

슬기로운 **예측** **생활**

전력 수요량
예측모델

이문형

오재환

김성수

김성현





1. 팀원소개



2. 프로젝트
선정배경



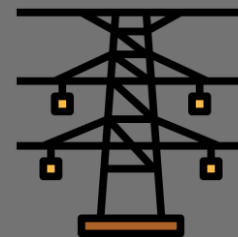
3. 개발내용



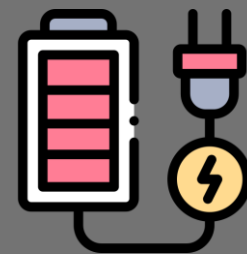
4. 활용방안




5. 개선점



CONTENTS





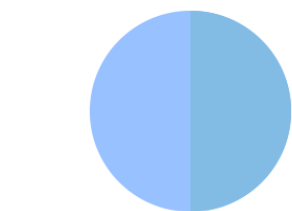
팀원 소개

슬기로운 예측생활

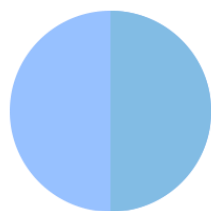
전력 수요량 예측모델



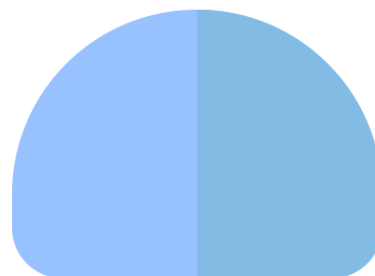
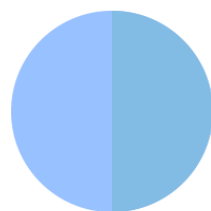
팀원 소개



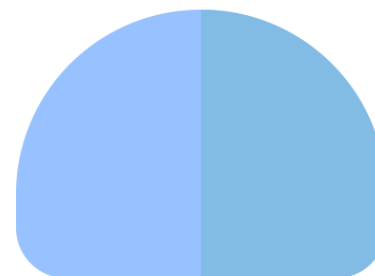
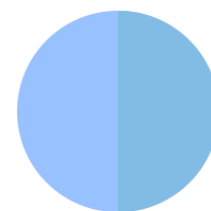
이문형




오재환



김성수



김성현



프로젝트 선정배경

슬기로운 예측생활

전력 수요량 예측모델



프로젝트 선정배경





프로젝트 선정배경





프로젝트 선정배경 - 문제점



에너지의 수출입 규제



전력 생산 시 저장의 어려움



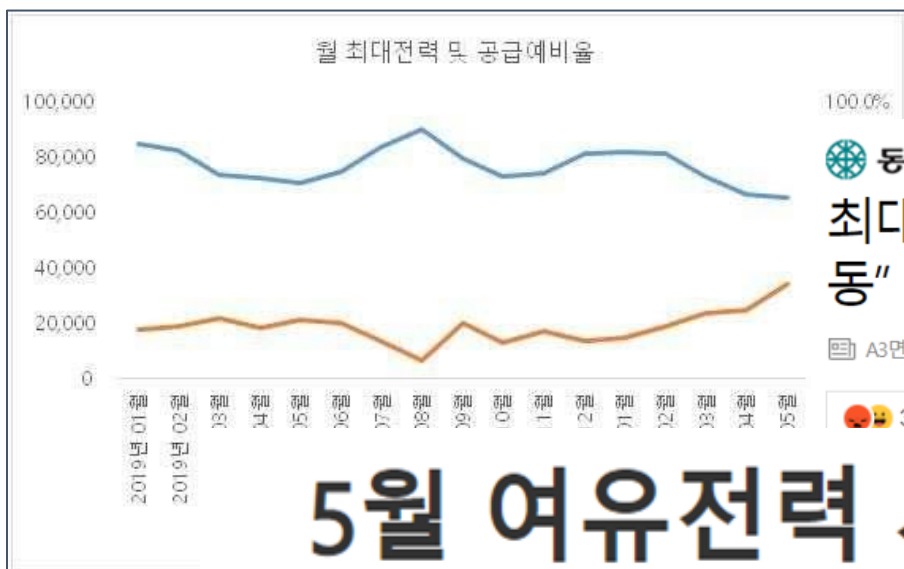
생산된 만큼 사용하지 않으면 곧바로 낭비가 되는 특성



잠깐의 전력 공급 중단이 경제, 사회적으로 큰 피해를 초래함



프로젝트 선정배경 - 실 사례



동아일보

최대 전력수요, 정부 당초 예측 빗나가... "원전 2기 추가 가동"

A3면 1단 | 기사입력 2018.07.20. 오전 3:01 | 최종수정 2018.07.20. 오전 4:06

기사원문

스크랩



본문듣기 · 설정

3,614

3,370

요약본

가



5월 여유전력 사상 최대...전기가 남는다

CBS노컷뉴스 유동근 기자 2020.06.07. 17:25



프로젝트 선정배경 - 관련 연구



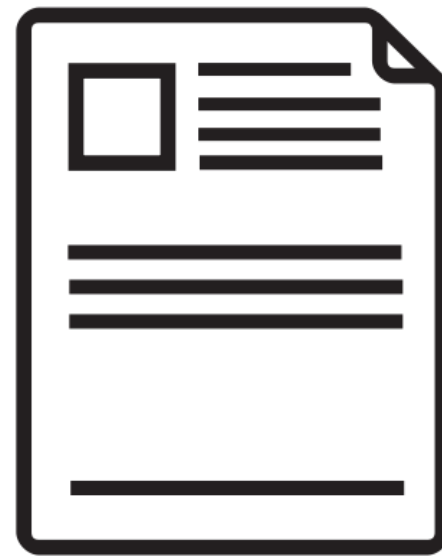
특허 사례

- ◆ 기상변화에 따른 전력수요지수 예측 시스템 및 그 방법(2016), (주)나우드림
- ◆ 사전 전력수요예측에 의한 전력수요관리 방법 및 시스템(2018), 한국전력공사
- ◆ 단위 가구 피크 전력 수요량 예측 방법 및 시스템(2019), 전자부품연구원



국내 연구 사례

- ◆ 권보성, 송경빈. (2020). 심층 신경망을 이용한 동·하계 최대 전력수요의 앙상블 예측 방법. 전기학회논문지, 69(6), 765-771.
- ◆ 이지원, 김형준, 김문겸. (2020). 빅데이터를 이용한 인공지능 기반 시간별 전력수요 예측. 전기학회논문지, 69(6), 792-799.
- ◆ 이용건, 오재영, 김기백. (2020). 머신러닝 해석 기법을 이용한 전력 수요 예측 모델 해석. 전기학회논문지, 69(3), 480-485.





개발 내용

슬기로운 예측생활

전력 수요량 예측모델



개발 내용 - ① 프로젝트 개요



분석 개요

- ▣ 인천시 전력 사용 기록(전력거래소, 1시간 단위)
 - ▣ 기상 관측 자료(기상청),
 - ▣ 날짜 데이터 등 공공데이터
- = 각 가정 및 회사의 일별 전력 사용량을 예측**



분석 내용

356세대의 2018.06.01 ~ 2018.06.30, 30일간,

'일간 전력사용량' 예측

※ 이번 분석에서는 생활 습관, 산업 활동, 경제적 요인 등 전력사용량에 영향을 미치는 다른 변수는 고려하지 않음



개발 내용 - ② 프로젝트 계획

목표 및 분석 모델



성과 지표 SMAPE

(Symmetric Mean Absolute Percentage Error)

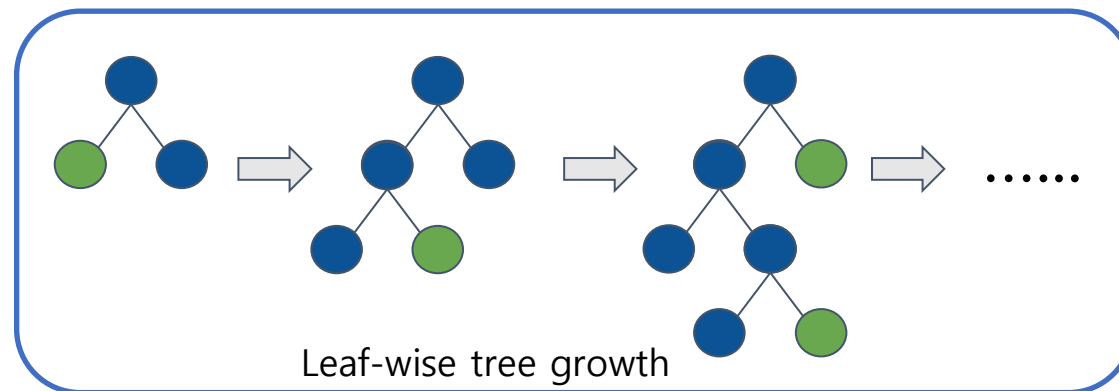
$$\text{SMAPE} = \frac{100\%}{n} \sum_{t=1}^n \frac{|F_t - A_t|}{(|A_t| + |F_t|)/2}$$

20 이내 달성



분석 모델

① 메인 모델 : LightGBM - 머신러닝 앙상블 알고리즘



② 비교 모델 : Decision Tree, Random Forest Tree, Bagging, Extra Tree, Adaboost, Catboost, Deep Neural Network



개발 내용 - ② 프로젝트 계획



개발 환경

① 협업 도구 :

▨ GitHub

▨ Google Presentation

② 사용 언어 및 IDE :

▨ Python & Library

▨ Jupyter Notebook

bunchacha / power-forecasting

<> Code ⓘ Issues 🔗 Pull requests ⚙️ Actions 📁 Projects 📖 Wiki 🛡️ Security 📊 Insights ⚙️ Settings

🔗 master 4 branches 0 tags Go to file Add file Code

jaehwan0995 finish_code 62a2673 1 hour ago 46 commits

.ipynb_checkpoints	remove csv	yesterday
csv	before 1week	yesterday
DL_preprocessing.ipynb	Create DL_preprocessing.ipynb	5 hours ago
DeepLearning_0728.ipynb	finish_code	1 hour ago
lightgbm_0724.ipynb	lightGBM model_ver1	4 days ago
lightgbm_0727.ipynb	xgboost catboost 수정	8 hours ago
lightgbm_0728.ipynb	finish_code	1 hour ago
tree_0728.ipynb	finish_code	1 hour ago
ver_4_miniProj.ipynb	EDA 완료	4 hours ago
weather_d_practice.csv	parameter test	5 hours ago
weather_practice.csv	parameter test	5 hours ago



개발 내용 - ② 프로젝트 계획



개발 일정

구분	6월 넷째 주	7월 첫째 주	7월 둘째 주	7월 셋째 주	7월 넷째 주
데이터 집계 및 전처리					
데이터 탐색 및 분석					
데이터 분할 (train, test set 파티셔닝)					
부스팅 알고리즘 스터디 (lightgbm, xgboost, catboost)					
데이터 분석 및 예측 (메인 모델 : lightgbm , 연습 : MLR, ARIMA, Random Forest Tree, k-NN, Neural Network, Deep Learning 등) ※ 분석 기법은 추후 변경 가능함					
결과보고서 작성 및 개발 완료					



개발 내용 - ③ 데이터 집계



(1) 전력 데이터

= 인천시내 1500세대(또는 상가)의 시간당 전력 사용량(kWh)

	NX1	NX2	NX3	NX4	NX5	NX6
count	3064	3065	3041	3006	0	3041
mean	0.371506	0.410571	0.519877	0.260363		0.268325
std	0.190358	0.272845	0.236184	0.231955		0.149223
min	0.049	0.066	0.053	0.017		0.034
25%	0.271	0.225	0.416	0.114		0.162
50%	0.329	0.338	0.485	0.1765		0.228
75%	0.43	0.536	0.581	0.339		0.346
max	4.104	4.837	6.232	5.624		2.718



개발 내용 - ③ 데이터 집계

⚡ (2) 기상 관측 자료

- 인천_시간별_기상자료(16-18).csv

- 인천_일별_기상자료(16-18).csv

	지점	기온(°C)	강수량(mm)	풍속(m/s)	습도(%)	적설(cm)	전운량(10분위)
count	17088	17057	1435	17088	17081	510	13313
mean	112	12.40511813	1.305505226	3.108760534	68.03291376	1.353333333	5.152407421
std	0	10.72658485	3.645712266	1.616314135	20.03519318	1.321686526	4.088560691
min	112	-17.1	0	0	12	0	0
25%	112	3.3	0	2	52	0.4	0
50%	112	13.6	0.1	2.8	68	0.8	6
75%	112	21.5	1.1	4	85	1.9	9
max	112	33.9	58.3	11.6	100	8	10

	지점	평균기온(°C)	최저기온(°C)	최고기온(°C)
count	721	721	721	721
mean	112	12.57059639	9.34590846	16.34895978
std	0	10.49633512	10.68690076	10.55503446
min	112	-14.6	-17.1	-11.9
25%	112	3.5	0	7.4
50%	112	14.2	10.3	18.2
75%	112	21.9	18.4	25.5
max	112	30.9	27.6	34



개발 내용 - ④ 데이터 전처리 (기상)

1. 변수 선택

날씨, 전운량 feature 제거 : 결측치가 너무 많음. 추가 예측모델을 만들어 봤지만 성능이 떨어졌음

년, 월, 주, 일, 요일, 공휴일, 주말 여부를 feature로 추가로 고려함

2. 이상치 처리

Tukey Fences 방법론을 사용해, 사분위 범위(IQR) 기반으로 이상치를 NaN 값으로 변경함

3. 결측치 처리

'온도, 습도'는 직전 값으로 처리함

'강수량, 적설'은 0으로 처리함



개발 내용 - ④ 데이터 전처리 (기상)

4. 시간별 데이터를 일별 데이터로 변환함

'풍속, 습도'는 하루 평균값으로 변환함

'강수량, 적설'은 하루 총합값으로 변환함

5. 정규화

정규화 방식으로 Min-Max 전략을 사용함 (Deep Neural Network 모델에서 사용)

6. 원 핫 인코딩

카테고리형 feature들에 대하여 원 핫 인코딩을 적용함 (Deep Neural Network 모델에서 사용)



개발 내용 - ④ 데이터 전처리 (전력)



전력 사용량 데이터에서도 결측치, 이상치 이슈 존재

=> 발생 이유를 반영하여 전력 데이터 전처리를 수행함



NaN이 발생한 경우, 직전 시간의 전력 사용량 값이 상당히 큰 경향이 있음.
이는 미터링 데이터 수집 시스템의 특징으로 보임.

(1) 전력 측정기의 통신 문제

=미터기 인식 오류로 누적된 값이 들어옴

(2) 해당 시간대에 사용을 하지 않았음

(3) 처음에 NaN이 있는 경우는 미터기 설치가 안된 경우임





개발 내용 - ④ 데이터 전처리 (전력)



전력 사용량 데이터 전처리 절차

1) 이상치 처리

Tukey Fences 방법론을 사용해, 사분위 범위(IQR) 기반으로 이상치를 NaN 값으로 변경함

2) 모든 NaN 값을 0으로 처리함

3) 유효한 데이터만 사용하기 위하여 시계열 앞뒤로 0인 행을 제거함

4) 전력 사용량 중간에 있는 0을 (전체 평균값, 직전 값, 구간 평균값)으로 처리함





개발 내용 - ④ 데이터 전처리 (전력)



전력 사용량 데이터 전처리 절차

전력 사용량 데이터는 label 데이터 이면서도, 시계열 예측을 위한 feature로 활용함

5) 시간별 전력 사용량을 일별 전력 사용량으로 변환함

6) 변수 선택

(직전 1주일 치의 전력 사용량 및 1주일 구간 평균값,

직전 2주일 치의 전력 사용량 및 2주일 구간 평균값,

직전 3주일 치의 전력 사용량 및 3주일 구간 평균값) 을 feature로 추가로 고려함

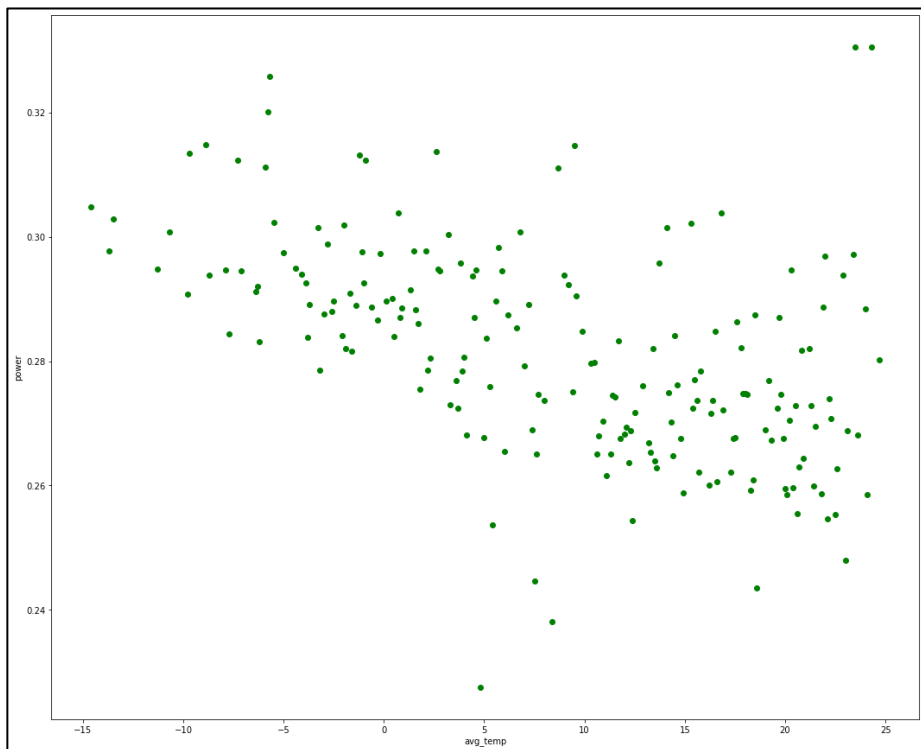
따라서, 라벨링 결측치 처리 규칙과 변수 선택 조건에 따라 총 9개의 데이터 셋이 생성. (3*3개)



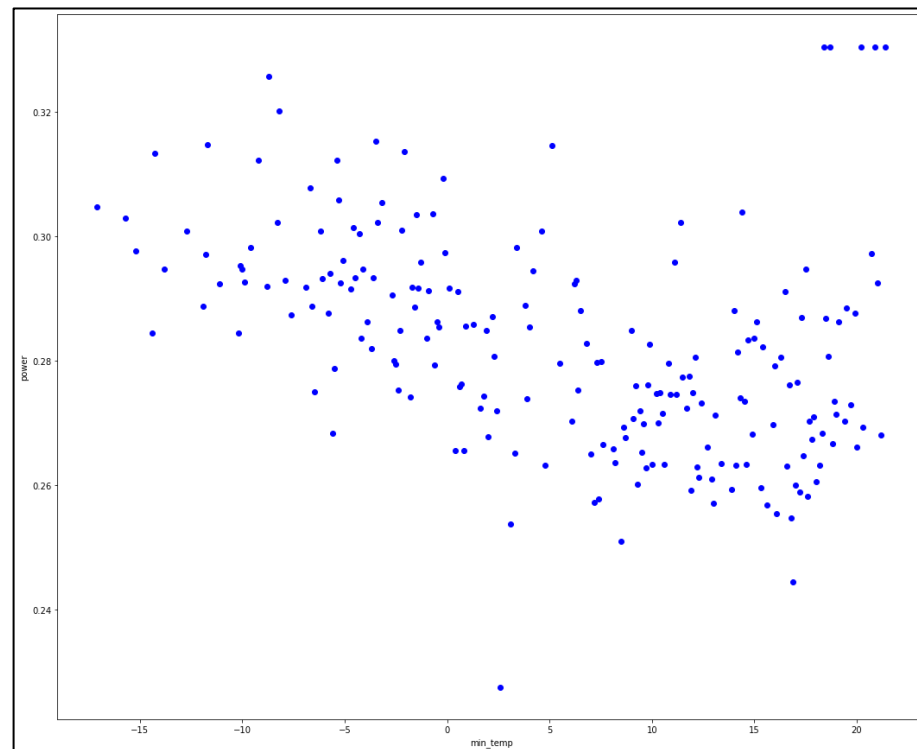
개발 내용 - ⑤ 데이터 탐색(EDA)

1. 기상 데이터와 전력량과의 관계 - 상관 관계가 없음

(원인으로는 산업용 비중이 높은 인천시의 지역적 특성으로 파악가능)



평균 기온과 전력량 관계



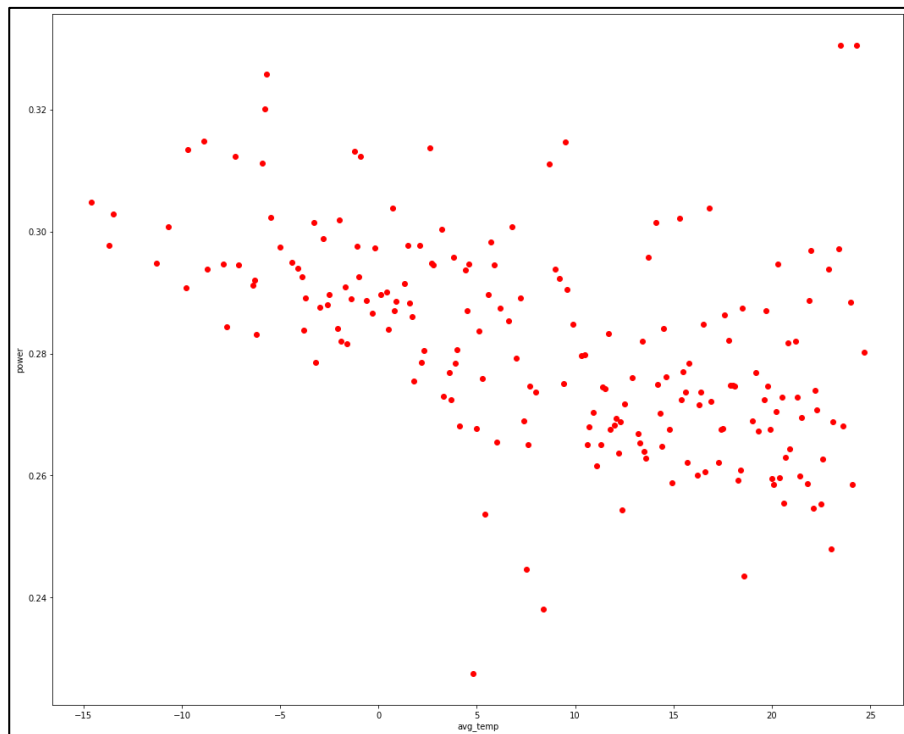
최저 기온과 전력량 관계



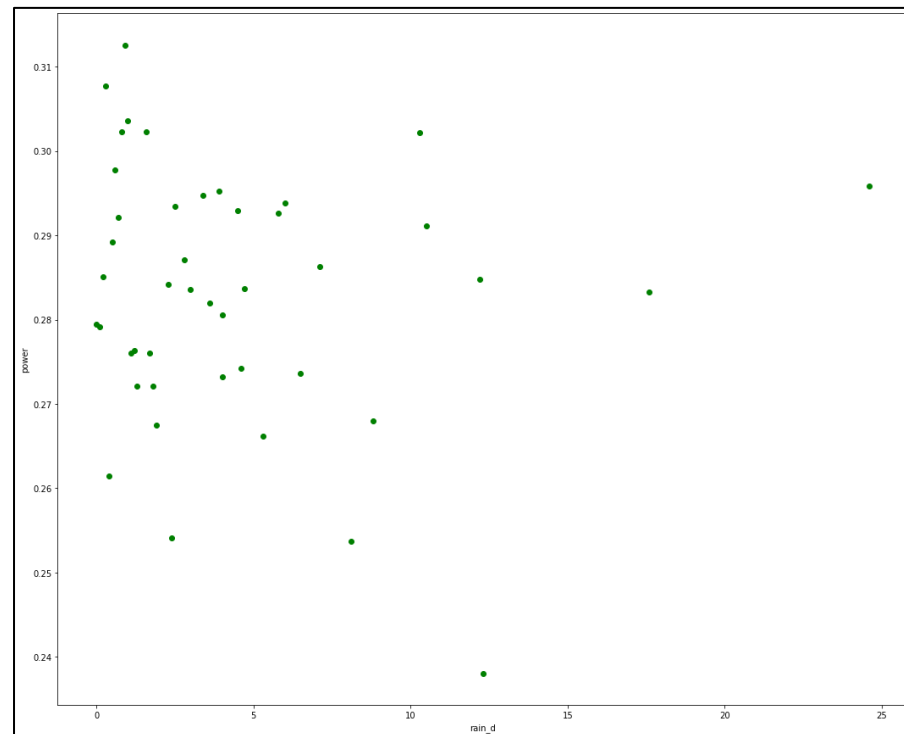
개발 내용 - ⑤ 데이터 탐색(EDA)

1. 기상 데이터와 전력량과의 관계 - 상관 관계가 없음

(원인으로는 산업용 비중이 높은 인천시의 지역적 특성으로 파악가능)



평균 기온과 전력량 관계



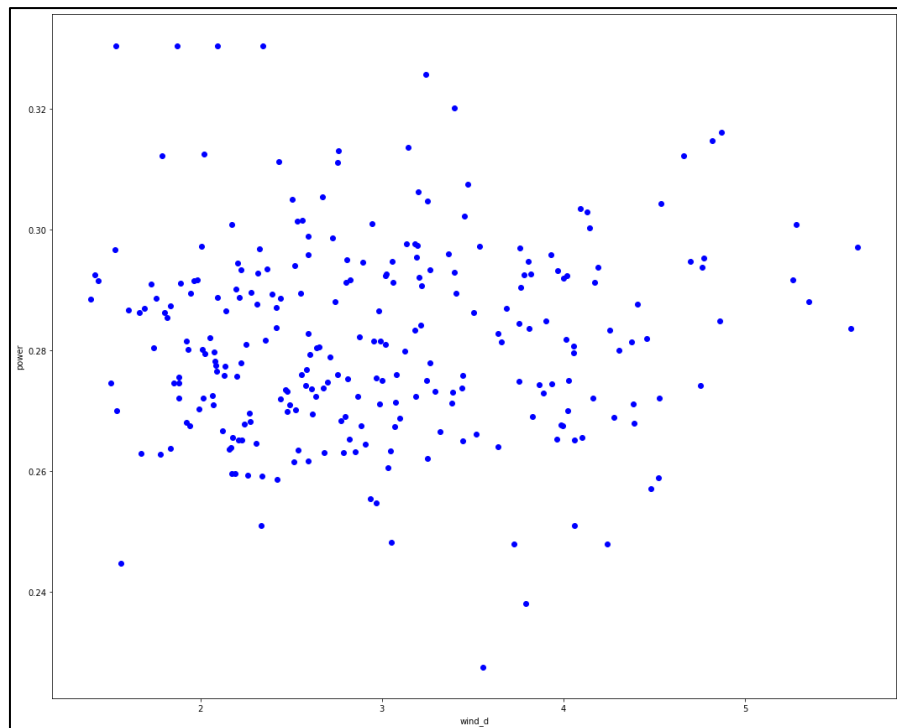
강수량과 전력량 관계



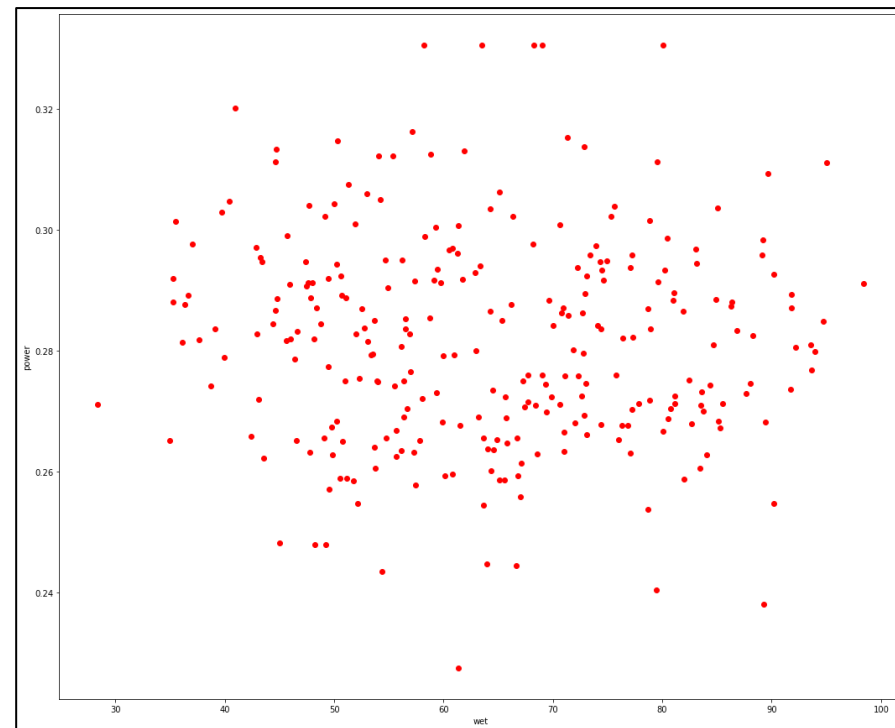
개발 내용 - ⑤ 데이터 탐색(EDA)

1. 기상 데이터와 전력량과의 관계 - 상관 관계가 없음

(원인으로는 산업용 비중이 높은 인천시의 지역적 특성으로 파악가능)



풍속과 전력량 관계



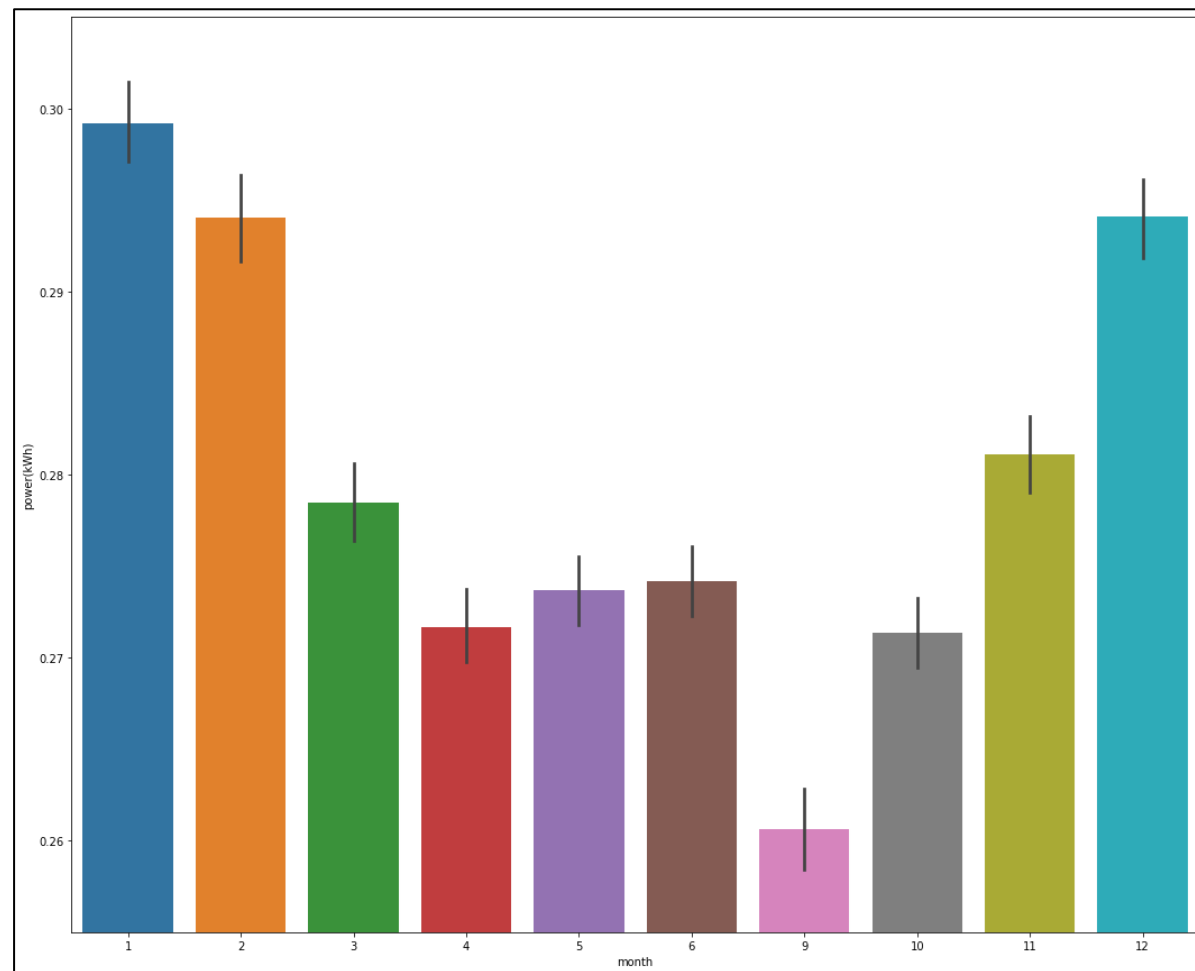
습도와 전력량 관계



개발 내용 - ⑤ 데이터 탐색(EDA)

2. 날짜 데이터와 전력량과의 관계 - 월

- 겨울철(11~2월)의 전력량 高



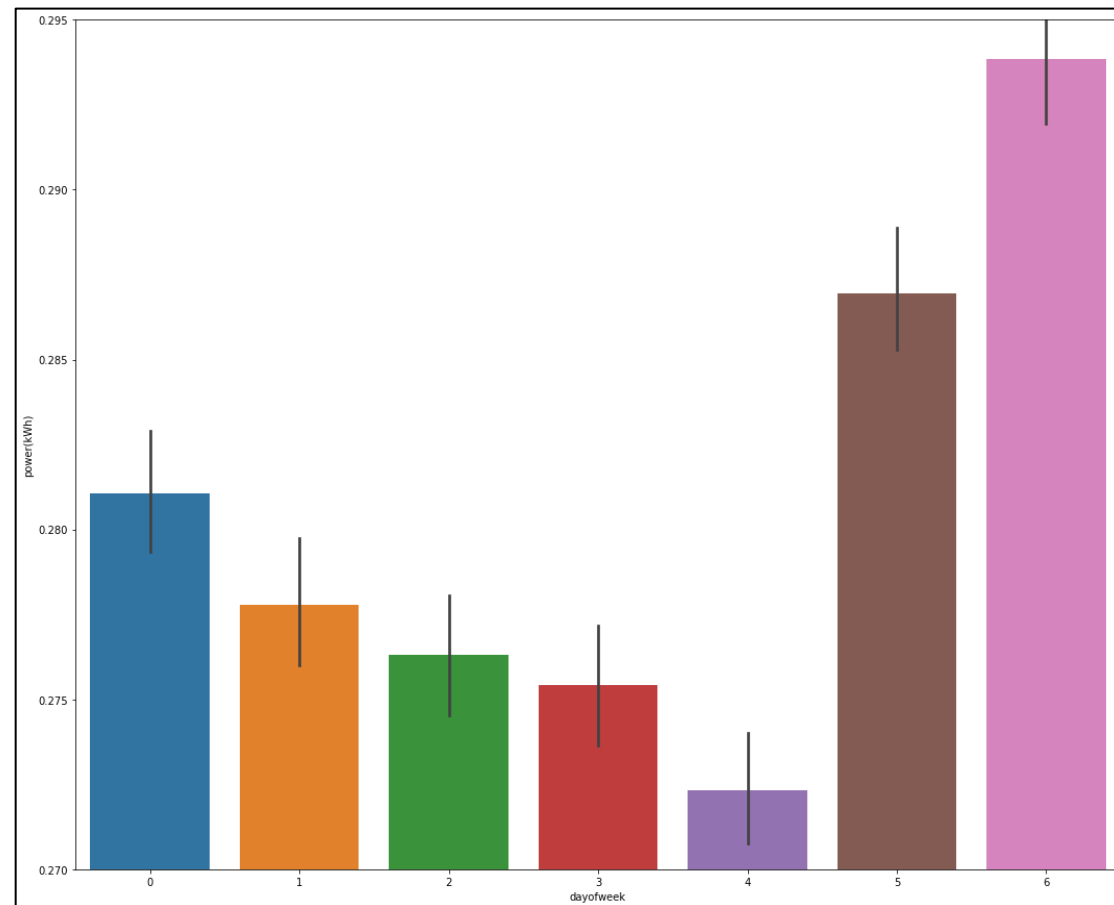


개발 내용 - ⑤ 데이터 탐색(EDA)

2. 날짜 데이터와 전력량과의 관계 - **요일**

- 월~금요일까지 전력량 高, 감소함

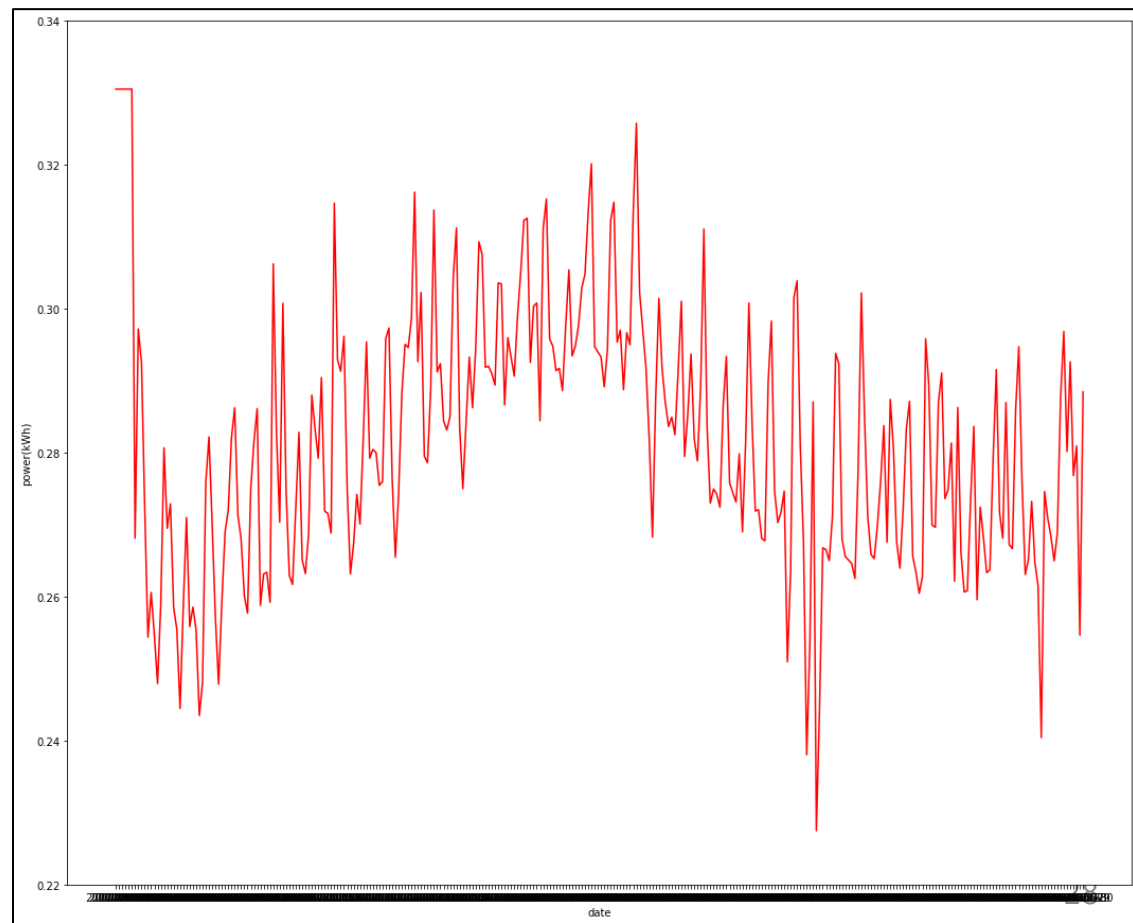
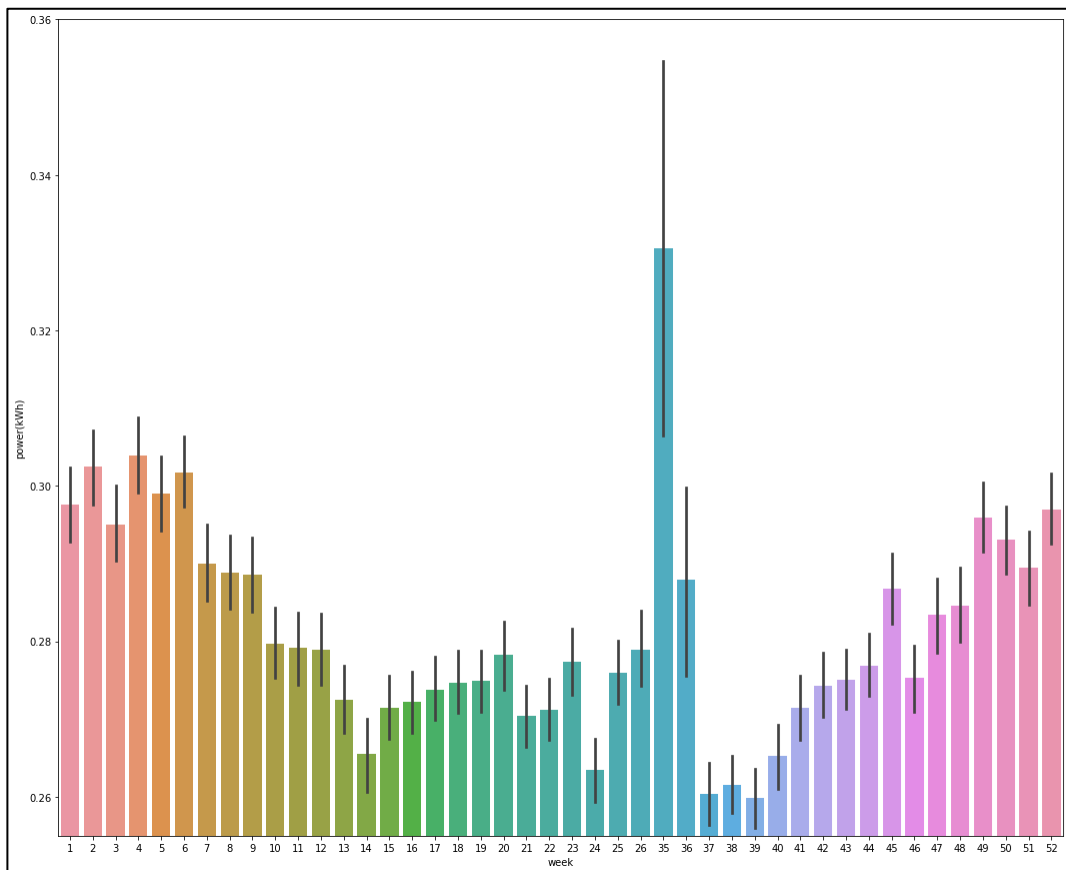
- 토~일요일까지 전력량 底, 증가함





개발 내용 - ⑤ 데이터 탐색(EDA)

2. 날짜 데이터와 전력량과의 관계 - 주, 일

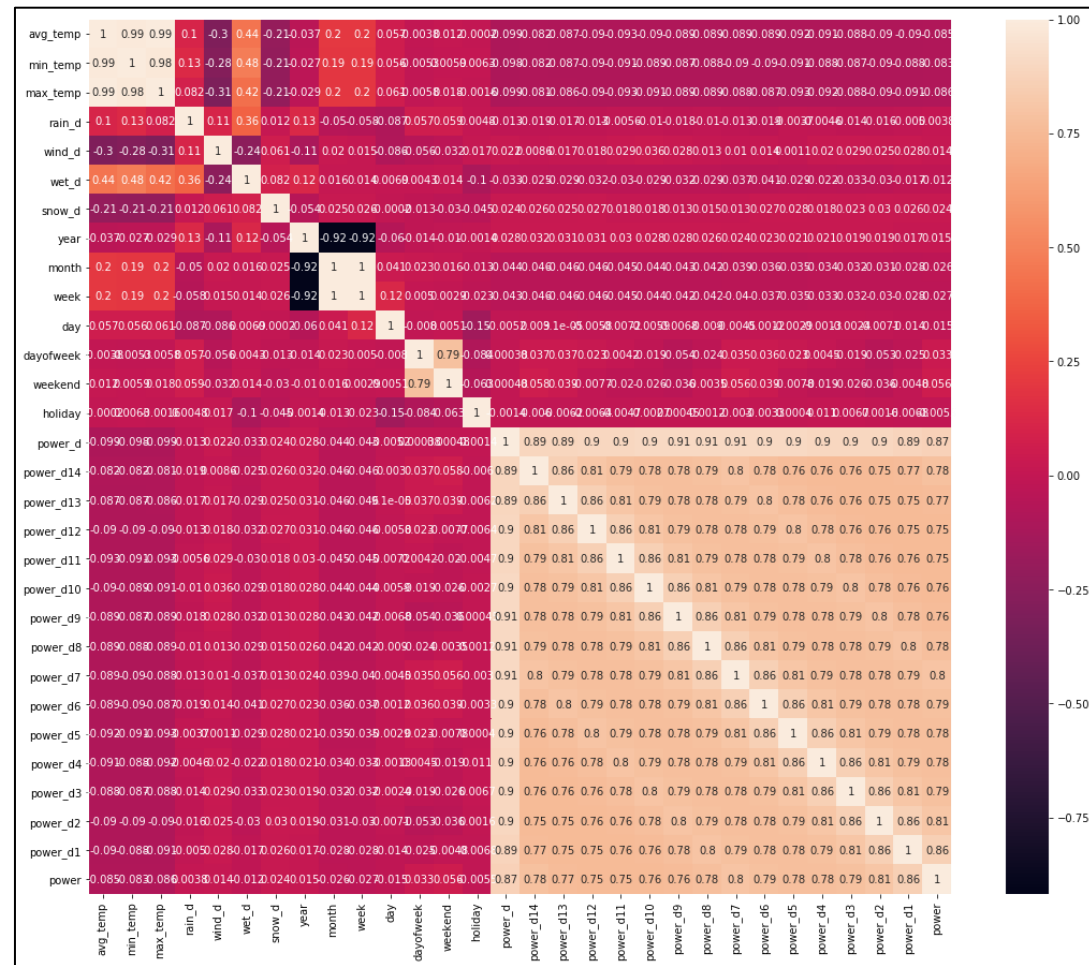




개발 내용 - ⑤ 데이터 탐색(EDA)

3. 모든 속성과 전력량과의 관계

- 기상 데이터와 전력량은 낮은 양의 상관 관계
- 날짜 데이터와 전력량은 낮은 양의 상관 관계
- 과거 전력량은 높은 양의 상관 관계

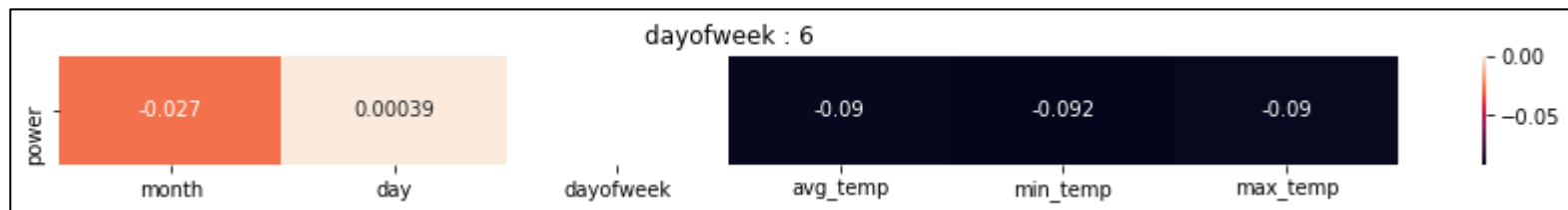
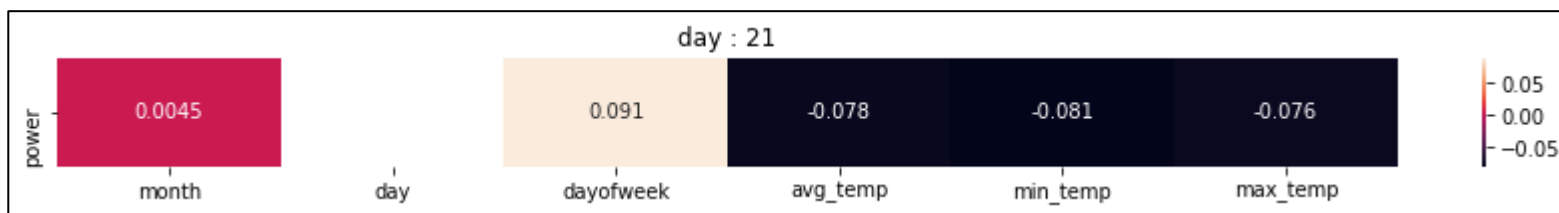
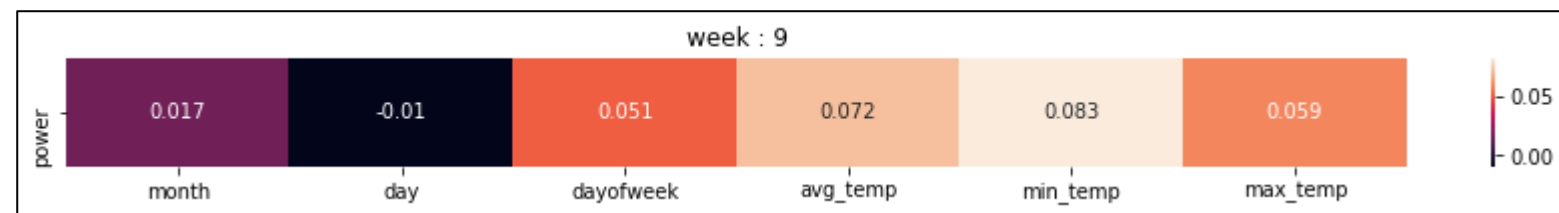
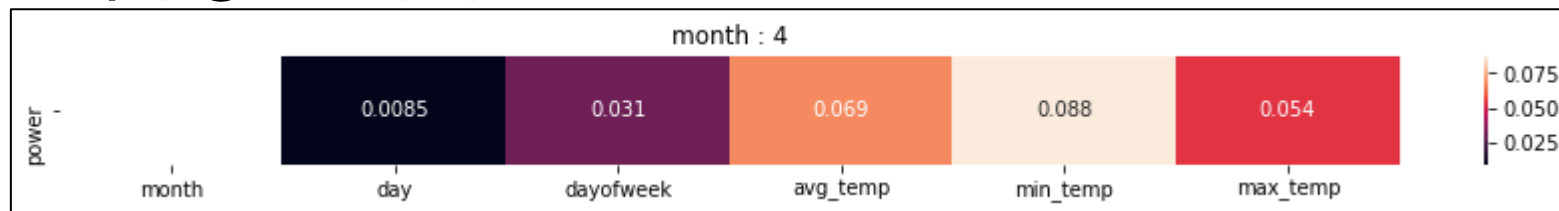




개발 내용 - ⑤ 데이터 탐색(EDA)

월, 주, 일, 요일 기준으로 통합 후 전력과 상관관계 비교

- 대체적으로 상관관계가 0에 근사함





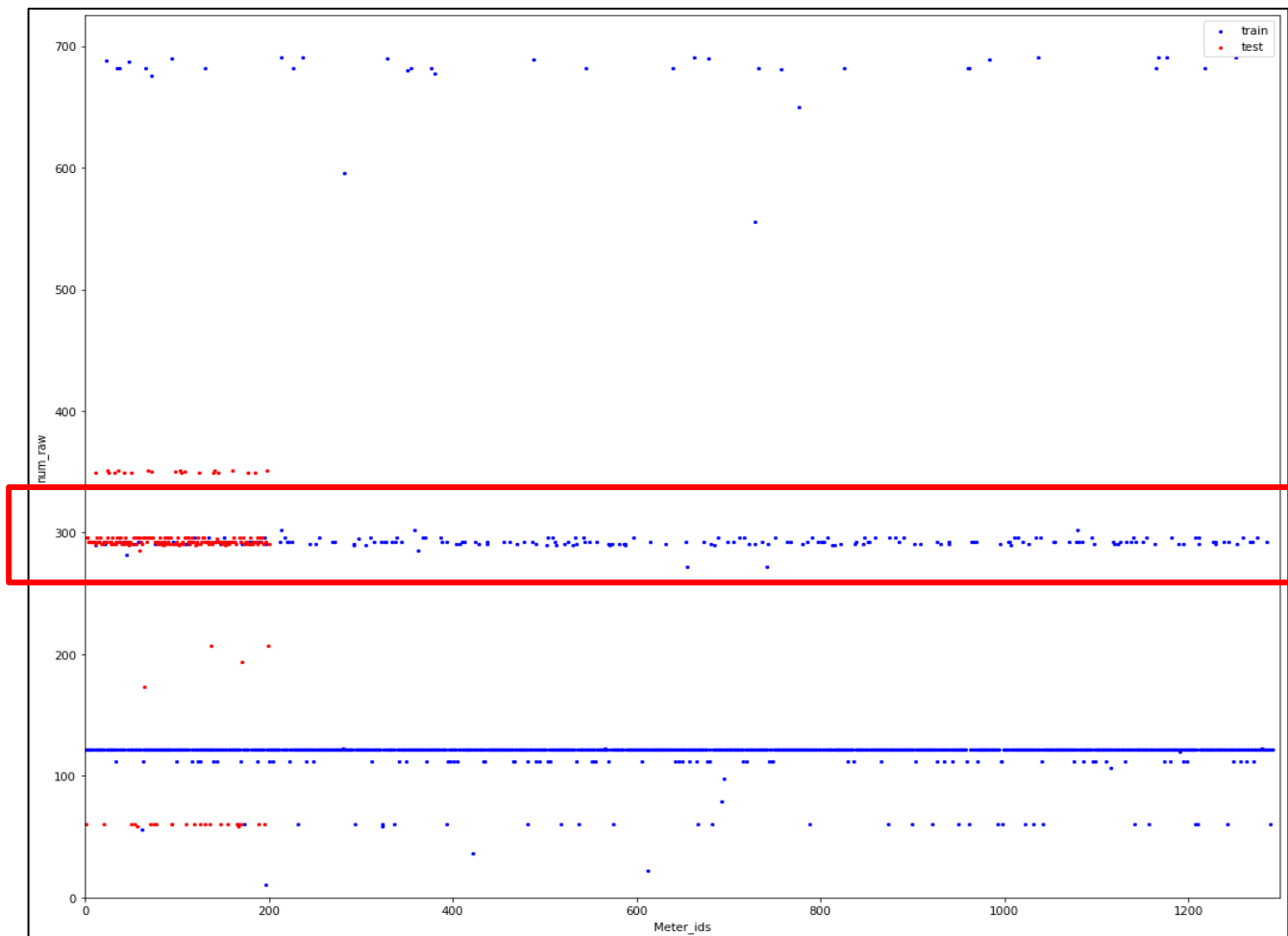
개발 내용 - ⑥ 데이터 파티셔닝

⚡ 데이터 수의 분포가 균등한 세대만을 뽑아 데이터 셋으로 설정.

⚡ data leakage 문제로 Train + Test 데이터를 통합하여 전체 데이터 셋으로 합침.

⚡ 356세대만 데이터 분석에 사용함.

⚡ 그 후에 시계열 기준으로 Train(80%, 약 개)
Validation(10%, 약 개) Test(10%, 약 개)로 분리함.





개발 내용 - ⑦ 모델링

lightgbm을 메인 모델로 선택한 이유

1. 트리 기반 모델의 장점 : 높은 설명력(기상, 날짜 데이터와 전력 사용량과의 관계를 파악하고 싶었음), 데이터 scale에 구애 받지 않음
2. 부스팅 알고리즘의 장점 : 테이블 형태로 저장되어 있는 정형화된 데이터(속성이 정확한 의미가 있을 때)를 분석할 때, 데이터 경진대회 및 선행 연구에서 winning 알고리즘임
3. lightgbm의 장점 : 부스팅 알고리즘 중에서도 높은 성능, 효율적인 학습속도와 메모리 사용량, 대량 데이터 셋(10000건 이상)에 적합함



개발 내용 - ⑦ 모델링



lightgbm 하이퍼 파라미터 튜닝

- 알고리즘 성능에 큰 영향을 준다고 판단되는 파라미터부터 순차적으로 Grid Search를 진행함
- 전처리 조합 중에서 [라벨링 결측치 처리 구간 평균값, 속성 추가: 직전 2주일 치의 전력 사용량 및 2주일 구간 평균값]의 경우가 실험 성과지표가 가장 좋았음

lightGBM

```
In [3]: 1 train_lgb = lgb.Dataset(x_train, label = y_train)
2 val_lgb = lgb.Dataset(x_val, label = y_val)
3
4 params = {'random_seed':seed, 'bagging_seed':seed,
5           'feature_fraction_seed':seed, 'data_random_seed':seed,
6           'drop_seed':seed, # seed 고정
7
8           'boosting_type':'gbdt', 'objective':'regression_l1', # 부스팅 타입과 loss function
9           'bagging_fraction':0.1, 'feature_fraction':0.8, # 샘플링 행과열의 비율
10          'max_bin':300, 'num_leaves':100, # 잎사귀 수
11          'lambda_l1':0.0, 'lambda_l2':15.0, # 정규화 lambda 값
12          'num_iterations':400, 'learning_rate':0.05, # iteration과 학습률
13          'max_depth':-1 # 트리 깊이
14          }
15 model = lgb.train(params, train_lgb, valid_sets = val_lgb,
16                  num_boost_round = 2000, early_stopping_rounds = 100,
17                  verbose_eval = 200)
```

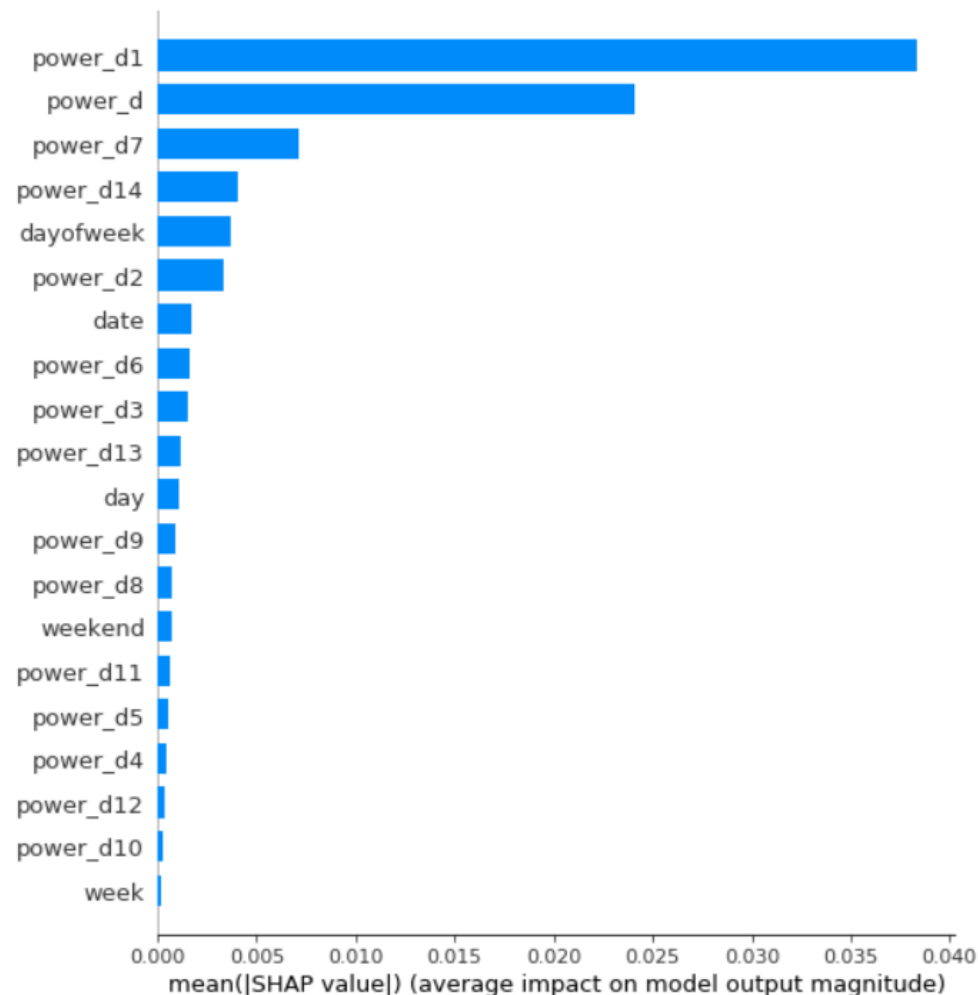


개발 내용 - ⑧ 모델링 평가



속성 중요도

- shap value : 일관성이 반영된 앙상블 모델의 속성 중요도 지표
- 대체적으로, 과거 전력량 데이터가 전력 예측 모델에 활용되는 것을 확인
- **기상 데이터를 제외**하고 분석한 경우가 가장 우수한 결과를 보임





개발 내용 - ⑧ 모델링 평가

⚡ 예측오차 값 (SMAPE) : 12.72

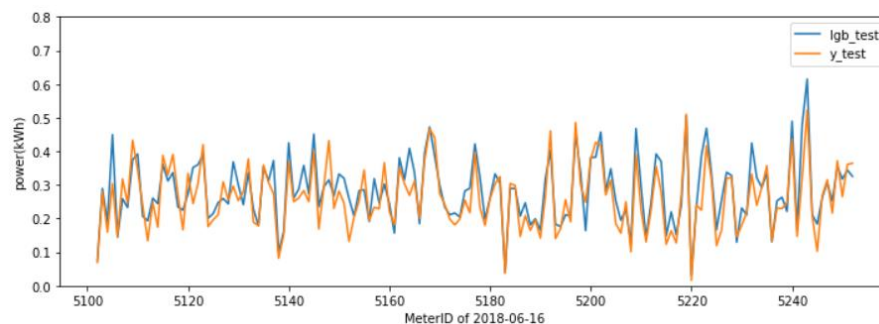
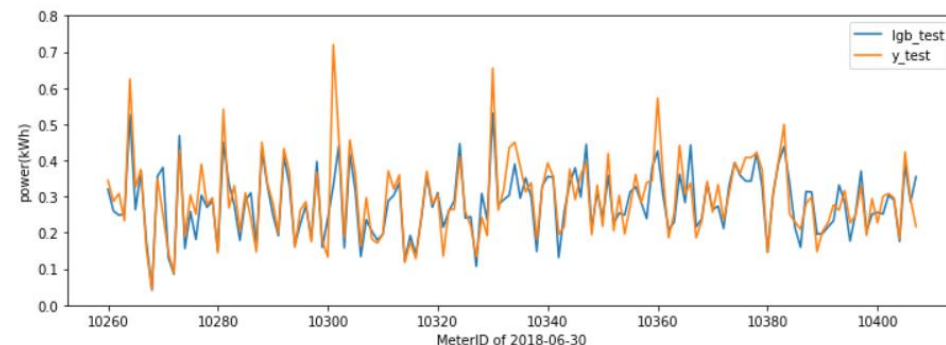
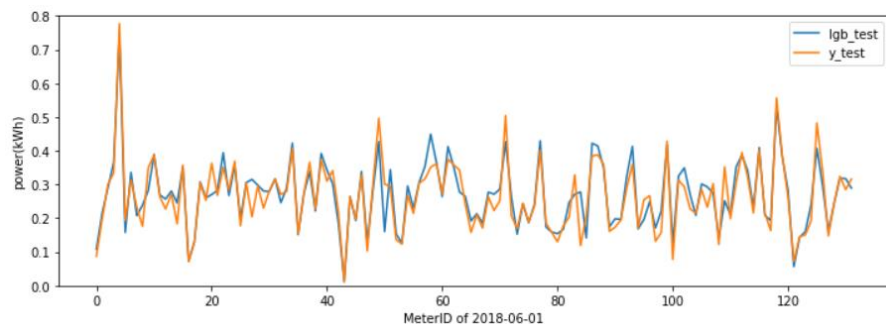
예측 모델	test score (SMAPE)
Decision tree	13.97
Random forest	13.48
Bagging	13.55
Extra tree	18.45
Adaboost	20.06
Catboost	18.99
Deep Neural Network	16.71
Lightgbm (메인 모델)	12.72

<각 모델 오차율 비교>



개발 내용 - ⑧ 모델링 평가

⚡ 시각화 (실제값과 예측값 비교)



<특정 날짜에 대한 예측 결과 비교>



활용 방안

슬기로운 예측생활

전력 수요량 예측모델



활용 방안

Energy Demand Forecasting | Dashboard Example



국가기관 및 기업



개인화 서비스



활용 방안

-생활적 측면:

사회 안전망 서비스



#. A 노인복지센터는 독거노인 B씨 집의 전력사용량이 평소와 달리 급격히 줄자 응급구조단과 함께 B씨의 집을 방문했다. 갑자기 온 복통에 쓰러져 앓고 있던 B씨를 발견, 병원으로 옮겨 큰 일을 막을 수 있었다.



활용 방안

-생활적 측면

스마트 홈 서비스



적용 예시: 피크부하 저감

에너지 소비가 많은 냉장고의 특정 기능(제빙 등)을
(예상된)피크시간대를 피해 가동하는 방식을 통해 2~4%의 에너지
비용을 절감

출처: 미국 ACEEE(American Council for an Energy-Efficient
Economy)의 연구 보고서



활용 방안

-상업적 측면

보도자료 Press Release



배포일시	2014.8.21.(목) 15:00(총 5매)	보도시점	즉시
담당부서	기상산업정보화국 기상융합서비스팀	담당자	팀장 오미림 사무관 김진석
		전화번호	02-2181-0928 02-2181-0416

전력수요 예측 기상 거대자료(빅데이터)로 오차 줄여

- 기상 거대자료(빅데이터) 활용 시, 연간 1,200억 원의 경제적 효과
- 전력수요 예측 오차를 무려 25%나 개선할 수 있어 효과 톡톡히

- 이는 예측 오차를 무려 25% 정도 개선할 수 있는 것으로, 연간 약 1,100GWh 정도의 발전량을 줄일 수 있어 약 1,200억 원의 비용을 절감할 수 있게 된다.[붙임 1]



활용 방안

-상업적 측면

규제샌드박스 심의 통과 사업 현황

자료: 산업통상자원부

사업자	사업	규제 완화 내용
한국전력공사	전력데이터 공유센터	개인정보 가린 전력데이터를 공개해 상권 분석에 활용
	에너지마켓 플레이스	통신 중개업과 판매업을 허용해 전력수요관리 사업자와 수요자를 연결
알에스케어서비스	수동휠체어 전동보조키트	인증 기준을 마련해 전동휠체어보다 싼 휠체어 보조기구 도입
엔에프	중앙집중식 산소발생	순도가 93%로 낮은 편인 산소도 의약품으로 인정해 병원서 활용
정렙코스메틱	프로바이오틱스 화장품	생균을 화장품 원료로 사용

#. 식당을 창업한 K씨는 전력데이터를 활용한 상권 분석의 효과를 톡톡히 보고 있다. 해당 지역의 전주에 부착된 센서에 기록된 유동인구 데이터와 상가의 전력사용량을 종합해 입점을 결정했고, 예상했던 수익을 올리고 있다.

개선점 및 제안

슬기로운 예측생활

전력 수요량 예측모델



개선점 및 제안



기상 데이터의 상관성이 낮은 이유

=> 인천시의 지역적 특성(산업용 비중이 높음)이 반영



따라서 지역별 / 용도별 (주거용, 상업용, 산업용) 모델로 세분화 할 수 있음




결측치 데이터의 영향력 최소화를 위한 기술적 보완점 필요

ex) 센싱 기술력 개선, 센싱 데이터에 적합한 전처리 알고리즘 개발 필요



추후 모델 앙상블, RNN 모델의 도입 고려



Q&A

슬기로운 예측생활

전력 수요량 예측모델