



# 가스 공급량 예측 모델 개발

가스레인지 X 메디치교육센터





# 팀 소개

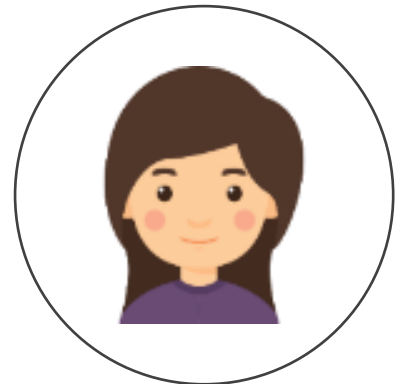
가스레민저 X 메디치교육센터



유현주



정찬웅



서민지



# CONTENTS

---

## 01

### 프로젝트 개요

- 프로젝트 배경
- 프로젝트 목표

## 02

### 데이터 분석

- 데이터 설명
- EDA
- 데이터 전처리
- Feature 추가

## 03

### 모델링 및 분석 결과

- 외부 데이터 Modeling
- 공급량 데이터 Modeling
- 성능평가
- 결론
- 시사점 및 한계점

## 04

### 프로젝트 수행 소감

- 느낌 및 소감



# CONTENTS

---

## 01

프로젝트 개요



# 01-1 프로젝트 배경

가스레인지 X 메디치교육센터



## 에너지 대란 나비효과...물류 이어 곡물값에도 영향

질소·요소 등 주요 생산원료 천연가스·석탄 대란에 급등  
북미 비료가격 사상 최고치...애그플레이션 더 심해질듯



출처 : <https://view.asiae.co.kr/article/20211108102137914-21>

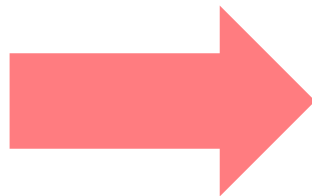


# 01-1 프로젝트 배경

가스레인지 X 메디치교육센터



에너지 공급  
=> 전 가치 사슬에 영향을 미침



4차 산업에 있어서  
에너지 공급량 예측 필요



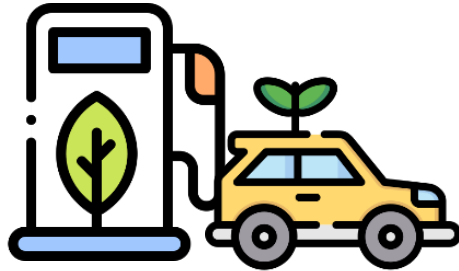
# 01-1 프로젝트 배경

가스레인지 X 메디치교육센터



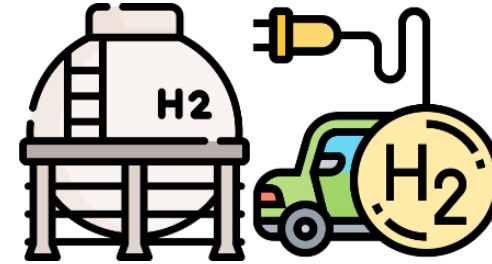
## - 천연가스의 역할

1)



화석연료 중  
가장 환경 친화적

2)



수소에너지원 추출 및  
탄소중립 사회로의 발판 구축



# 01-1 프로젝트 배경

가스레인지 X 메디치교육센터



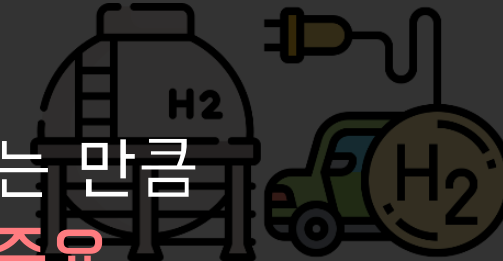
## - 천연가스의 역할

1)



화석연료 중  
가장 환경 친화적

2)



수소에너지원 추출 및  
탄소중립 사회로의 발판 구축

천연가스의 역할이 대두되는 만큼  
**천연가스 공급량 예측 중요**





# 01-2 프로젝트 목표

가스레인저 X 메디치교육센터



⇒ 천연가스 공급량 예측 모델 구축



# CONTENTS

---

## 02

데이터 분석



## 02-1 데이터 설명

가스레인지 X 메디치교육센터



### - 원본데이터

|        | 연월일        | 시간  | 구분  | 공급량         |
|--------|------------|-----|-----|-------------|
| 0      | 2013-01-01 | 1   | A   | 2497.129000 |
| 1      | 2013-01-01 | 2   | A   | 2363.265000 |
| 2      | 2013-01-01 | 3   | A   | 2258.505000 |
| 3      | 2013-01-01 | 4   | A   | 2243.969000 |
| 4      | 2013-01-01 | 5   | A   | 2344.105000 |
| ...    | ...        | ... | ... | ...         |
| 368083 | 2018-12-31 | 20  | H   | 681.033000  |
| 368084 | 2018-12-31 | 21  | H   | 669.961000  |
| 368085 | 2018-12-31 | 22  | H   | 657.941000  |
| 368086 | 2018-12-31 | 23  | H   | 610.953000  |
| 368087 | 2018-12-31 | 24  | H   | 560.896000  |

연월일 : 2013 ~ 2018년도의 데이터

시간 : 1 ~ 24

구분 : A, B, C, D, E, G, H (총 7개)

공급량 : 원자료의 계량단위는  $m^3$ ,  $ton$ , 혹은 MJ(열량단위)

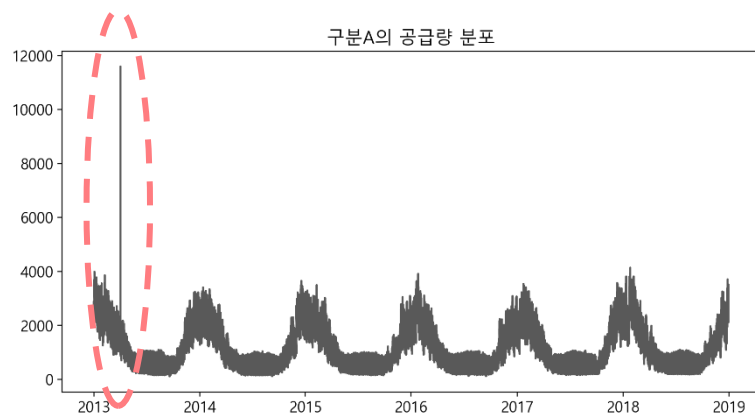


## 02-2 EDA

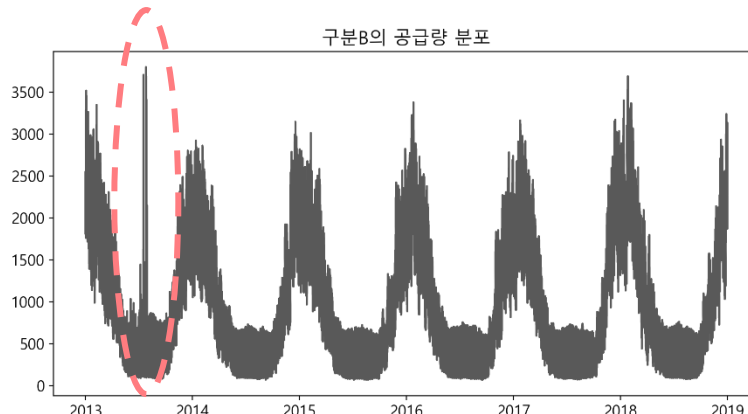


### - 구분(공급사) 별 공급량 확인

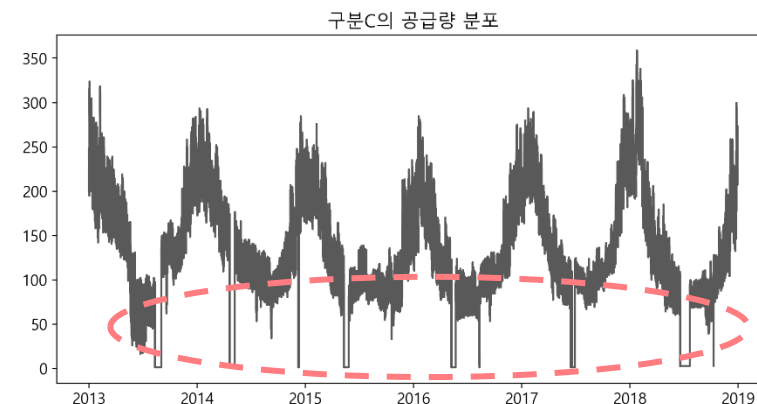
: 구분 A, B, C



구분 A는 튀는 값을 제외하고,  
일정한 U 모양을 그림  
튀는 값(이상치)는 2013년 4월 3일 경 존재



구분 B는 튀는 값을 제외하고,  
일정한 U 모양을 그림  
튀는 값(이상치)는 2013년 7월 26일 경 존재



구분 C의 분포는 불규칙한 모양을 그림  
타 구분에 비해 크기의 분포가 10배정도 작으며  
불규칙한 튀는 값 존재



## 02-2 EDA

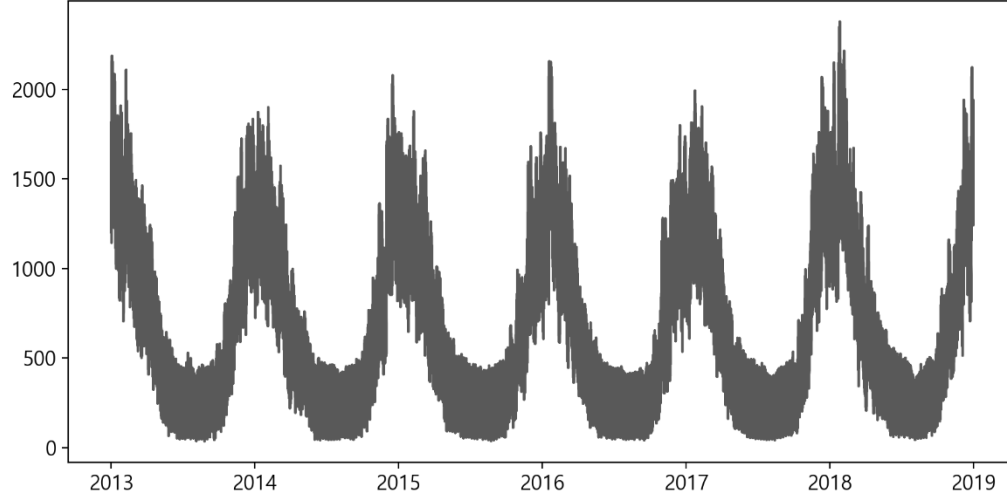
가스레인지 X 메디치교육센터



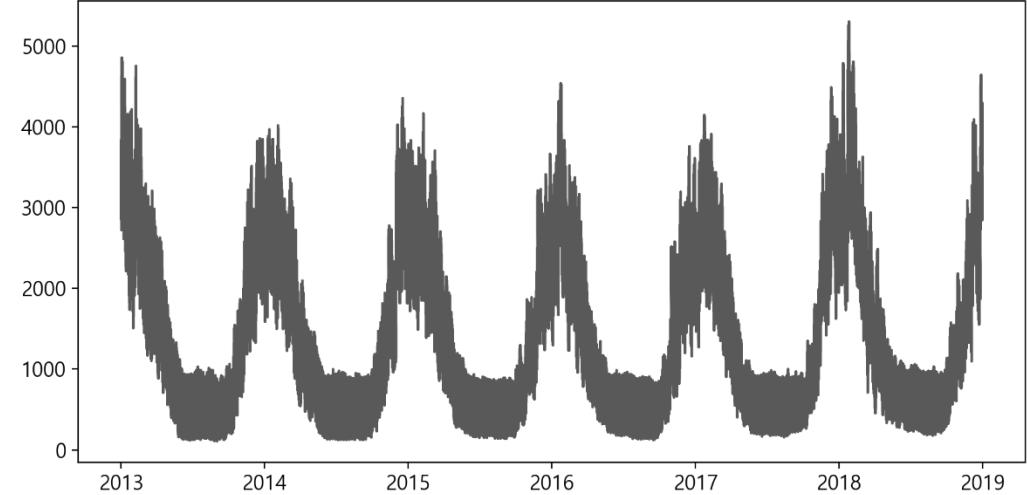
### - 구분(공급사) 별 공급량 확인

: 구분 D, E

구분D의 공급량 분포



구분E의 공급량 분포



구분 D와 E의 경우 튀는 값 없이 공급량의 분포가 일정한 U 모양을 그림



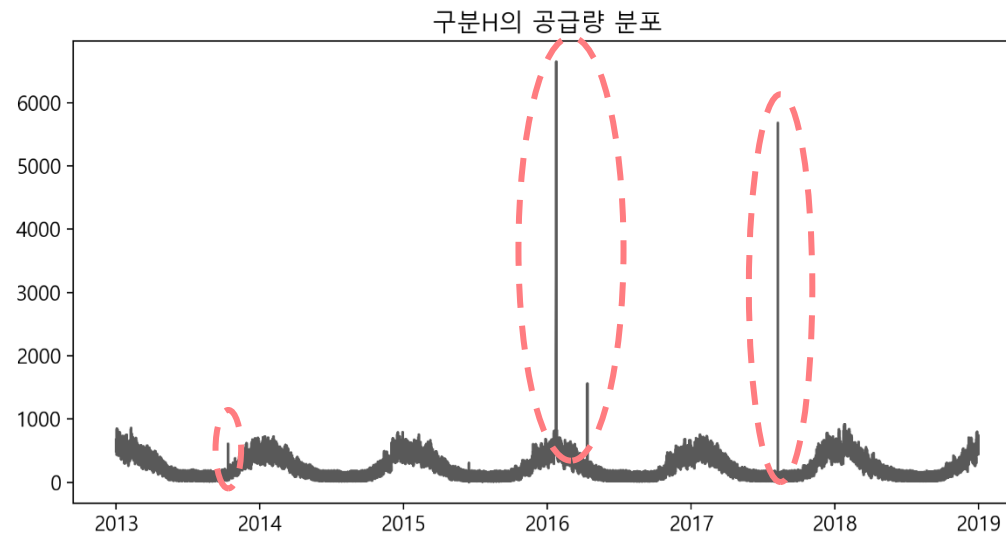
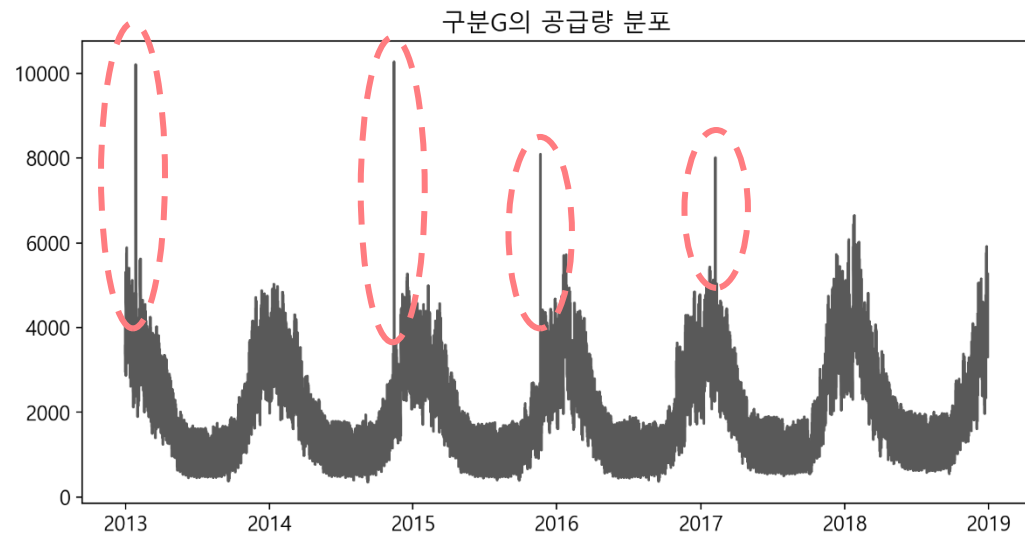
## 02-2 EDA

가스레인지 X 메디치교육센터



### - 구분(공급사) 별 공급량 확인

: 구분 G, H



구분 G와 H의 경우 튀는 몇 개의 값들을 제외하고 일정한 U모양을 그림



## 02-3 데이터 전처리

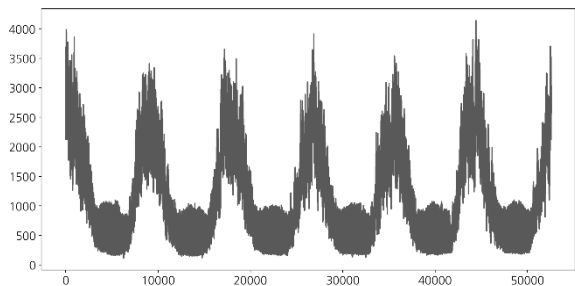
가스레인지 X 메디치교육센터



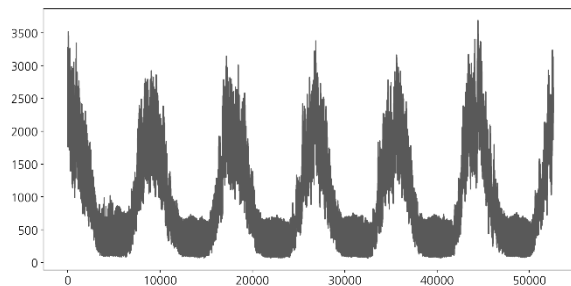
### - 공급량 이상치 처리

: 1시간 전후 값과의 차이를 기준으로 튀는 값을 선별하고 평균, 선형보간, prophet 등 다양한 방법으로 이상치 처리 시도  
: 가장 이상치 처리가 잘 된 선형 보간법의 결과

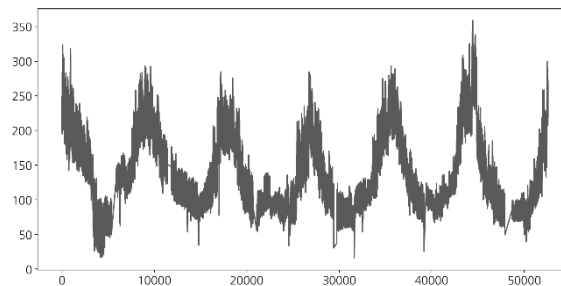
구분 A



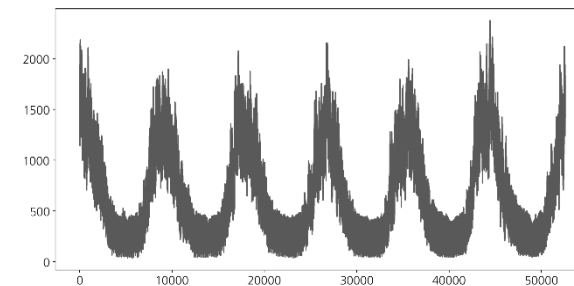
구분 B



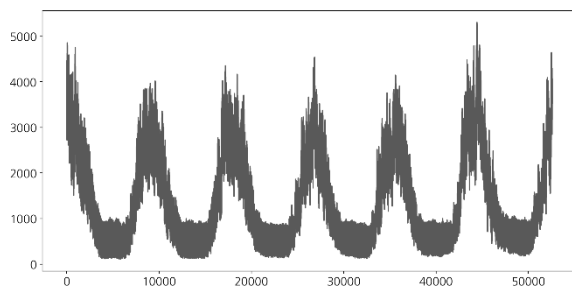
구분 C



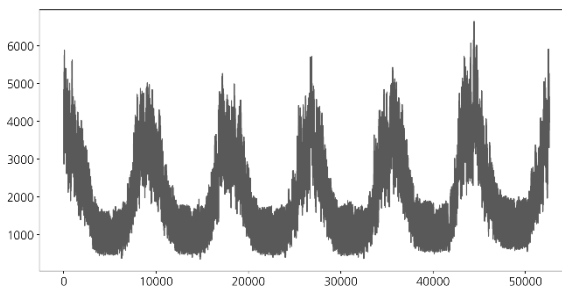
구분 D



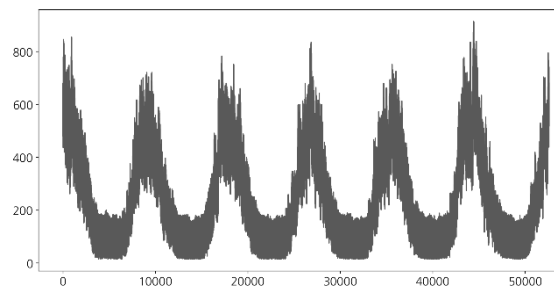
구분 E



구분 G



구분 H





## 02-4 Feature 추가

가스레인저 X 메디치교육센터



### - 파생변수

날짜 파생변수 : 연, 월, 일, 요일, 분기, 음력\_분기, 계절, 음력 월, 음력 일, 절기, 평/휴일 여부

|        | 연월일        | 시간  | 구분  | 공급량      | year | month | day | weekday | lunar      | lun_month | lun_day | quarter | lun_quarter | season | 절기명 | 휴일여부 |
|--------|------------|-----|-----|----------|------|-------|-----|---------|------------|-----------|---------|---------|-------------|--------|-----|------|
| 0      | 2013-01-01 | 1   | 0   | 2497.129 | 2013 | 1     | 1   | 1       | 2012-11-20 | 11        | 20      | 1       | 4           | winter | 소한  | Y    |
| 1      | 2013-01-01 | 2   | 0   | 2363.265 | 2013 | 1     | 1   | 1       | 2012-11-20 | 11        | 20      | 1       | 4           | winter | 소한  | Y    |
| 2      | 2013-01-01 | 3   | 0   | 2258.505 | 2013 | 1     | 1   | 1       | 2012-11-20 | 11        | 20      | 1       | 4           | winter | 소한  | Y    |
| 3      | 2013-01-01 | 4   | 0   | 2243.969 | 2013 | 1     | 1   | 1       | 2012-11-20 | 11        | 20      | 1       | 4           | winter | 소한  | Y    |
| 4      | 2013-01-01 | 5   | 0   | 2344.105 | 2013 | 1     | 1   | 1       | 2012-11-20 | 11        | 20      | 1       | 4           | winter | 소한  | Y    |
| ...    | ...        | ... | ... | ...      | ...  | ...   | ... | ...     | ...        | ...       | ...     | ...     | ...         | ...    | ... | ...  |
| 383203 | 2019-03-31 | 20  | 6   | NaN      | 2019 | 3     | 31  | 6       | 2019-02-25 | 2         | 25      | 1       | 1           | spring | 청명  | N    |
| 383204 | 2019-03-31 | 21  | 6   | NaN      | 2019 | 3     | 31  | 6       | 2019-02-25 | 2         | 25      | 1       | 1           | spring | 청명  | N    |
| 383205 | 2019-03-31 | 22  | 6   | NaN      | 2019 | 3     | 31  | 6       | 2019-02-25 | 2         | 25      | 1       | 1           | spring | 청명  | N    |
| 383206 | 2019-03-31 | 23  | 6   | NaN      | 2019 | 3     | 31  | 6       | 2019-02-25 | 2         | 25      | 1       | 1           | spring | 청명  | N    |
| 383207 | 2019-03-31 | 24  | 6   | NaN      | 2019 | 3     | 31  | 6       | 2019-02-25 | 2         | 25      | 1       | 1           | spring | 청명  | N    |

383208 rows x 16 columns





## 02-4 Feature 추가

가스레인지 X 메디치교육센터



### - 파생변수

공급량 파생변수 : 전년도 월별 공급량 평균/표준편차, 공급량 lagging

|        | 연월일        | 시간  | 구분  | 공급량      | supply_2160 | supply_8760 | 전년도_월별공급량_mean | 전년도_월별공급량_std |
|--------|------------|-----|-----|----------|-------------|-------------|----------------|---------------|
| 0      | 2013-01-01 | 1   | 0   | 2497.129 | NaN         | NaN         | NaN            | NaN           |
| 1      | 2013-01-01 | 2   | 0   | 2363.265 | NaN         | NaN         | NaN            | NaN           |
| 2      | 2013-01-01 | 3   | 0   | 2258.505 | NaN         | NaN         | NaN            | NaN           |
| 3      | 2013-01-01 | 4   | 0   | 2243.969 | NaN         | NaN         | NaN            | NaN           |
| 4      | 2013-01-01 | 5   | 0   | 2344.105 | NaN         | NaN         | NaN            | NaN           |
| ...    | ...        | ... | ... | ...      | ...         | ...         | ...            | ...           |
| 383203 | 2019-03-31 | 20  | 6   | NaN      | 681.033     | 244.162     | 320.007126     | 97.041542     |
| 383204 | 2019-03-31 | 21  | 6   | NaN      | 669.961     | 248.059     | 320.007126     | 97.041542     |
| 383205 | 2019-03-31 | 22  | 6   | NaN      | 657.941     | 231.181     | 320.007126     | 97.041542     |
| 383206 | 2019-03-31 | 23  | 6   | NaN      | 610.953     | 199.022     | 320.007126     | 97.041542     |
| 383207 | 2019-03-31 | 24  | 6   | NaN      | 560.896     | 190.212     | 320.007126     | 97.041542     |

383208 rows x 8 columns



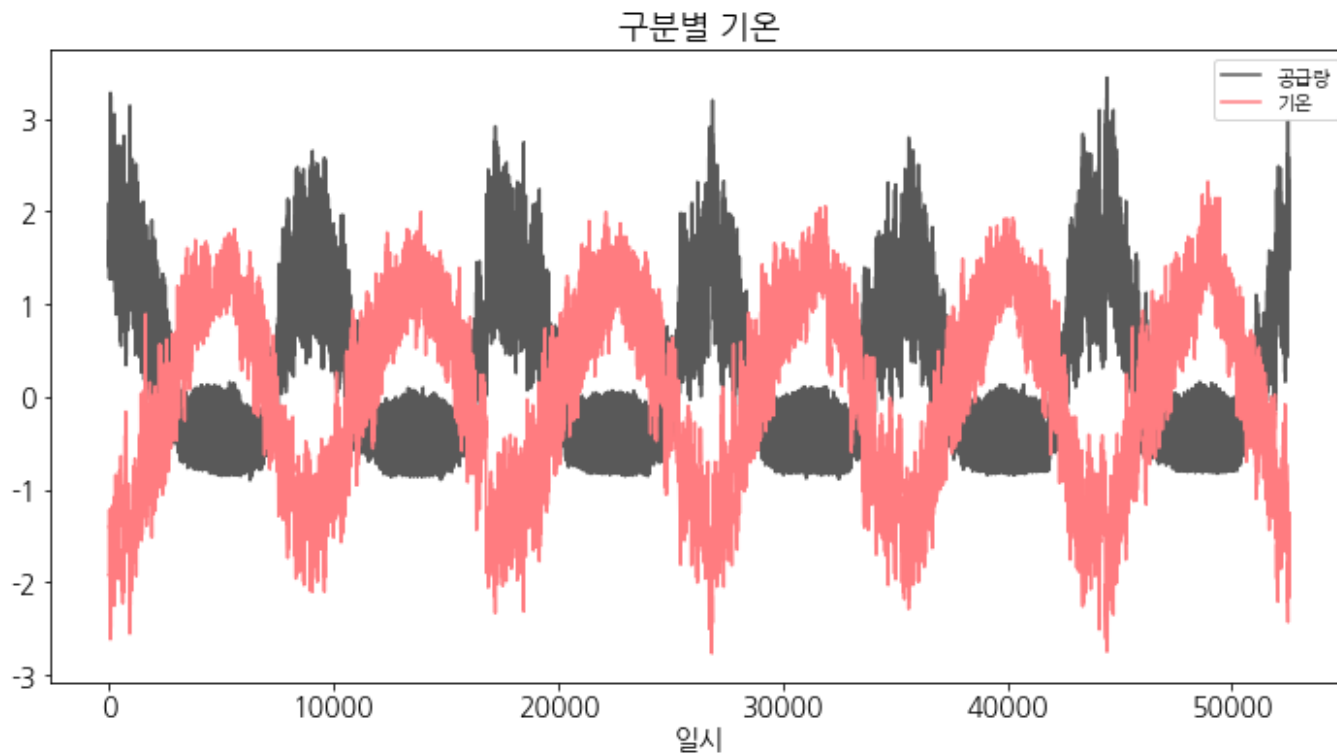
## 02-4 Feature 추가

가스레인지 X 메디치교육센터



### ① 외부데이터 - 기온

- : 기온과 공급량이 음의 상관관계성을 보임
- : 주어진 공급량 데이터의 출처 지역을 알 수 없기 때문에 인구가 제일 많은 **서울지역의 기온**을 사용하기로 결정
- : 결측치의 경우, 선형 보간법으로 대체



\* 공급량과 기온의 차이가 크기 때문에 모두 스케일링 하여 나타냄



## 02-4 Feature 추가

가스레인지 X 메디치교육센터



### ② 외부데이터 - 기온 반응(수도권), 산업 기온 반응(서울) 함수

|       | 기온   | 기온반응도   | 산업_기온반응도 |
|-------|------|---------|----------|
| 0     | -8.5 | 0.81834 | 0.28609  |
| 1     | -8.4 | 0.82012 | 0.28578  |
| 2     | -8.1 | 0.82544 | 0.28469  |
| 3     | -8.2 | 0.82367 | 0.28508  |
| 4     | -8.2 | 0.82367 | 0.28508  |
| ...   | ...  | ...     | ...      |
| 54739 | 4.9  | 0.69123 | 0.08023  |
| 54740 | 4.2  | 0.72546 | 0.09372  |
| 54741 | 3.6  | 0.75204 | 0.10547  |
| 54742 | 3.0  | 0.77606 | 0.11733  |
| 54743 | 2.6  | 0.79064 | 0.12528  |

54744 rows x 3 columns

- 기온반응함수, 산업 기온 반응 함수
- : 기온이 도시가스수요에 미치는 영향을  
각 개별 기온대별로 나타내어 일종의 함수 형태로 제시한 것  
⇒ 구분별 기온에 따라 기온반응함수와 산업 기온 반응 함수 추가



## 02-4 Feature 추가

가스레인지 X 메디치교육센터



### ③ 외부데이터 - 난방지수

: 한국지역난방공사의 난방지수(최근 1년간 시간당 최대 열공급 실적을 1000으로 두고 지수화한 값)에 대한 정보를 날짜별, 시간대별, 지사별 난방지수

|        | 일시                  | 기온   | 기온반응도   | 산업_기온반응도 | 난방지수    |
|--------|---------------------|------|---------|----------|---------|
| 0      | 2013-01-01 00:00:00 | -8.5 | 0.81834 | 0.28609  | 10695.0 |
| 1      | 2013-01-01 01:00:00 | -8.4 | 0.82012 | 0.28578  | 10186.0 |
| 2      | 2013-01-01 02:00:00 | -8.1 | 0.82544 | 0.28469  | 9765.0  |
| 3      | 2013-01-01 03:00:00 | -8.2 | 0.82367 | 0.28508  | 9468.0  |
| 4      | 2013-01-01 04:00:00 | -8.2 | 0.82367 | 0.28508  | 9378.0  |
| ...    | ...                 | ...  | ...     | ...      | ...     |
| 368083 | 2018-12-31 19:00:00 | -3.7 | 0.87942 | 0.24080  | 10267.0 |
| 368084 | 2018-12-31 20:00:00 | -4.6 | 0.87371 | 0.25341  | 10752.0 |
| 368085 | 2018-12-31 21:00:00 | -5.4 | 0.86584 | 0.26323  | 11190.0 |
| 368086 | 2018-12-31 22:00:00 | -5.2 | 0.86803 | 0.26091  | 10962.0 |
| 368087 | 2018-12-31 23:00:00 | -5.5 | 0.86469 | 0.26436  | 10529.0 |



## 02-4 Feature 추가

가스레인지 X 메디치교육센터



### ④ 외부데이터 - 천연가스의 수입물가지수, 수입금액지수, 수입물량지수

- : 천연가스 해외 수입 의존도가 높기 때문에 수입 상품의 종합적인 가격수준을 측정하는 통계지표 3가지 사용
- : 한국은행 경제통계시스템으로부터 월별 수입물가, 수입금액, 수입물량지수 추출

|     | 일자         | 수입물량지수_천연가스 | 수입금액지수_천연가스 | 수입물가지수(원화) | 소비량  |
|-----|------------|-------------|-------------|------------|------|
| 0   | 2013-01-01 | 138.70      | 193.49      | 143.53     | 6487 |
| 1   | 2013-02-01 | 142.15      | 204.18      | 148.58     | 5291 |
| 2   | 2013-03-01 | 144.35      | 210.42      | 151.37     | 4781 |
| 3   | 2013-04-01 | 113.39      | 166.03      | 151.01     | 4348 |
| 4   | 2013-05-01 | 98.08       | 140.76      | 148.94     | 3612 |
| ... | ...        | ...         | ...         | ...        | ...  |
| 70  | 2018-11-01 | 137.00      | 145.58      | 115.62     | 4772 |
| 71  | 2018-12-01 | 162.74      | 172.60      | 114.82     | 6381 |
| 72  | 2019-01-01 | 132.04      | 143.13      | 117.27     | 6580 |
| 73  | 2019-02-01 | 133.88      | 148.47      | 120.01     | 5339 |
| 74  | 2019-03-01 | 96.80       | 99.68       | 112.27     | 4984 |

75 rows × 5 columns



## 02-4 Feature 추가

가스레인저 X 메디치교육센터



### ⑤ 외부데이터 - 천연가스 에너지 소비량

- : 천연가스 소비량은 수요를 나타내는 한 지표이기에 공급량 예측에 영향을 줄 것이라 판단
- : 한국은행 경제통계시스템으로부터 월 별 천연가스 에너지 소비량 데이터 추출

|     | 일자         | 수입물량지수_천연가스 | 수입금액지수_천연가스 | 수입물가지수(원화) | 소비량  |
|-----|------------|-------------|-------------|------------|------|
| 0   | 2013-01-01 | 138.70      | 193.49      | 143.53     | 6487 |
| 1   | 2013-02-01 | 142.15      | 204.18      | 148.58     | 5291 |
| 2   | 2013-03-01 | 144.35      | 210.42      | 151.37     | 4781 |
| 3   | 2013-04-01 | 113.39      | 166.03      | 151.01     | 4348 |
| 4   | 2013-05-01 | 98.08       | 140.76      | 148.94     | 3612 |
| ... | ...        | ...         | ...         | ...        | ...  |
| 70  | 2018-11-01 | 137.00      | 145.58      | 115.62     | 4772 |
| 71  | 2018-12-01 | 162.74      | 172.60      | 114.82     | 6381 |
| 72  | 2019-01-01 | 132.04      | 143.13      | 117.27     | 6580 |
| 73  | 2019-02-01 | 133.88      | 148.47      | 120.01     | 5339 |
| 74  | 2019-03-01 | 96.80       | 99.68       | 112.27     | 4984 |

75 rows x 5 columns



## CONTENTS

---

# 03

모델링 및 분석 결과



## 03-1 외부데이터 Modeling

가스레인지 X 메디치교육센터



### - 외부 데이터 예측 모델

- : 외부 데이터 활용 시 2019년 이후 데이터를 활용하는 것은 data leakage에 해당
- : 2018년 12월 31일 이전 데이터를 활용하여 예측을 진행한 후 사용

#### ○ 활용한 모델





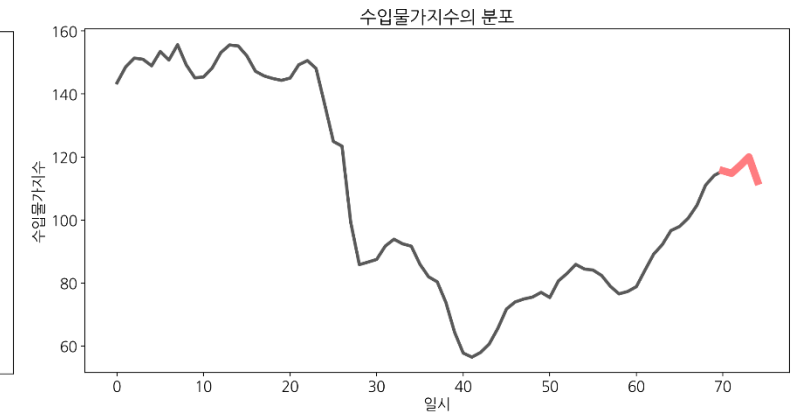
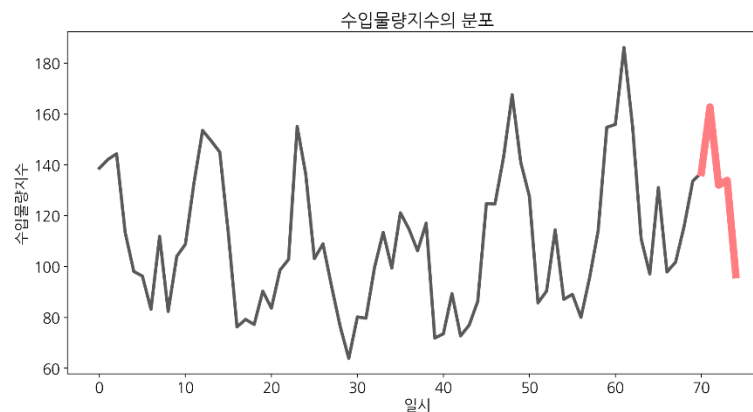
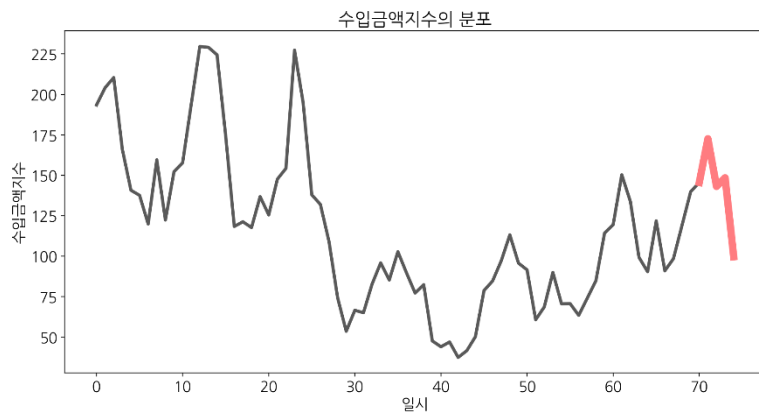
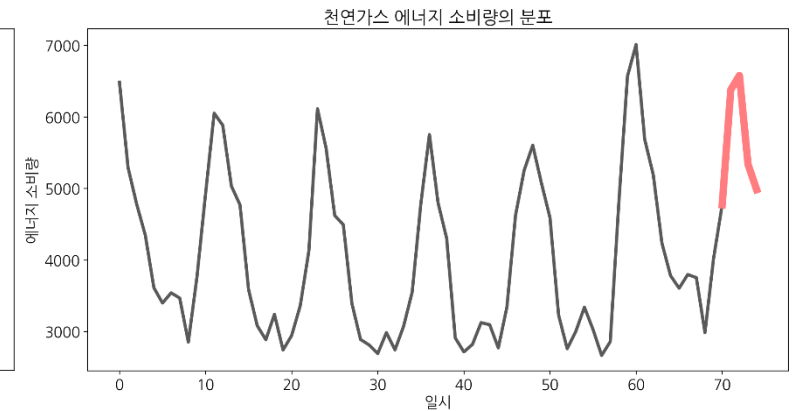
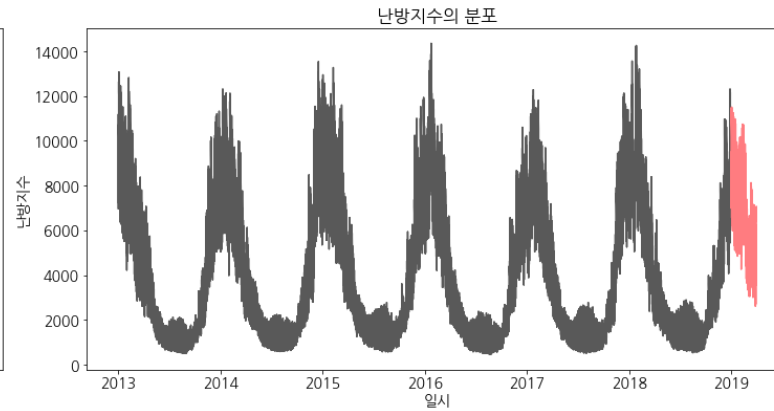
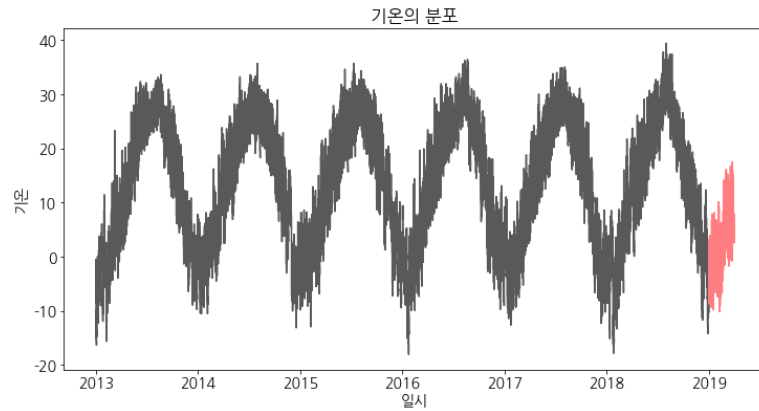


# 03-1 외부데이터 Modeling

가스레인지 X 메디치교육센터



## - 외부 데이터 예측 결과





## 03-2 공급량 Modeling

가스레인저 X 메디치교육센터

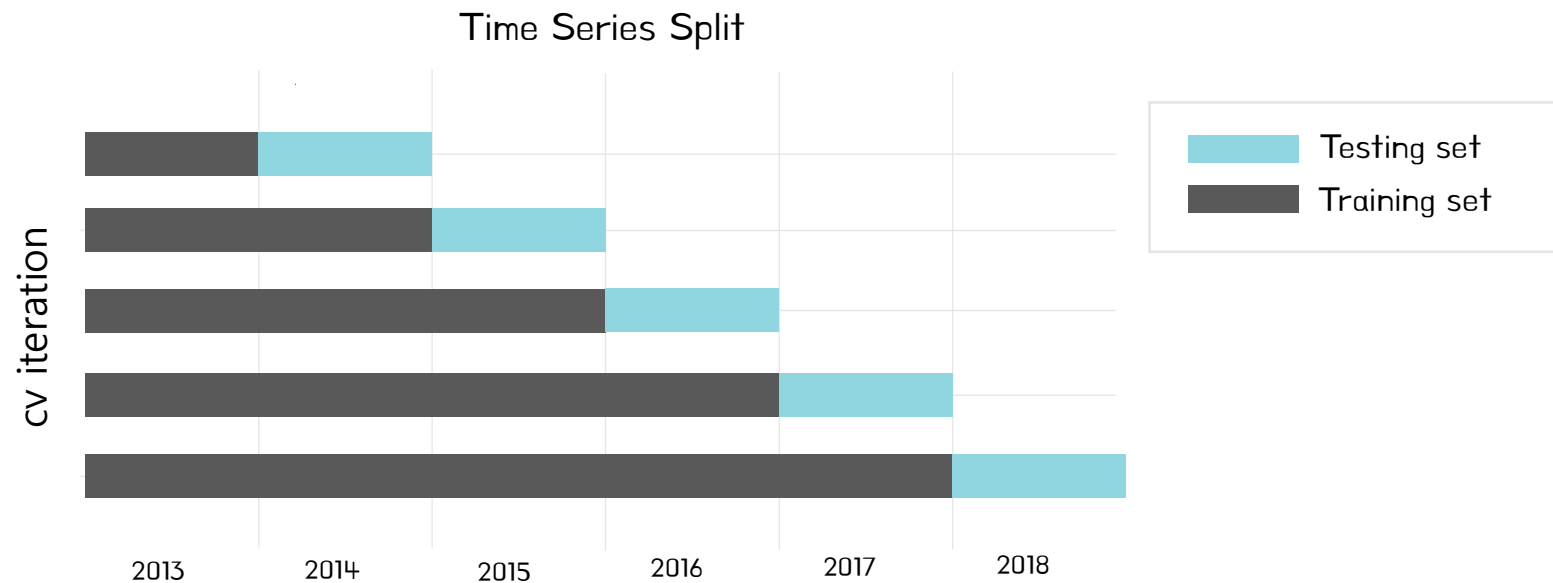


### - 구분 별 모델링 구축

: 구분에 따라 공급량이 다르기 때문에 구분별로 모델링을 진행

### - Cross Validation

: 과적합 방지를 위해 cross validation을 연도별로 진행





## 03-2 공급량 Modeling

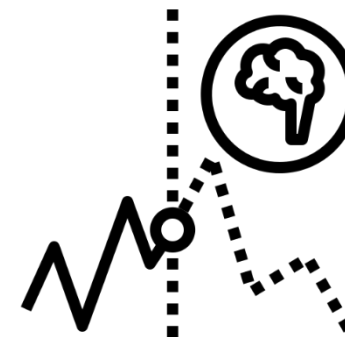
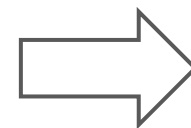
가스레인지 X 메디치교육센터



### - 성능 평가 지수

$$NMAE = \frac{1}{N} \sum_m \frac{|true_i - predict_i|}{true_i}$$

### - 공급량 예측 모델





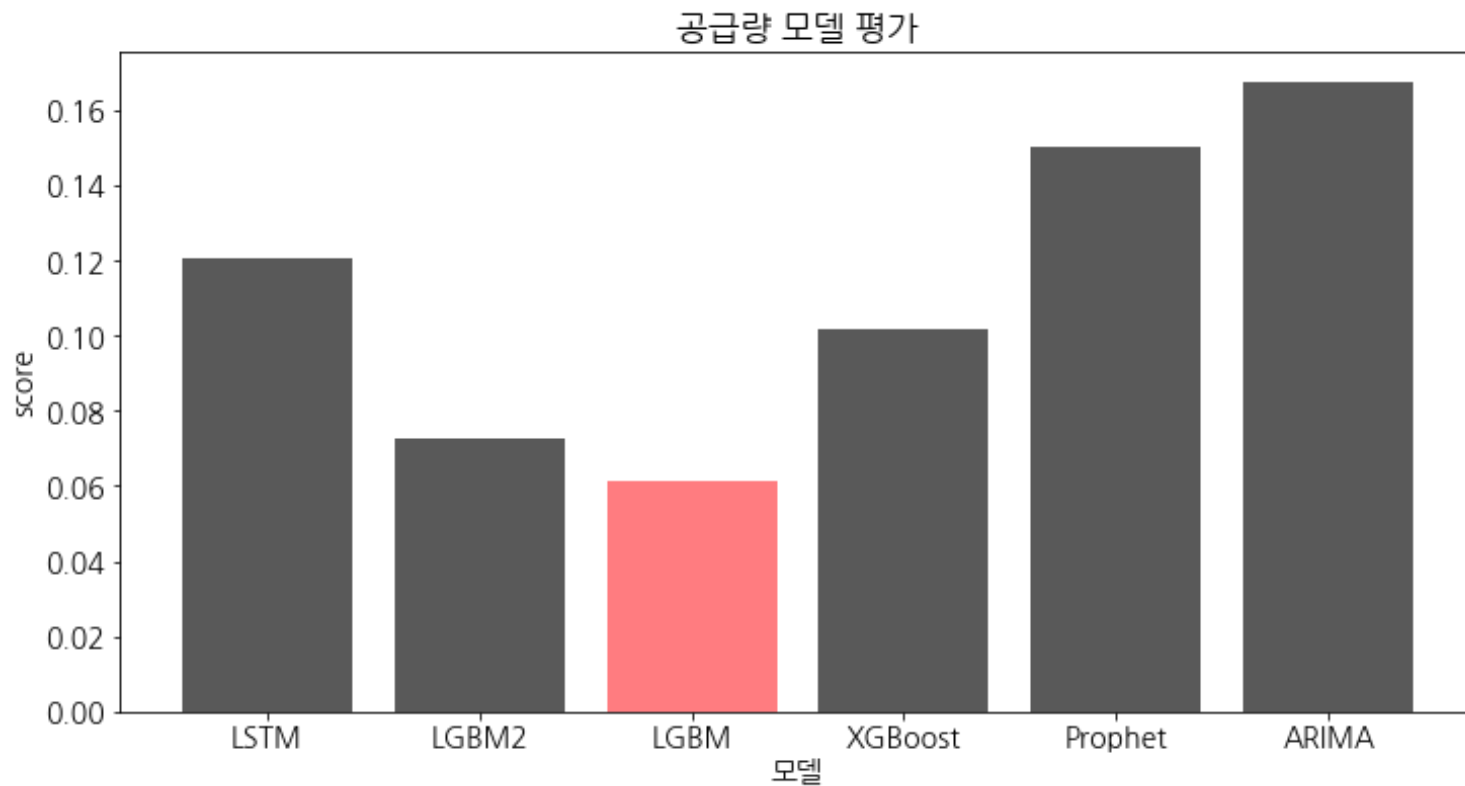
## 03-3 성능평가

가스레인지 X 메디치교육센터



### - 성능 평가

: Feature Selection을 진행하며 각각의 모델로 모델링 진행





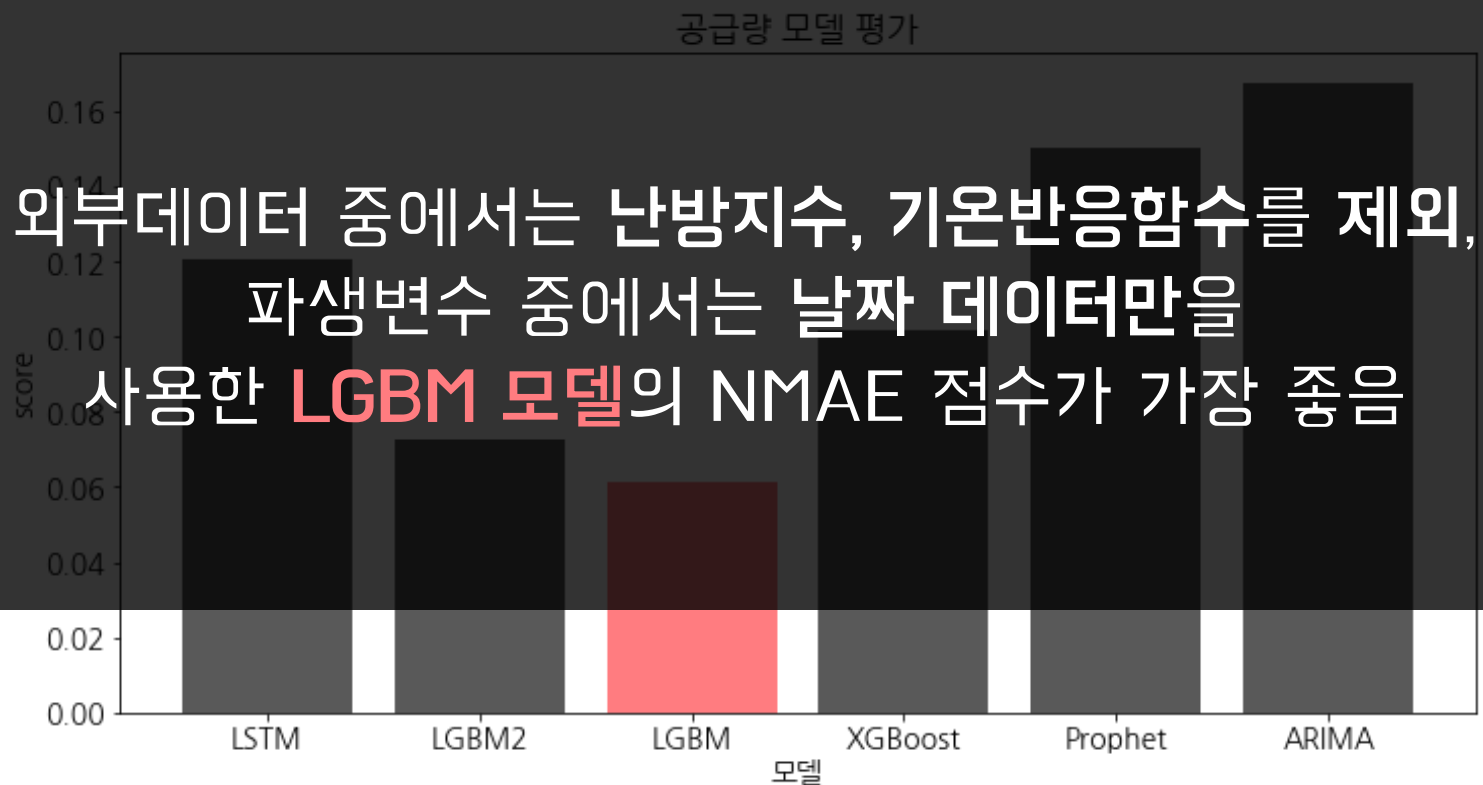
## 03-3 성능평가

가스레인지 X 메디치교육센터



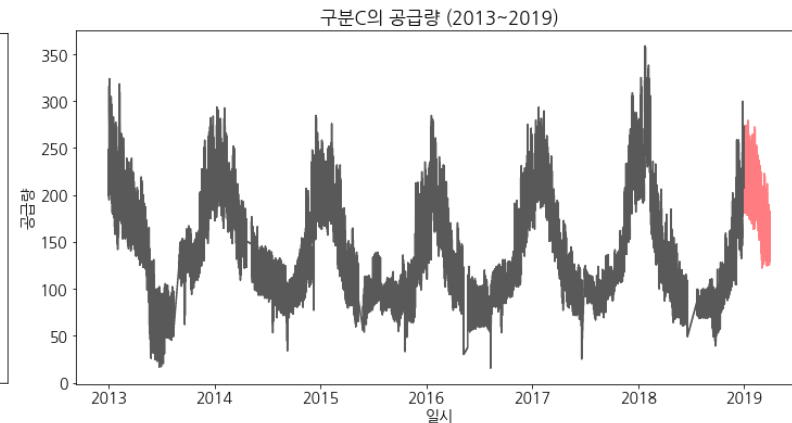
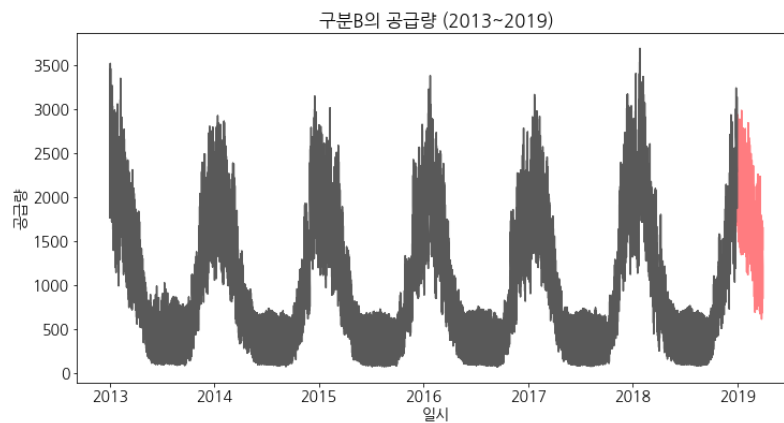
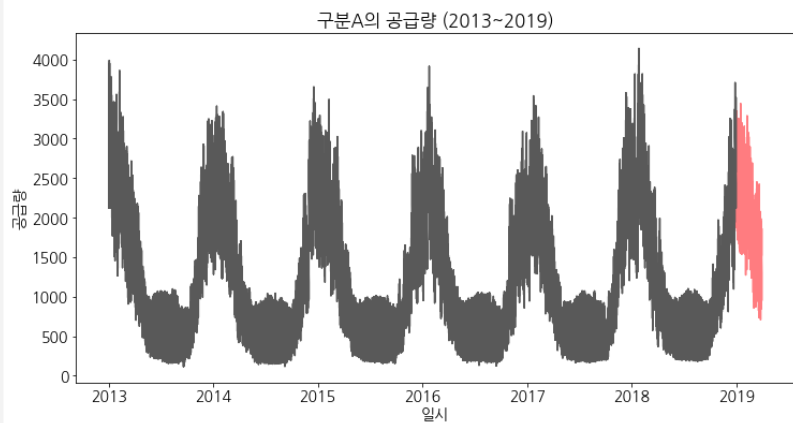
### - 성능 평가

: Feature Selection을 진행하며 각각의 모델로 모델링 진행





## - LGBM 공급량 모델 예측 결과



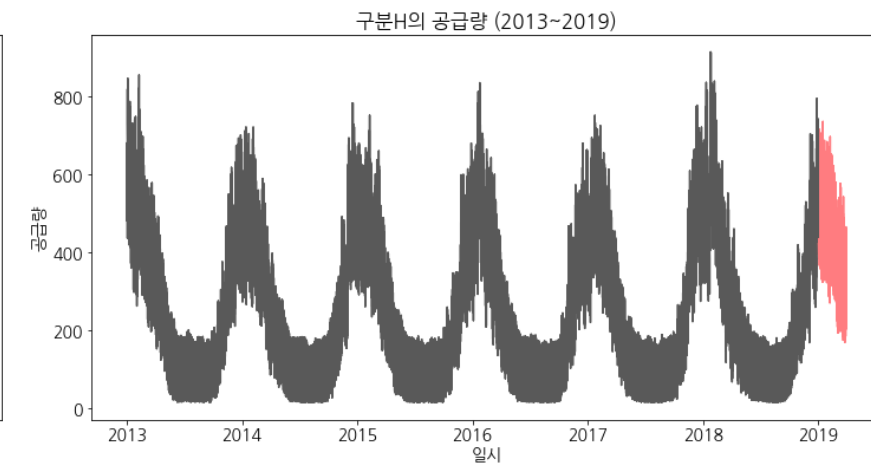
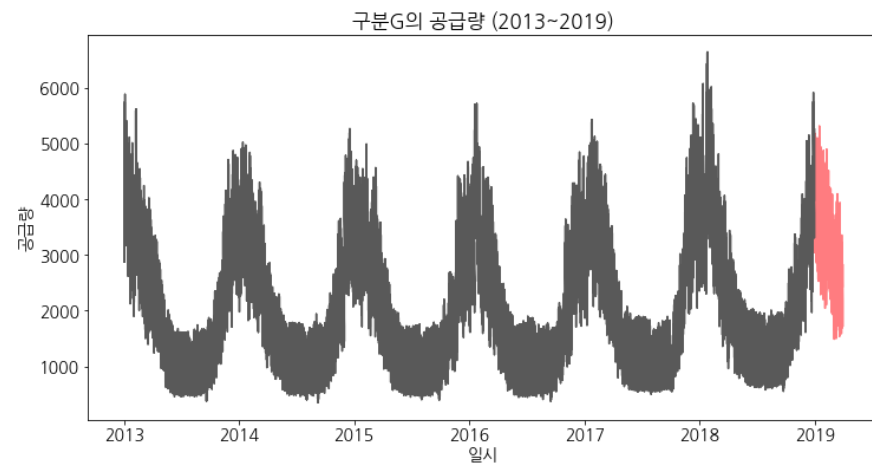
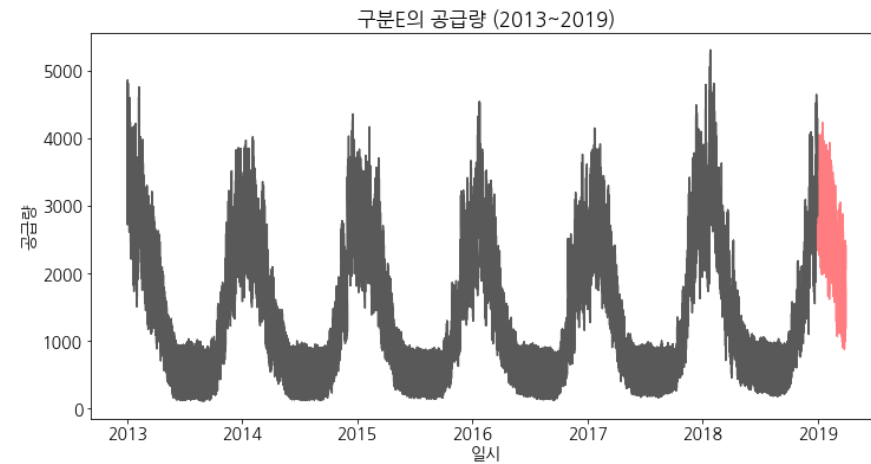
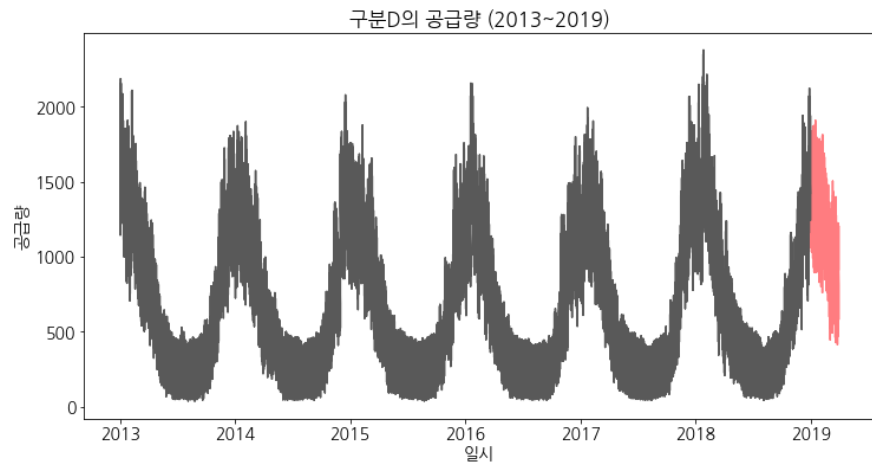


## 03-4 결론

가스레인지 X 메디치교육센터



### - LGBM 공급량 모델 예측 결과



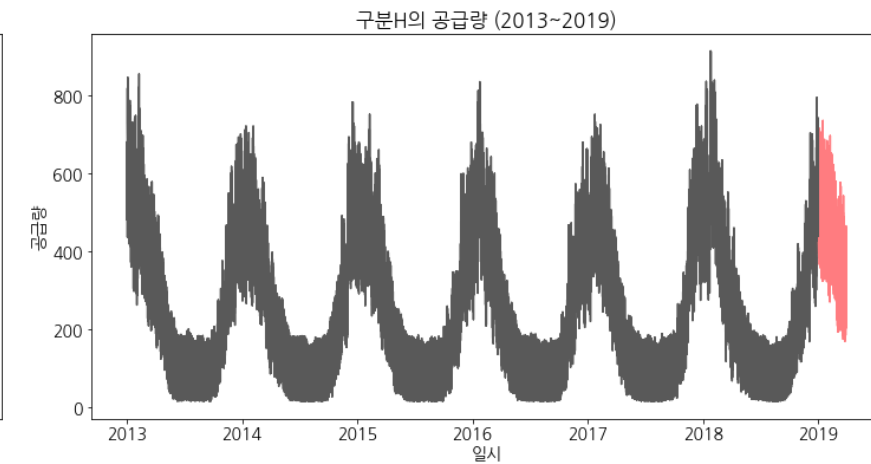
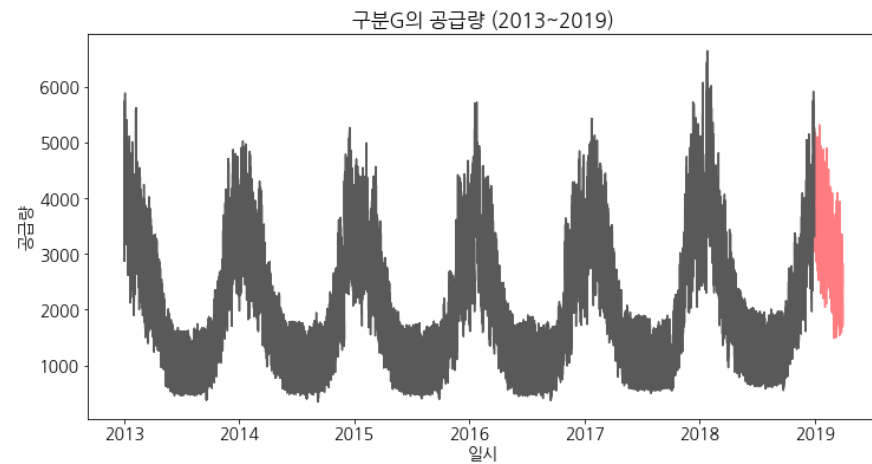
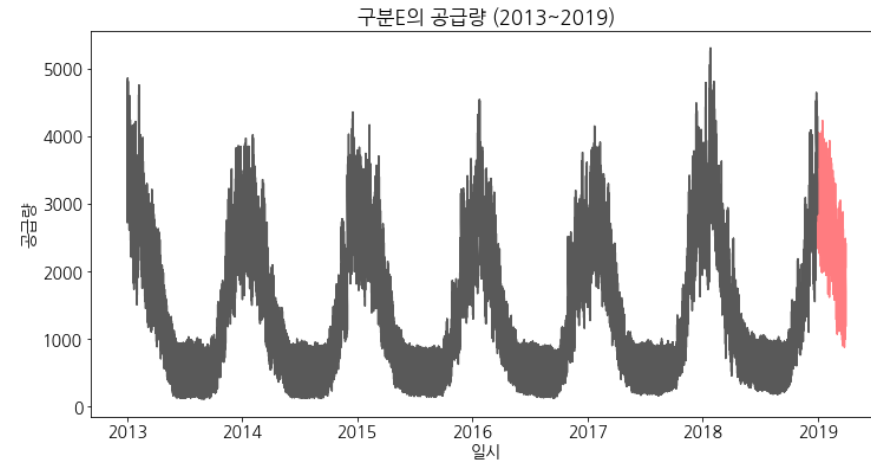
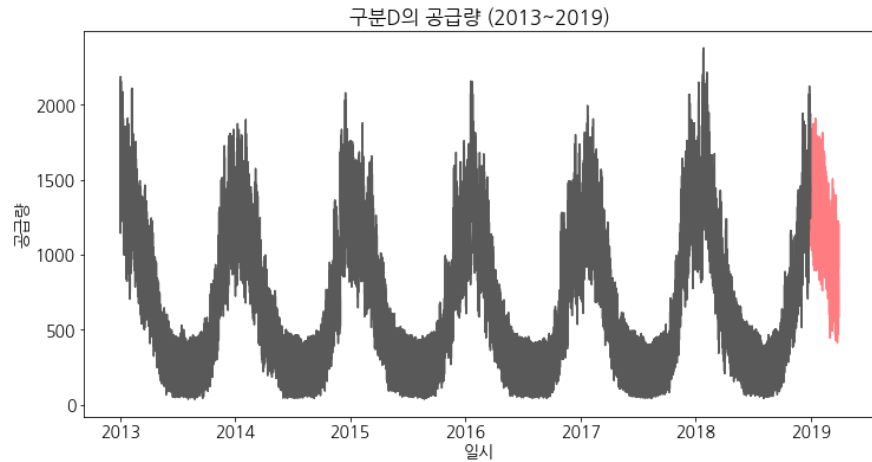


## 03-4 결론

가스레인지 X 메디치교육센터



### - LGBM 공급량 모델 예측 결과





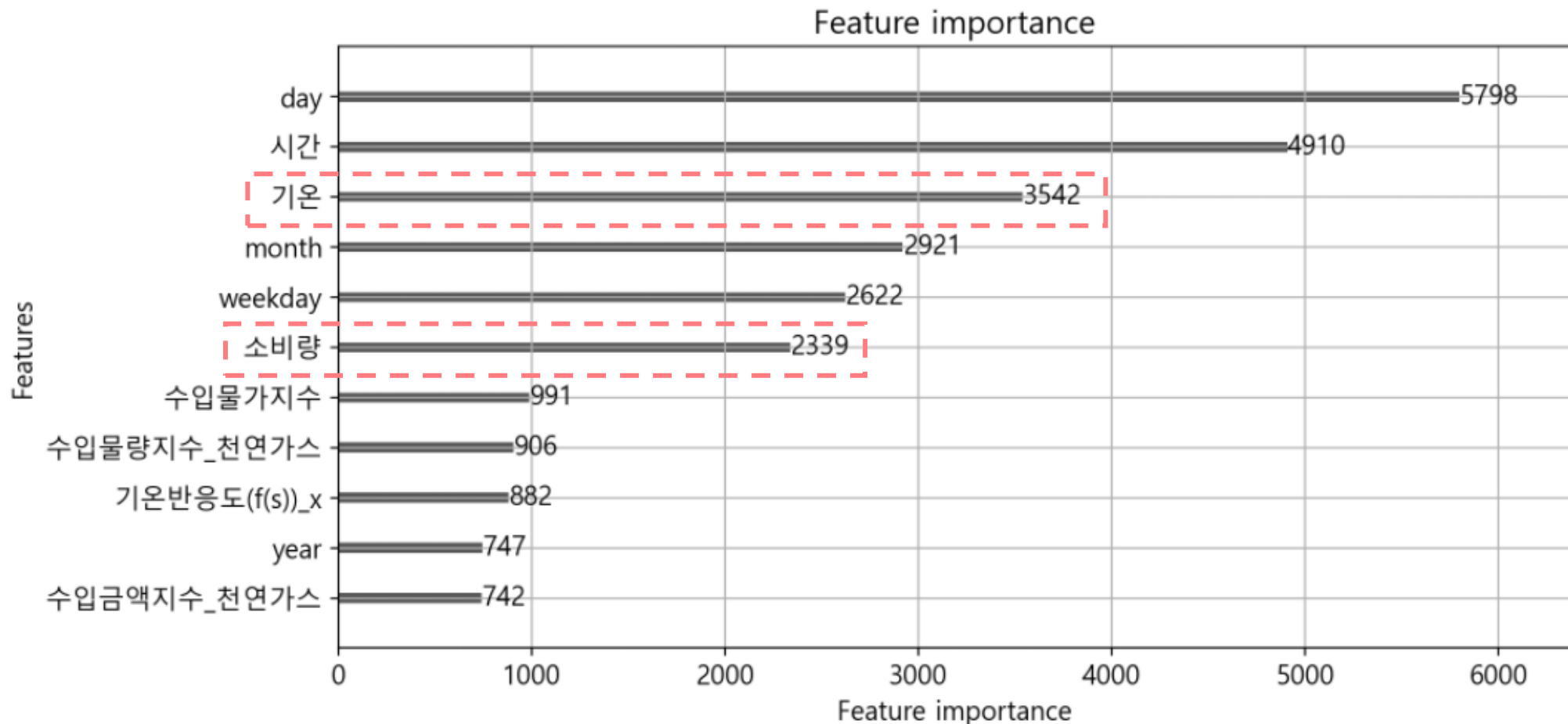


## 03-5 Feature Importance

가스레인지 X 메디치교육센터



### - Feature Importance



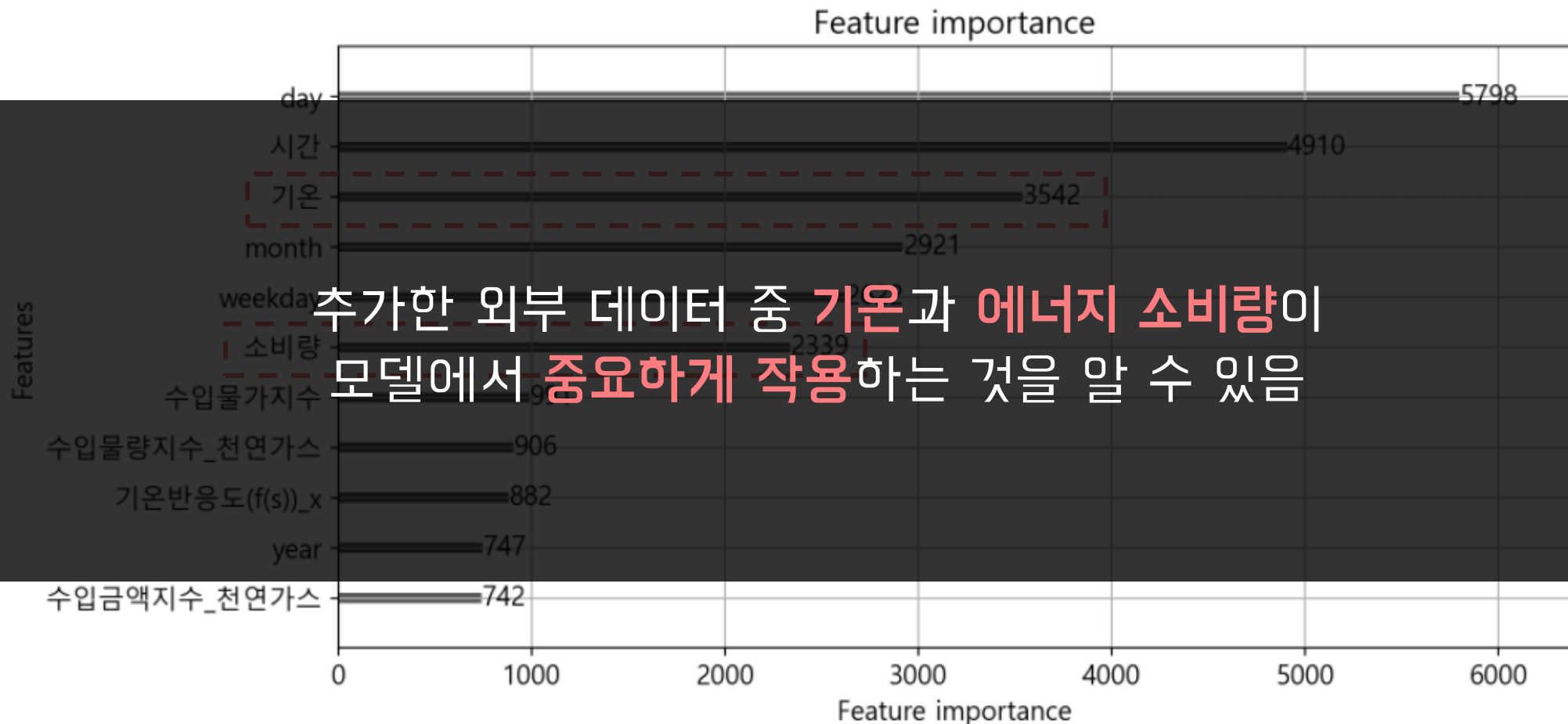


## 03-5 Feature Importance

가스레인지 X 메디치교육센터



### - Feature Importance



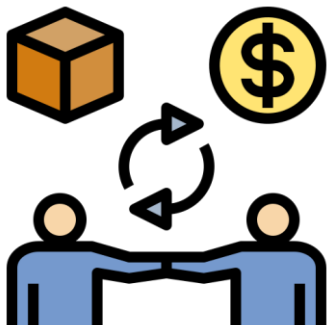


## 03 시사점 및 한계점

한국가스공사 x 가스레인지



### - 시사점



보다 정확한 수요전망대비  
공급량 예측 가능  
→ 에너지의 안정적인 수급 가능



천연가스 수입 시 스팟 구매 물량과  
중장기 계약물량의 가격차이 (5~10%차이)  
→ 비용 절감에 따른 국부 강화



에너지 대란으로  
발생가능한 물류, 곡물 등의  
인플레이션 최소화

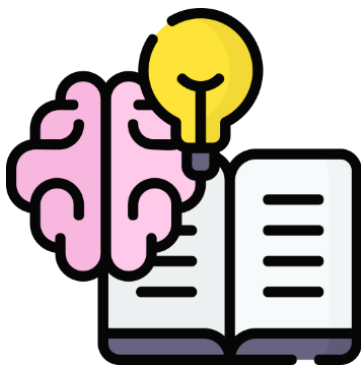


## 03-5 시사점 및 한계점

가스레인지 X 메디치교육센터



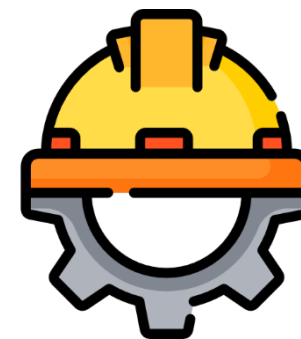
### - 한계점



프로젝트 주제에 대한  
도메인 지식 한계



외부 데이터 변수를 예측하여 사용하여  
변수의 예측값과 실제값의 오차로 인한  
분석 모형의 성능 하락 우려



추가된 데이터 사이  
강한 상관관계가 나타남  
⇒ feature engineering을 통한  
공분산성 해결 필요



## CONTENTS

---

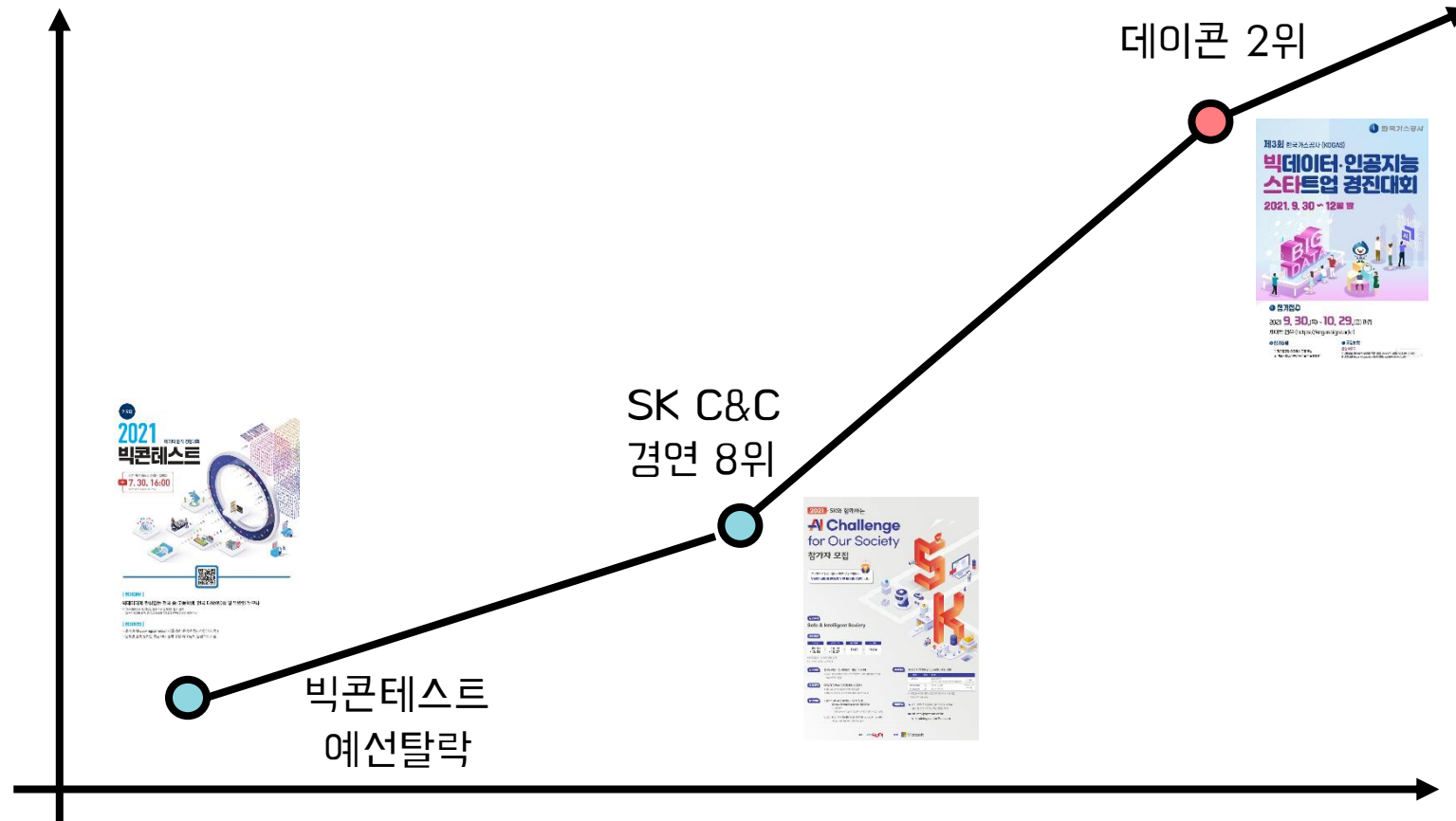
# 04

프로젝트 수행 소감



## 04 프로젝트 수행소감

가스레인지 X 메디치교육센터



**Q & A**

**감사합니다:)**





## 참고자료

---

### - 기사 인용

<http://www.kharn.kr/news/article.html?no=2652>

<http://amenews.kr/news/view.php?idx=45488>

<http://www.e2news.com/news/articleView.html?idxno=216926>

<http://www.gasnews.com/news/articleView.html?idxno=82668>

### - 분석 및 모델 관련 자료

<https://scikit-learn.org/stable/>

<https://www.tensorflow.org/?hl=ko>

<https://facebook.github.io/prophet/>

<https://www.statsmodels.org/stable/index.html>

### - 이미지 및 아이콘 출처 자료

<https://www.flaticon.com/>



# APPENDIX



## - 상관관계 확인

- : 기온반응도, 난방지수 및 공급량 파생변수는 공급량과 양의 상관관계를 가짐
- : 계절, 기온은 공급량과 음의 상관관계를 가짐
- : 추가한 외부데이터 사이 높은 양의 상관관계를 보임

