전력 수요 예측을 통한 안정적인 관광사업 유지’



제주 지역: 서비스업에서의 전력 수요량이 큰 비중을 차지(4~50%) -> 관광 및 서비스업에서의 전력 수요 예측은 매우 중요한 문제  
관광산업 및 서비스업 분야의 전력 사용량 데이터를 활용하여 예측 모델을 구현한다면 보다 원활하게 관광사업을 지속할 수 있을 것

## | 사업화 방안

### ‘환경 친화적인 전기 사용량 조절 앱 개발’

국민들을 대상으로 에너지 공공요금 인상에 대해 질의한 설문조사에 따르면, 60.4%가 서민층에 부담이 되므로 최대한 억제해야 한다는 부정적 반응

-> 많은 사람들이 전기요금 인상에 대해 부담을 느끼고 있음. 해당 앱에 대한 수요 역시 있을 것으로 보임

해당 앱을 통해 균형 잡힌 전력 사용과 더불어 전력 사용량의 감소를 통한 환경적 이점까지 기대할 수 있음





## I 기대효과

### 1. 다각도로 전력 수요를 분석하여 제주 지역의 안정적인 전력 수급 계획에 도움

여러 가지 기상 요소에 더해 데이터셋의 연도적 특성(코로나 19 사태로 인한 생활패턴의 변화)도 반영

-> 제주 지역만의 특성뿐만 아니라 예기치 않게 발생하는 사회적 변화를 반영하는 방법 또한 제시

### 2. 제주 지역의 신재생에너지 보급과 확대에 있어 효율적인 설비 가동을 가능하게 할 것

각종 기상 조건은 신재생에너지 발전량에 큰 영향

-> 에너지 발전량의 기복이 심한 편이고, 이로 인해 전력 수급에도 차질을 빚을 우려가 있음

기상 요소를 고려하여 예측된 전력 수요량을 통해 신재생에너지 발전을 효과적으로 지속하기 위한 발판을 마련

## I 분석 한계 및 개선 방안

데이터 안심구역 내 분석 환경의 제한

데이터의 한계



코로나19의 영향을 보다 더 자세하게 반영

GDP와 같은 경제 지표 관련 변수 활용

고객들의 특성, 생활패턴, 경제활동 등을 반영

딥러닝 모델 사용

## I 활용데이터 및 참고문헌

### • 공공 데이터 및 외부 데이터

#### 1. 산업통상자원부 산하기관 공공데이터

- 한국전력거래소>주요사업>전력관련정보>하루전 발전계획용 수요예측
- 한국전력공사>전기자료>전력통계>한국전력통계>8.2 행정구역별 발전설비 및 발전량

#### 2. 외부 데이터

- 기상자료개방포털>데이터>기상관측>지상>종관기상관측>제주도(고산, 서귀포, 성산, 제주) >기온, 풍속, 습도

### • 참고 문헌

- [1] 김혜민, 김인겸, 박기준, 유승훈, "기온이 전력수요에 미치는 영향 분석", 에너지공학 24.2 (2015): 167-173.
- [2] 신이레, 윤상후, "전력수요예측을 위한 기상정보 활용성평가", 한국데이터정보과학회지 27.6 (2016): 1601-1607.
- [3] 정현철, 정재성, 강병오, "ARIMA 모델 기반 생활 기상지수를 이용한 동 · 하계 최대 전력 수요 예측 알고리즘 개발", 전기학회논문지 67.10 (2018): 1257-1264.
- [4] 박주환, 조영호, 이두희, "코로나 19 영향을 고려한 머신러닝 기반의 전력사용량 예측 알고리즘 구현", 대한전기학회 학술대회 논문집 2021.4 (2021): 211-212.
- [5] 정희원, "제주계통의 전력수요예측 및 적정 BESS 용량 산정에 대한 연구", 대진대학교 대학원 석사학위논문, 2017.12
- [6] 한국전력공사, "2022 년 한국전력통계", 제 92 호, 2023.5
- [7] 산업통상자원부, "신재생에너지생산량", 2021.12
- [8] 산업통상자원부, "전력수급동향", 2021.12
- [9] 국가통계포털(KOSIS), "용도별 전력사용량", 2019-2021